



# An iterative action minimizing method for computing optimal paths in stochastic dynamical systems

Brandon S. Lindley\*, Ira B. Schwartz

U.S. Naval Research Laboratory, Code 6792, Plasma Physics Division, Nonlinear Systems Dynamics Section, Washington, DC 20375, USA

## HIGHLIGHTS

- A new efficient method for computing switching rates for a large class of stochastic dynamical systems is presented.
- The method computes optimal transition paths for general stochastic problems in finite and infinite dimensions.
- The method is validated on physical switching problems as well as noise-induced extinction problems.

## ARTICLE INFO

### Article history:

Received 15 October 2012

Received in revised form

22 February 2013

Accepted 7 April 2013

Available online 12 April 2013

Communicated by J. Garnier

### Keywords:

Stochastic dynamical systems

Transition-path theory

Optimal paths

Stochastic differential equations

## ABSTRACT

We present a numerical method for computing optimal transition pathways and transition rates in systems of stochastic differential equations (SDEs). In particular, we compute the most probable transition path of stochastic equations by minimizing the effective action in a corresponding deterministic Hamiltonian system. The numerical method presented here involves using an iterative scheme for solving a two-point boundary value problem for the Hamiltonian system. We validate our method by applying it to both continuous stochastic systems, such as nonlinear oscillators governed by the Duffing equation, and finite discrete systems, such as epidemic problems, which are governed by a set of master equations. Furthermore, we demonstrate that this method is capable of dealing with stochastic systems of delay differential equations.

Published by Elsevier B.V.

## 1. Introduction

One important aspect of the study of dynamical systems is the study of noise on the underlying deterministic dynamics [1,2]. Although one might expect the deterministic dynamics to be only slightly perturbed in the presence of small noise, there are now many examples where noise causes a dramatic measurable change in behavior, such as noise induced switching between attractors in continuous systems [3], and noise induced extinction in finite size systems [2].

In systems transitioning between coexisting stable states, much research has been done primarily because switching can be now investigated for a large variety of well-controlled micro- and mesoscopic systems, such as trapped electrons and atoms, Josephson junctions, and nano- and micro-mechanical oscillators [4–13]. In these systems, observed fluctuations are usually due to thermal or externally applied noise. However, as systems become smaller, an increasingly important role may be played also by non-Gaussian

noise. It may come, for example, from one or a few two-state fluctuators hopping at random between the states, in which case the noise may be often described as a telegraph noise. It may also be induced by Poisson noise [14].

In finite size populations or systems, extinction occurs in discrete, finite populations undergoing stochastic effects due to random transitions or perturbations. The origins of stochasticity may be internal to the system or may arise from the external environment [15,16], and in most cases is non-Gaussian [17,18].

Extinction depends on the nature and strength of the noise [19], outbreak amplitude [20] and seasonal phase occurrence [21]. For large populations, the intensity of internal population noise is generally small. However, a rare, large fluctuation can occur with non-zero probability and the system may be able to reach the extinct state. Since the extinct state is absorbing due to effective stochastic forces, eventual extinction is guaranteed when there is no source of reintroduction [22,23,1].

Models of finite populations, which include extinction processes, are effectively described using the master equation formalism, and predict probabilities of rare events [17]. For many problems involving extinction in large populations, if the probability distribution of the population is quasi-stationary, the probability of extinction is a function that decreases exponentially with

\* Corresponding author. Tel.: +1 202 404 8358; fax: +1 202 767 0631.

E-mail address: [brandon.lindley\\_ctr@nrl.navy.mil](mailto:brandon.lindley_ctr@nrl.navy.mil) (B.S. Lindley).

increasing population size. The exponent in this function scales as a deterministic quantity called the action [24]. It can be shown that a trajectory that brings the system to extinction is very likely to lie along a most probable path, called the optimal path. It is a major property that a deterministic quantity such as the action can predict the probability of extinction, which is inherently a stochastic process, and is also formulated in continuous systems driven by noise [25,26].

Locating the optimal path is important since the quantity of interest, whether it is the switching or extinction rate, depends on the probability to traverse this path. Therefore, a stochastic control strategy based on the switching or extinction rates can be determined by its effect on the optimal path [26].

The optimal path formalism converts the entire stochastic problem to a mechanistic dynamical systems problem with definitive properties. First, the optimal path is a solution to a Hamiltonian dynamical system. In the case of continuous stochastic models, the dimension of the system is twice that of the original stochastic problem. The other dimensions are conjugate momenta, and typically represent the physical force of the noise which induces escape from a basin of attraction to either switch or go extinct. Finally, due to the symplectic structure of the resulting Hamiltonian system, it can be shown that both attractors and saddles of the original system become saddles of the Hamiltonian system.

One of the main obstacles to finding the optimal path is that it is an inherently unstable object. That is, if one starts near the path described by the Hamiltonian system, then after a short time, the dynamics leave the neighborhood of the path. In addition, although the path may be hyperbolic near the saddle points, it may not be hyperbolic along the rest of the path. Solving such problems using shooting methods for simple epidemic models [27–29], or mixed shooting using forward and backward iteration [30,31] will in general be inadequate to handle even the simplest unstable paths in higher dimensions. Therefore, it is the goal of this paper to exemplify a robust numerical method to solve for the optimal path using a general accurate discrete formulation applied to the Hamiltonian two point boundary value problem.

The method we employ here to compute the optimal paths is similar to the generalized minimum action method (gMAM), [32] which is a blend of the string method [33–35] and minimum action method [36]. Both are iterative methods which globally minimize the action along the path. These other techniques differ from ours primarily in that our formulation of the problem allows a direct, fully explicit iterative scheme, while the gMAM, in particular, employs a semi-implicit scheme. The numerical scheme presented here should provide an easy to employ alternative to the methods discussed above for stochastic optimization problems which are formulated as Hamiltonian two-point boundary value problems.

The paper is organized as follows. We first briefly present the general SDE problem, and the formulation of the corresponding deterministic Hamiltonian system by treating the switching as a rare event. We then present the details of the numerical approximation technique we will use to find the path which maximizes the probability of switching. Using this technique, we then demonstrate finding the optimal path from a stable focus to the unstable saddle for the unforced Duffing equation, then compute optimal extinction pathways for a simple epidemic model, and finally adapt our method to find optimal transition paths in stochastic delay differential systems.

## 2. General problem

Consider a general stochastic differential equation of the form

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t)) + \mathbf{G}(\mathbf{x}(t))\boldsymbol{\xi}(t), \quad (1)$$

where  $\mathbf{x} \in \mathbb{R}^n$  represents the physical quantity in state space and the matrix  $\mathbf{G}^1$  is given by  $\mathbf{G}(\mathbf{x}(t)) = \text{diag}\{g_1(\mathbf{x}(t)), g_2(\mathbf{x}(t)), \dots, g_n(\mathbf{x}(t))\}$ , where the  $g_i$ 's are general nonlinear functions. We suppose the noise  $\boldsymbol{\xi} \in \mathbb{R}^n$  is a vector having a Gaussian distribution with intensity  $D$ , and independent components. It is characterized by its probability density functional  $\mathcal{P}_\xi = e^{-\mathcal{R}_\xi/D}$ ,

$$\mathcal{R}_\xi[\boldsymbol{\xi}(t)] = \frac{1}{2} \int dt dt' \boldsymbol{\xi}(t)\boldsymbol{\xi}(t'). \quad (2)$$

We wish to determine the path with the maximum probability of traveling from the initial state  $\mathbf{x}_A$  to the final state  $\mathbf{x}_B$ , where the initial and final states are equilibria of the noise-free (i.e.  $\boldsymbol{\xi} = \mathbf{0}$ ) version of Eq. (1) given by  $\mathbf{f}(\mathbf{x}) = \mathbf{0}$ . These states typically characterize a generic problem of study in stochastic systems, such as switching between attractors, escape from a basin of attraction, or extinction of population. We assume the noise intensity  $D$  is sufficiently small so that in our analysis sample paths will limit on an optimal path as  $D \rightarrow 0$ . We also remark that  $\xi$  is formally the time derivative of a Brownian motion, sometimes referred to as white noise [37].

For  $D$  sufficiently small compared to an effective barrier, and examining the tail of the distribution for a large fluctuation (which is assumed to be a rare event), the probability of observing such a large fluctuation scales exponentially by [3,38],

$$\mathcal{P}_x = e^{-R/D}, \quad R = \min \mathcal{R}(\mathbf{x}, \boldsymbol{\xi}, \mathbf{p}), \quad (3)$$

where,

$$\mathcal{R}(\mathbf{x}, \boldsymbol{\xi}, \mathbf{p}) = R_\xi + \int \mathbf{p} \cdot [\dot{\mathbf{x}} - \mathbf{f}(\mathbf{x}) - \mathbf{G}(\mathbf{x}(t))\boldsymbol{\xi}], \quad (4)$$

where the Lagrange multipliers  $\mathbf{p}$  also correspond to the conjugate momenta of the equivalent Hamilton–Jacobi formulation of this problem.<sup>2</sup> The exponent  $R$  of Eq. (3) is called the action, and corresponds to the minimizer of the action in the Hamilton–Jacobi formulation which occurs along the optimal path. This path will minimize the integral of Eq. (4), and is found by setting the variations along the path  $\delta\mathcal{R}$  to zero. The transition rate exponent is proportional to the ratio of the action to the noise intensity,  $R/D$ .

The resulting equations of motion for the states and Lagrange multipliers are given by

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{f}(\mathbf{x}(t)) + \mathbf{G}^2(\mathbf{x})\mathbf{p} \\ \dot{\mathbf{p}} &= -\mathbf{G}(\mathbf{x}) \frac{\partial \mathbf{G}}{\partial \mathbf{x}}(\mathbf{x})\mathbf{p}\mathbf{p} - \frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}}\mathbf{p}. \end{aligned} \quad (5)$$

Here  $[\frac{\partial \mathbf{G}}{\partial \mathbf{x}}(\mathbf{x})\mathbf{p}\mathbf{p}]_i = [\frac{\partial \mathbf{G}}{\partial \mathbf{x}}(\mathbf{x})]_{ijk} [\mathbf{p}]_j [\mathbf{p}]_k$ , where Einstein summation is assumed over repeated indices. Note that  $\mathbf{p} = \mathbf{0}$  is invariant, and recovers the noise free case of Eq. (1). The above equations can be shown to satisfy the motion of a Hamiltonian system with Hamiltonian

$$H(\mathbf{x}, \mathbf{p}) = \frac{(\mathbf{G}^2(\mathbf{x})\mathbf{p}) \cdot \mathbf{p}}{2} + \mathbf{p} \cdot \mathbf{f}(\mathbf{x}). \quad (6)$$

That is, the dynamics of a given path satisfy  $\dot{\mathbf{x}} = \frac{\partial H(\mathbf{x}, \mathbf{p})}{\partial \mathbf{p}}$  and  $\dot{\mathbf{p}} = -\frac{\partial H(\mathbf{x}, \mathbf{p})}{\partial \mathbf{x}}$ . The conjugate momenta can be thought of as the effective force the noise imparts on the dynamics along the path.

In addition to solving the Hamiltonian system of dynamics, the full problem specification of an optimal path requires boundary conditions for both state  $\mathbf{x}$  and momenta  $\mathbf{p}$ . The boundary conditions of the optimal path consist of two steady states,  $X_A = (\mathbf{x}_A, \mathbf{p}_A)$

<sup>1</sup> Throughout the paper, boldface lower-case letters will indicate vectors, while boldface upper-case letters will indicate matrices.

<sup>2</sup> The vector multiplication here is assumed to be an inner product.

and  $X_B = (\mathbf{x}_B, \mathbf{p}_B)$  at equilibrium. Typically, the boundary conditions are derived from the equations of motion defined by Eq. (6). Since they are derived as steady state conditions, they are asymptotic boundary conditions that are infinite limits in the temporal line. In addition, since we have assumed the Hamiltonian is time-independent, energy is conserved, and the path must lie on a fixed energy surface. Notice, in the case where the Hamiltonian is time invariant, and given that the action is minimized along the optimal path, the Hamilton–Jacobi equations require a zero-energy constraint; i.e.,  $H(\mathbf{x}, \mathbf{p}) = 0$ . We now describe how to solve the Hamiltonian system as a two point boundary value problem on a restricted energy surface.

### 2.1. Stability of steady state solutions

As stated above, we seek the optimal path between two steady state solutions  $X_A = (\mathbf{x}_A, \mathbf{p}_A)$  and  $X_B = (\mathbf{x}_B, \mathbf{p}_B)$  where  $\mathbf{x}_A$  and  $\mathbf{x}_B$  are steady state solutions of the zero-noise case of Eq. (1), and thus satisfy  $\mathbf{f}(\mathbf{x}_A) = \mathbf{f}(\mathbf{x}_B) = \mathbf{0}$ . Depending upon the linear stability of the zero-noise steady states, the stochastic transition from  $\mathbf{x}_A$  to  $\mathbf{x}_B$  may or may not be a rare event. For example, if the stability matrix of the general SDE, Eq. (1),  $\mathbf{A} = \frac{\partial \mathbf{f}(\mathbf{x}_A)}{\partial \mathbf{x}} = \mathbf{f}'(\mathbf{x}_A)$  has at least one eigenvalue with positive real part, then it is an unstable, or saddle, point in the stochastic equation. Further, if  $\mathbf{f}'(\mathbf{x}_B)$  has eigenvalues with all negative real parts, then it is a stable focus, and the transition from  $\mathbf{x}_A$  to  $\mathbf{x}_B$  will be deterministic restricted to  $\mathbf{p} = \mathbf{0}$  and therefore not a rare event. The rare event, in this case, would be a transition from  $\mathbf{x}_B$  to  $\mathbf{x}_A$ , meaning noise drives the system out of the stable focus and onto the unstable/saddle point.

It is worth noting that all steady state solutions of the Hamiltonian system Eq. (5) which correspond to the deterministic steady state solutions of Eq. (1), i.e.  $\mathbf{f}(\mathbf{x}) = \mathbf{0}$ , are saddle points. It is easy to see why this is true in the case of additive noise, i.e.  $\mathbf{G} = \mathbf{I}$ . We can classify the steady states of Eq. (5) by calculating the eigenvalues of its stability matrix  $\mathbf{Q}$

$$\mathbf{Q} = \begin{bmatrix} \mathbf{f}'(\mathbf{x}_{A,B}) & \mathbf{I} \\ -\mathbf{f}''(\mathbf{x}_{A,B})\mathbf{p}_{A,B} & -\mathbf{f}'(\mathbf{x}_{A,B}) \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{I} \\ \mathbf{0} & -\mathbf{A} \end{bmatrix}, \quad (7)$$

at the steady states. The eigenvalues  $\gamma_k$  of  $\mathbf{Q}$  are given by,

$$\begin{bmatrix} \mathbf{A} & \mathbf{I} \\ \mathbf{0} & -\mathbf{A} \end{bmatrix} \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{A}\mathbf{z}_1 + \mathbf{z}_2 \\ -\mathbf{A}\mathbf{z}_2 \end{bmatrix} = \gamma_k \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{bmatrix}. \quad (8)$$

If we assume that  $\mathbf{z}_2 = \mathbf{0}$ , then  $\mathbf{A}\mathbf{z}_1 = \gamma_k \mathbf{z}_1$ . Thus, if  $\lambda_k \in \mathbb{R}$  is an eigenvalue of  $\mathbf{A}$ , the stability matrix of the SDE problem given by Eq. (1), then  $\gamma_k = \lambda_k$  is an eigenvalue of  $\mathbf{Q}$  with eigenvector  $[\mathbf{z}_1, \mathbf{0}]^T$ . Similarly, if  $\mathbf{z}_1 = \mathbf{0}$ , then  $-\mathbf{A}\mathbf{z}_2 = \gamma_k \mathbf{z}_2$ , which implies that  $\gamma_k = -\lambda_k$  is an eigenvalue of  $\mathbf{Q}$  with eigenvector  $[\mathbf{0}, \mathbf{z}_2]^T$ . Since  $\pm\lambda_k$  are eigenvalues of  $\mathbf{Q}$ , we can conclude that every steady state solution of Eq. (5) has eigenvalues with both positive and negative real parts and, thus, is a saddle. Thus, every steady state solution whose linearization has non-zero real part of the general stochastic equation, Eq. (1) regardless of stability becomes a saddle point when the system is converted to a Hamiltonian system. That is, deterministic attractors and saddles map to saddles in the Hamiltonian formulation.

### 3. Numerical scheme

Our numerical approach involves using a finite differences scheme to write the system of ODEs as a high dimensional algebraic system, to which we apply a modified Newton's Method to minimize the residual error until a solution is reached (to within some desired tolerance). As before, assume that the Hamiltonian system in  $\mathbb{R}^{2n}$  admits two steady states,  $X_A$  and  $X_B$ .

Then we seek the optimal path on the zero-energy surface that connects  $X_A$  to  $X_B$ . In this formulation, one would expect such a path to exhibit several properties when parametrized along  $t \in (-\infty, \infty)$ . First, we assume that the path starts at  $X_A$  at  $t = -\infty$ . Since this point is an equilibrium solution, we should expect that the solution stays very near this value, that is, there exists  $\epsilon > 0$  such that  $|X_A - X(t)| < \epsilon$  for  $-\infty < t \leq -T_\epsilon$ , has a transition region from  $-T_\epsilon < t < T_\epsilon$  and finally stays near  $X_B$ , the second steady state for,  $T_\epsilon \leq t < \infty$ . Numerically, we will approximate the solution on the finite domain  $-T_\epsilon \leq t \leq T_\epsilon$  so that the value at  $X(\pm T_\epsilon)$  is arbitrarily close to the steady solution.

We map the interval  $[-T_\epsilon, T_\epsilon]^3$  onto  $[0, 1]$  using the linear transformation  $t = 2T_\epsilon \bar{t} - T_\epsilon$ , and drop the “bar” notation for readability. On this discrete time domain, we write the system of ODEs as a system of nonlinear algebraic equations using central differences.

The simplest method is to employ a finite step size  $h = 1/N$  and use a uniform time step subdividing  $[0, 1]$  into  $N + 1$  equal segments. In practice, however, the simple uniform step size is not always the best choice since the optimal path tends to stay very near the stable points throughout most of the domain, and sometimes makes a relatively sharp transition near the center of the domain. In this case, it is helpful to use a nonuniform grid to resolve the sharp transition region using a fine mesh, and to use a coarse mesh near the edges where the solution is mostly flat. That is, the mesh is designed to have high resolution in a fast time boundary layer, and coarse resolution in quasi-stationary regions. We remark that the boundary layer is fixed by the initial data and the vector field. Thus, for the nonuniform time step  $h_k$ , yielding the time series  $t_{k+1} = t_k + h_k$  and corresponding function values  $\mathbf{x}_k$ , the derivative is approximated by the operator  $\delta_h$ ,

$$\frac{d}{dt}\mathbf{x}_k \approx \delta_h \mathbf{x}_k \equiv \frac{h_{k-1}^2 \mathbf{x}_{k+1} + (h_k^2 - h_{k-1}^2) \mathbf{x}_k - h_k^2 \mathbf{x}_{k-1}}{h_{k-1} h_k^2 + h_k h_{k-1}^2}. \quad (9)$$

At this point, we can write the generic system of  $2n(N + 1)$  nonlinear algebraic equations:

$$\delta_h \mathbf{x}_k - \frac{\partial H(\mathbf{x}_k, \mathbf{p}_k)}{\partial \mathbf{p}} = 0 \quad \delta_h \mathbf{p}_k + \frac{\partial H(\mathbf{x}_k, \mathbf{p}_k)}{\partial \mathbf{x}} = 0, \quad (10)$$

$$k = 0, 1, \dots, N,$$

and solve this system using a general Newton's Method for nonlinear systems of equations. To properly apply Newton's method here, let  $\mathbf{q}_j = \{\mathbf{x}_{1,j}, \dots, \mathbf{x}_{N,j}, \mathbf{p}_{1,j}, \dots, \mathbf{p}_{N,j}\}^T$  be the extended vector of dimension  $1 \times 2nN$  containing the  $j$ -th Newton iterate (recalling that  $\mathbf{x}_{k,j}, \mathbf{p}_{k,j} \in \mathbb{R}^n$  are defined on the time series given by  $k = 0, 1, \dots, N$ ). Then  $j = 0$  will represent the initial guess. Let  $\mathcal{F} : \mathbb{R}^{2nN} \rightarrow \mathbb{R}^{2nN}$  be the function defined by Eq. (10) acting on  $\mathbf{q}_j$ . Then to find the zeros of  $\mathcal{F}(\mathbf{q})$  we employ a Newton scheme. A new Newton iterate is given by solving the linear system  $J_{\mathcal{F}}(\mathbf{q}_j)(\mathbf{q}_{j+1} - \mathbf{q}_j) = -\mathcal{F}(\mathbf{q}_j)$ , using any one of a variety of methods such as LU decomposition or the generalized minimal residual method (GMRES) with appropriate preconditioners. Throughout this paper we will use LU decomposition with partial pivots optimized for a sparse linear system. Here the Jacobian  $J_{\mathcal{F}}(\mathbf{q}_n)$  may be computed approximately using a central difference scheme. We note that the real Jacobian may also be used without any change to the algorithm or performance. However, if the model is high-dimensional, it is more efficient to use finite differencing. In addition, the solution to the boundary value problem lies in a Banach

<sup>3</sup> For the simulations below,  $|T_\epsilon| \geq 100$  unless otherwise specified. Ideally,  $T_\epsilon$  is picked large enough so that  $|\mathbf{x}_A - \mathbf{x}(-T_\epsilon)| < 10^{-16}$  and  $|\mathbf{x}_B - \mathbf{x}(T_\epsilon)| < 10^{-16}$ , i.e. the steady states are obtained up to machine precision for double precision numbers.

space, so it is natural to approximate the space with a finite dimensional formulation, which we have extended to the Jacobian.

Formally, this method is second order with respect to  $h_k$ . The initial guess for this algorithm is constructed by the knowledge that the optimal path spends most of its time near the stable equilibria, and has a brief but sometimes sharp transition between the two states. One choice that has worked in practice is using functions like  $\mathbf{q}_{0,k} = (\mathbf{x}_A - \mathbf{x}_B)/(1 + e^{Ct_k}) + \mathbf{x}_B$ , with  $k = 0, 1, \dots, N$  (where the  $C > 0$  parameter adjusts the sharpness of the jump), which have horizontal asymptotes at the appropriate critical values. Usually, though not always,  $\mathbf{q}_{0,k}$  is set so that for  $k = N+1, N+2, \dots, 2N$ ,  $\mathbf{q}_{0,k} = \mathbf{0}$ . Another possible choice for  $\mathbf{q}_{0,k}$  with  $k > N$ , i.e. the  $\mathbf{p}$  variables, is an initial guess generated using a similar function that connects the appropriate boundary conditions.

Note that the zero-energy surface constraint  $H(\mathbf{x}, \mathbf{p}) = 0$  is not imposed. Rather, the initial guess will start out near  $X_A$  (at  $t = 0$ ) and  $X_B$  (at  $t = 1$ ) which lie asymptotically close to the zero energy surface, and thus the final solution will have to lie on this surface since such a solution is time invariant (i.e.  $d/dtH(\mathbf{x}, \mathbf{p}) = 0$ ). At each iterate, both the residual error and the Hamiltonian are checked at each point, and both must reach a desired tolerance in the  $\mathcal{L}_\infty$  norm before the procedure is completed.

Once the optimal path is computed, the action (i.e. the exponent of the distribution) along the optimal path may be obtained with a simple integral,

$$R = \int_{t=0}^1 \left[ \mathbf{p}(t) \cdot \frac{d\mathbf{x}(t)}{dt} - H(\mathbf{x}(t), \mathbf{p}(t)) \right] dt. \quad (11)$$

Exhaustive convergence tests on the residual error for a variety of test problems for both the uniform and non-uniform grids have demonstrated the second order convergence for this method. We have noted some dependence on the initial guess in terms of the overall speed of convergence (or divergence for a particularly bad guess). The method generally produces a unique solution (up to possibly a horizontal shift in the time series seen in a few examples, which does not affect the path integrals of interest). This method has been reliable for a wide parameter regime for each of the test problems, with the limitations to be discussed below on a case by case basis.

This method, which we will henceforth refer to as the Iterative Action Minimization Method (IAMM), has several distinct advantages over other methods, the foremost of which is straightforward scalability to higher dimensions. For very high dimensional problems, the systems will eventually become too large to treat easily with a single processor, but this algorithm has proven efficient for up to six dimensional problems with a single processor. Further, this method lends itself to infinite dimensional problems, such as time delay stochastic differential equations, as will be demonstrated below.

#### 4. Noise induced transitions

We demonstrate the numerical techniques by examining several bistable dynamical systems. Using methods discussed above, we explicitly approximate the optimal path between the two states and then numerically integrate along the path directly to compute the action.

##### 4.1. Switching in the duffing equation

One of the standard nonlinear dynamical systems which exhibits bi-stability is Duffing's equation. This equation is used to model certain types of nonlinear damped oscillators, and here we consider the singularly perturbed and unforced version [39],

$$\begin{aligned} \dot{x} &= y + g_1(x, y)\xi_1(t), \\ \epsilon \dot{y} &= \alpha x - \beta x^3 - \delta y + g_2(x, y)\xi_2(t) \\ \dot{\epsilon} &= 0. \end{aligned} \quad (12)$$

Here  $\alpha$  and  $\beta$  control the size and nonlinear response of the restoring force, while  $\delta$  controls the friction or damping on the system. The terms  $\xi_1$  and  $\xi_2$  are uncorrelated white noise sources applied to the acceleration and velocity respectively. When the perturbation  $\epsilon \ll 1$ , fast and slow manifolds can be identified, while  $\epsilon = 1$  gives the unconstrained case. Rescaling time  $t' = (1/\epsilon)t$ , applying Eq. (6), and following the methodology above, we can write the system in the following general form,

$$\begin{aligned} H(x, y, p_x, p_y) &= \frac{\epsilon p_x^2 g_1(x, y)^2}{2} + \frac{p_y^2 g_2(x, y)^2}{2} \\ &+ \epsilon p_x y + p_y(\alpha x - \beta x^3 - \delta y) = 0, \end{aligned} \quad (13)$$

with corresponding equations of motion,

$$\begin{aligned} \dot{x} &= \epsilon(p_x g_1(x, y)^2 + y) \\ \dot{y} &= p_y g_2(x, y)^2 + \alpha x - \beta x^3 - \delta y \\ \dot{p}_x &= -\epsilon p_x^2 g_1(x, y) \frac{\partial}{\partial x} g_1(x, y) - p_y(\alpha - 3\beta x^2) \\ \dot{p}_y &= -p_y^2 g_2(x, y) \frac{\partial}{\partial y} g_2(x, y) - \epsilon p_x + \delta p_y. \end{aligned} \quad (14)$$

We will restrict our focus to the additive white noise case,  $g_1(x, y) = g_2(x, y) = 1$ .

Note that the Hamiltonian system emits three known steady states  $X_A = (0, 0, 0, 0)$  and  $X_B = (\pm\sqrt{\alpha/\beta}, 0, 0, 0)$ , all of which are saddle points. These steady states correspond to zero-noise critical points of Eq. (12),  $\mathbf{x}_A = (0, 0)$ , a saddle point, and  $\mathbf{x}_B = (\pm\sqrt{\alpha/\beta}, 0)$ , the centers of the stable foci. The zero-noise ( $p_x = p_y \equiv 0$ ) path from  $x_A$  to  $x_B$  is deterministic, as any solution perturbed from the saddle-node point will move along the solution curves and end up at either stable focus. A more interesting case is the optimal path from one of the focus points to the saddle-node point, which will require non-trivial momentum. Such momenta model the small noise effects which organize to force the trajectory across the basin of attraction, thus escaping from one attractor to the other.

Fig. 1 shows both the deterministic (zero-noise) and optimal paths as computed using the IAMM developed above. For the deterministic path, the algorithm correctly predicts that the noise will be zero along this path, and that the action will be zero as the probability of going from  $X_A$  to  $X_B$  is one, and thus not a rare event. On the other hand, the path from  $X_B$  to  $X_A$  involves nontrivial action, and the effect of the noise along the optimal path is shown. This path lies on the zero-energy surface, and maximizes the probability of traveling from  $X_B$  to  $X_A$  for arbitrarily small noise intensities,  $D$ . Fig. 2 shows the residual error at each iterate used to generate the data for Fig. 1 until the convergence criteria is reached.

When  $\epsilon \ll 1$ , we can derive an analytical formulation for the action by using a center manifold analysis on Eq. (13) by following the work of [39]. Here the center manifold is given by  $y = h(x, \epsilon)$ , and we approximate it as,

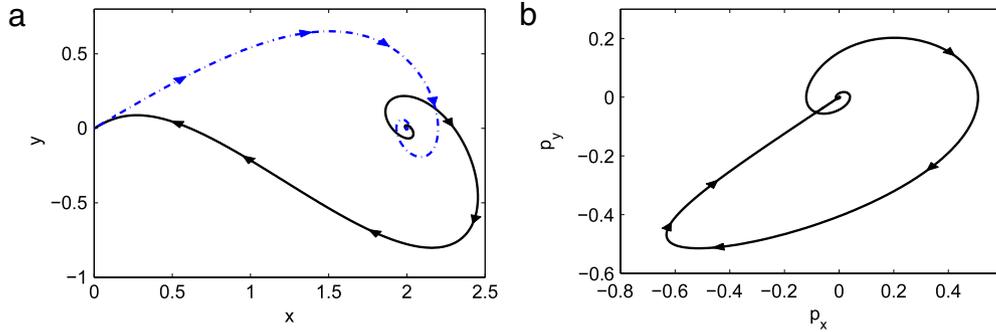
$$y = h(x, \epsilon) = h_0(x) + \epsilon h_1(x) + \mathcal{O}(\epsilon^2). \quad (15)$$

By substituting Eq. (15) into Eq. (12) and equating like powers of  $\epsilon$ , we arrive at a one dimensional form of Eq. (12) for the lowest order terms involving  $\epsilon$ ,

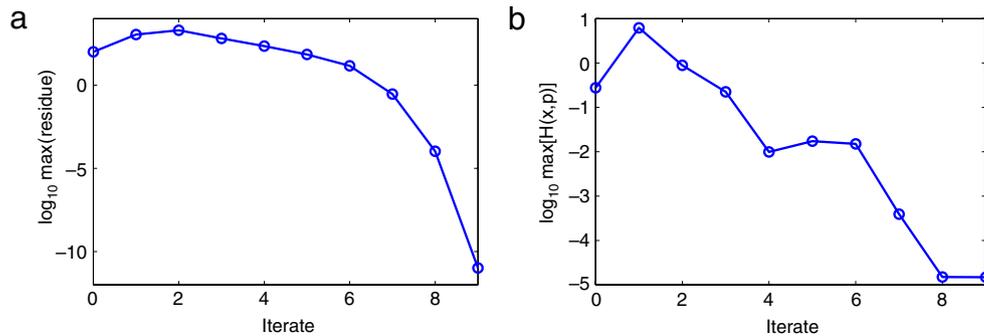
$$\dot{x} = \frac{\epsilon}{\delta} (\alpha x - \beta x^3 + \xi(t)). \quad (16)$$

Here the contribution of the uncorrelated noise terms  $\xi_{1,2}$  are contained in a single noise source  $\xi$ . The Hamiltonian form of (16) is,

$$H(x, p) = \frac{p^2}{2} + p \frac{\epsilon}{\delta} (\alpha x - \beta x^3). \quad (17)$$



**Fig. 1.** Projections of the optimal path (a) onto phase space and (b) onto momenta space. In (a) we show both the optimal (from  $(\sqrt{\alpha/\beta}, 0)$  to  $(0, 0)$ , labeled with a solid line) and deterministic path (from  $(0, 0)$  to  $(\sqrt{\alpha/\beta}, 0)$ , labeled with a dashed line) as predicted by our numerical method. Along the deterministic path, the action is zero and  $p_x$  and  $p_y$  are zero, but the corresponding noise components along the optimal path are shown in (b). Here  $\epsilon = 1$ ,  $\beta = .25$ ,  $\delta = 1$ , and  $\alpha = 1$ .



**Fig. 2.** The maximal residual error (a) and the Hamiltonian (b) of the numerical scheme as a function of iterate, where the initial guess is iterate 0. This figure shows the convergence for the data shown in Fig. 1 where the parameter values are  $\epsilon = 1$ ,  $\beta = .25$ ,  $\delta = 1$ , and  $\alpha = 1$ , and the error threshold is met after the ninth iteration. The maximal error is defined to be the maximum residue of all the time series for all the components.

We can find the nontrivial relationship ( $p \neq 0$ ) between  $p$  and  $x$  on the zero energy surface directly,  $p = -\frac{2\epsilon}{\delta}(\alpha x - \beta x^3)$ , and integrate along this path to predict the action along the optimal path. Since one may be interested in how the action scales relative to the distance between the two critical points, we substitute the values  $a^2 = \alpha$  and  $b^2 = \frac{1}{\beta}$  to better illustrate the scaling with respect to  $\beta$ . Thus, with this substitution, we seek the action from  $x = ab$  to  $x = 0$ . From Eq. (11), this is a simple integral,

$$R = -2\frac{\epsilon}{\delta} \int_{ab}^0 \left( a^2 x - \frac{x^3}{b^2} \right) dx = \frac{\epsilon}{2\delta} a^4 b^2. \quad (18)$$

Thus, we can predict, for example, that the action from  $(ab, 0)$  to  $(0, 0)$  should scale like the square of  $b$  near the center manifold. Indeed, varying the parameter  $b$  for the four-dimensional system, Eq. (14), predicts the same order scaling as seen below in Fig. 3. We also consider the scaling with respect to the damping parameter  $\delta$ , and again, near the center manifold, the predictions are born out by integrating along the optimal path. Indeed, the scaling for the action predicted from the lowest order terms in the center manifold analysis seems to persist in the two-dimensional model.

Since it is not clear if the same relationship will work further away from the center manifold, we check the scaling with respect to  $b$  and  $\delta$  when  $\epsilon = 1$  in Eq. (14) in Fig. 4. Interestingly, the leading order scaling of the action with respect to  $\beta$  is still 2, just as it was near the center manifold. Meanwhile, the scaling with respect to  $\delta$  is markedly different. Thus, we can predict both near and away from the center manifold how the action will scale as a function of the distance between the two equilibrium points.

#### 4.2. SIS epidemic models

Next we consider a simple susceptible–infected–susceptible (SIS) epidemic model in the general noise case, as considered

by [25]. This model is defined by a master equation which describes the probability of fluctuations between a susceptible or infected category in a population of  $N$  individuals. Assuming  $N$  is sufficiently large, we can write a mean field system of equations for the change of the population fractions of susceptible and infected individuals, denoted  $x_S$  and  $x_I$  respectively,

$$\begin{aligned} \dot{x}_S &= \mu - \beta x_S x_I + \kappa x_I - \mu x_S \\ \dot{x}_I &= \beta x_S x_I - \kappa x_I - \mu x_I. \end{aligned} \quad (19)$$

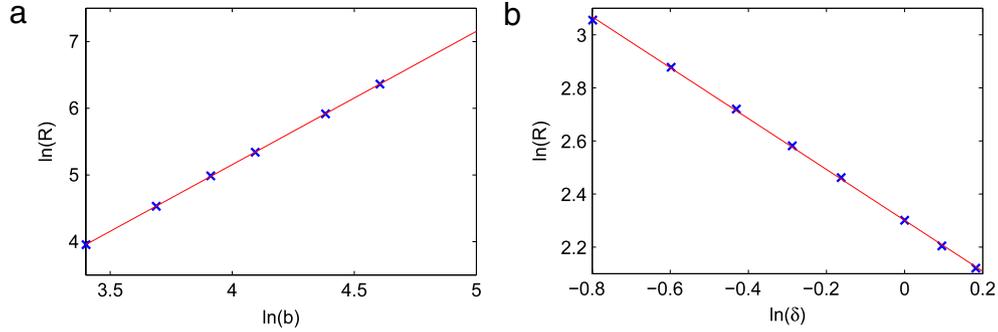
For simplicity, the population in the mean field is assumed constant, i.e. births and deaths are equal. Here  $\mu$  is the natural birth and death rate of both the susceptible and infected populations,  $\beta$  is the contact rate, and  $\kappa$  is the natural recovery rate of the infected population.

Since we have assumed the population fraction of susceptible and infected individuals are conserved, we have  $x_I = 1 - x_S$ . Under this constraint, assuming a small random fluctuation of only the infected individuals, we can use a methodology similar to the one described above (and worked out in detail in [25]) to write a Hamiltonian system,

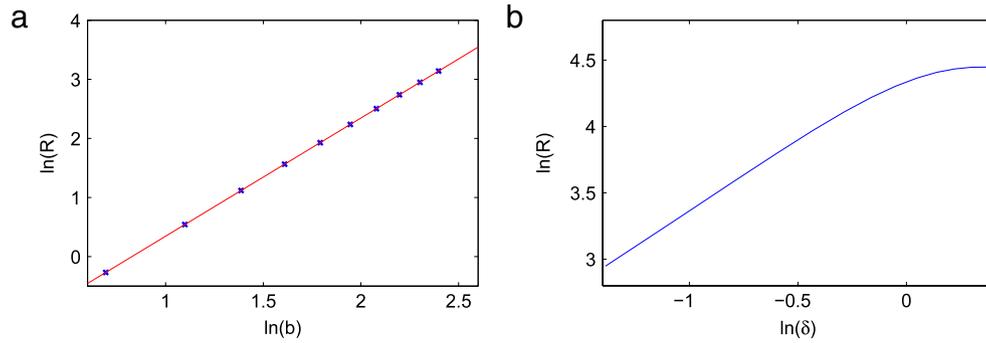
$$H(x_I, p_I) = \beta x_I (1 - x_I)(e^{p_I} - 1) + (\mu + \kappa)x_I(e^{-p_I} - 1). \quad (20)$$

We shall refer to Eq. (20) as the 1D SIS model equation. The optimal path will extend from the endemic state at  $x_I = 1 - 1/R_0$  to the extinct state  $x_I = 0$ , where  $R_0 = \beta/(\mu + \kappa)$ . Using the methods introduced above, the optimal path (at typical parameter values) is given below. Integrating the momentum along this path gives the action, which is proportional to the log of the probability of extinction. In this simple 1D SIS model, the predicted action scaling as a function of  $R_0$  is given by solving the Hamiltonian directly for  $p_I$  and solving the integral (11)

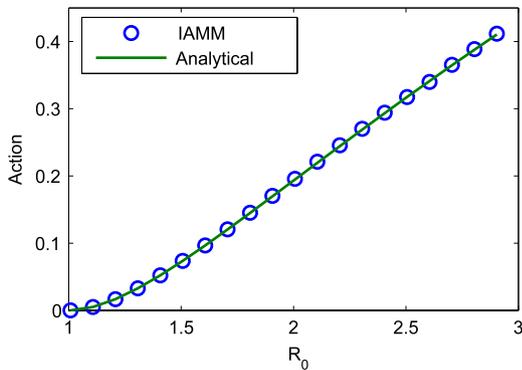
$$R = \int_{1-1/R_0}^0 -\ln[R_0(1 - x_I)] dx.$$



**Fig. 3.** The scaling of the action as a function of the parameters when  $\epsilon = .05$ . For the two sweeps, the action is obtained by using the IAMM on Eq. (14) with  $a = 1$  and  $\delta = 1$  while  $b$  is swept in (a) and with  $a = 1$  and  $b = 20$  while  $\delta$  is varied in (b). Since  $\epsilon$  is small, we expect the center manifold approximation of Eq. (18) to be valid. Indeed, the linear best fit in (a) is  $\ln(R) = 1.9989 \ln(b) - 2.8417$  while in (b) the equation is  $\ln(R) = -.958666 \ln(\delta) + 2.3008$ . The slope of both of these lines are close to the prediction given by Eq. (18).



**Fig. 4.** The scaling of the action as a function of the parameters when  $\epsilon = 1$ . For the two sweeps, the nonlinear term  $b$  is swept for fixed  $\delta = 1$  in (a) while  $b = 20$  is fixed while  $\delta$  is varied in (b). The linear best fit in (a) is  $\ln(R) = 2 \ln(b) - 1.655$  while in (b) the power law scaling vanishes as  $\delta$  is increased. The second order scaling of the action with respect to  $b$  is persistent even away from the center manifold.



**Fig. 5.** The action as a function of  $R_0$  as predicted by an analytical expression and our numerical method. Here,  $\kappa = 100$ , and  $\mu = .2$  while  $\beta$  is varied.

Since this is one of the rare cases the optimal path can be found analytically, it is a great test case for our method. A comparison of the analytical action to the numerical action is shown in Fig. 5.

In the case of an SIS epidemic model with independent fluctuations on both the susceptible ( $x_S$ ) and infected ( $x_I$ ) populations, the Hamiltonian form of the equations is given by,

$$\begin{aligned}
 H(x_S, x_I, p_S, p_I) = & \mu(e^{p_S} - 1) + \beta x_S x_I (e^{-p_S + p_I} - 1) \\
 & + \kappa x_I (e^{p_S - p_I} - 1) + \mu x_S (e^{-p_S} - 1) \\
 & + \mu x_I (e^{-p_I} - 1), \tag{21}
 \end{aligned}$$

and we refer to this form as the 2D SIS model. The two states of interest for this system are the endemic state,  $(1/R_0, 1 - 1/R_0, 0, 0)$  and the nontrivial extinct state,  $(1, 0, 0, \ln(1/R_0))$ . Using the procedure discussed above, we compute the optimal path from the endemic to the extinct state, and show a typical result in Fig. 6.

Instead of comparing the action scaling predicted here to an analytical formulation, we instead compare it to a Monte-Carlo simulation of the master equation for the initial system with a fixed population size of 20,000 individuals. In the Monte-Carlo simulation, a Gillespie algorithm is employed on the SIS model, and from all the simulations, a probability of extinction is computed for several values of  $R_0$ . From these, a mean extinction time is derived, and the log of this mean extinction time should scale like the action predicted from the action integrals, from Eq. (11), computed using our approach. Fig. 7 shows a comparison of the two approaches, and a good agreement is seen between these two independent methods.

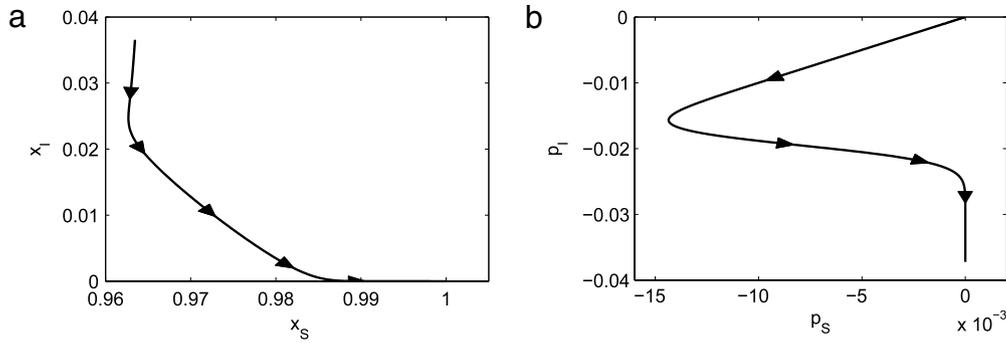
### 4.3. Finite time Lyapunov exponents

As demonstrated in [17], the optimal path to extinction coincides with ridges, i.e. maximal values, of the finite time Lyapunov exponents (FTLEs). Forgoston and others propose finding FTLE ridges as a method of computing optimal paths [17,18,40]. Here, we demonstrate that our approximation to the optimal path does indeed locally maximize the FTLE.

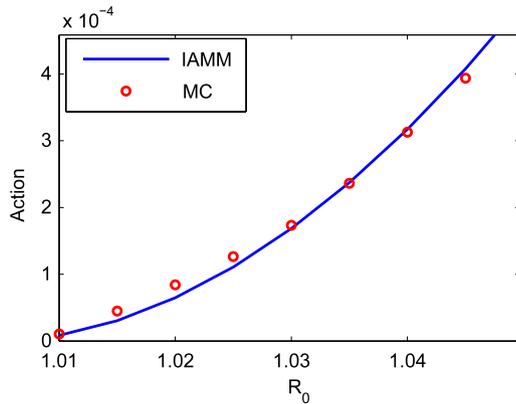
We proceed by using the methods outlined in [41–44] to approximate the FTLEs at points on our optimal path, and at nearby points transverse to the optimal path. We begin by picking a point on the optimal path (generated by our method), and on nearby points some small distance away from the path.

For the given vector field, we assume that we have a flow passing through initial point  $\mathbf{x}_0$ ,  $\phi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ , such that  $\phi_{t_0}^{t_0+T}(\mathbf{x}_0) = \mathbf{x}(t_0 + T; t_0, \mathbf{x}_0)$ . The local linear variation at  $\mathbf{x}_0$  is defined by  $\Delta(\mathbf{x}_0, t_0 + T) = \frac{\partial \phi_{t_0}^{t_0+T}(\mathbf{x}_0)}{\partial \mathbf{x}_0}$ .

Using a fourth order Runge–Kutta method, we can integrate all the initial points,  $\mathbf{x}_0$ , forward in time over a fixed interval, and



**Fig. 6.** Projection of the optimal path onto (a) population space and (b) momenta space for a sample SIS system with  $\beta = 104$ ,  $\kappa = 100$ , and  $\mu = .2$ .



**Fig. 7.** A comparison of the action predicted by computing the optimal path to a Monte-Carlo simulation of the original system. Here,  $\beta$  is varied while  $\kappa = 100$  and  $\mu = .2$ .

compute the finite time deformation rate of the local coordinates (i.e. Right Cauchy–Green Tensor)  $\mathbf{C}(\mathbf{x}_0) = \mathbf{\Delta}^T(\mathbf{x}_0, t_0 + T)\mathbf{\Delta}(\mathbf{x}_0, t_0 + T)$ . The maximal eigenvalue  $\lambda_{max}$  of  $\mathbf{C}(\mathbf{x}_0)$  will give the FTLE  $\sigma(y, t_i, T) = \frac{1}{T} \ln \sqrt{\lambda_{max}}$ .

Consider the case of the 1D SIS extinction model from above. Here, we will use a path computed above and compare the values near this path to the local FTLEs. In Fig. 8(a) we plot the FTLEs over a square domain, and show that a local maximum (ridge) is attained precisely where the computed optimal path predicts.

In higher dimensions, the maximal Lyapunov exponent is still exhibited along the optimal path [17]. For the 2D SIS model, note that the optimal path exists in four dimensions. Thus, the transverse direction is the set of points obtained by rotating a normal vector to the path around two Euler angles (which forms

a 3D sphere for a given radius  $r$ ). To illustrate FTLEs in this higher dimension framework, we must consider the “shell” around the initial starting point (and orthogonal to the path) at a fixed radius, and then compute FTLEs on this shell all along the optimal path. We expect that the maximal FTLE will occur along the optimal path, relative to nearby points on the transverse sphere. To illustrate this, we define  $u_{\perp}(t)$  as a unit vector orthogonal to the tangent vector of the optimal path at a given time, i.e.  $u_{\perp}(t) = (\dot{x}_i(t), -\dot{x}_s(t), 0, 0) / |(\dot{x}_i(t), -\dot{x}_s(t), 0, 0)|$ , and then examine the FTLE as a function of the two Euler angles for a given  $r$ . Fig. 8(b) shows just one cross section (over a short time interval), obtained by setting both rotation angles to either 0 or the antipodal angle  $\pi$ , of this high dimensional object, as a function of  $r$ , and demonstrates that the maximal FTLE is, indeed, along the optimal path.

## 5. Time delayed SDEs

One advantage of the IAMM is that it allows the solution of stochastic delay-differential equations of the form,

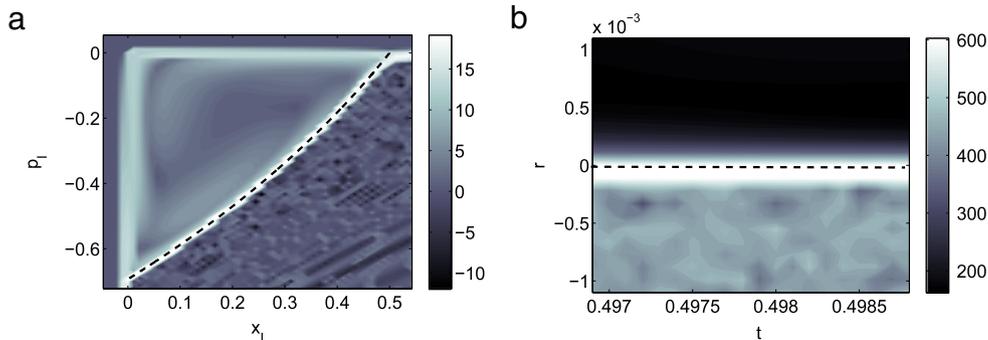
$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{x}(t - \tau)) + \mathbf{G}(\mathbf{x}(t))\boldsymbol{\xi}(t). \quad (22)$$

Schwartz et al. [45] have demonstrated that the methodology introduced in Section 2 can be adapted to write this system as a Hamiltonian system,

$$H(\mathbf{x}, \mathbf{x}_{\tau}, \mathbf{p}) = \frac{(\mathbf{G}^2(\mathbf{x})\mathbf{p}) \cdot \mathbf{p}}{2} + \mathbf{p} \cdot \mathbf{f}(\mathbf{x}, \mathbf{x}_{\tau}), \quad (23)$$

where  $\mathbf{x}_{\tau} = \mathbf{x}(t - \tau)$ . The equations of motion are given by,

$$\begin{aligned} \dot{\mathbf{x}} &= \frac{\partial H}{\partial \mathbf{p}}(\mathbf{x}, \mathbf{x}_{\tau}, \mathbf{p}) \\ \dot{\mathbf{p}} &= -\frac{\partial H}{\partial \mathbf{x}}(\mathbf{x}, \mathbf{x}_{\tau}, \mathbf{p}) - \frac{\partial H}{\partial \mathbf{x}_{\tau}}(\mathbf{x}(t + \tau), \mathbf{x}(t), \mathbf{p}(t + \tau)). \end{aligned} \quad (24)$$



**Fig. 8.** Figure (a) shows the numerically computed optimal path overlaid with the FTLE for the one-dimensional SIS model. The optimal path (dashed line) occurs precisely along a ridge of the FTLE. Figure (b) shows a single slice of the FTLE computation for the two-dimensional SIS model, where  $r$  is the distance from the optimal path in the transverse direction along a unit vector, and  $-r$  represented a distance in the antipodal direction. For both paths, the parameter values from Fig. 6 are used.

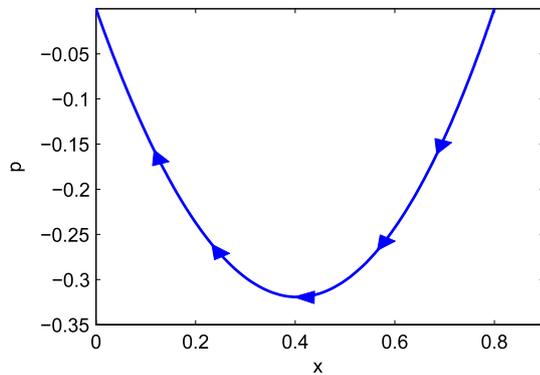


Fig. 9. The optimal path for Eq. (26) with  $\gamma = .2$  and  $\tau = .5$ .

Note the appearance of both delay and advance terms in Eq. (24). Because of the appearance of the delay term, the Hamiltonian is no longer time invariant, and unlike in the previous examples, where  $H(\mathbf{x}, \mathbf{p}) = 0$ , the zero-energy condition is not conserved.

We shall consider a one dimensional test case where  $f(x, x_\tau) = x(1-x) - \gamma x_\tau$ , where the steady states are given by  $x_A = 1 - \gamma$  and  $x_B = 0$ , for  $0 \leq \gamma \leq 1$ . Again, we will assume additive noise  $\mathbf{G} = 1$ , and derive the Hamiltonian system,

$$H(x, x_\tau, p) = \frac{p^2}{2} + p(x(1-x) - \gamma x_\tau), \quad (25)$$

and the corresponding equations of motion,

$$\begin{aligned} \dot{x} &= x(1-x) - \gamma x_\tau + p \\ \dot{p} &= -p(1-2x) + \gamma p(t + \tau). \end{aligned} \quad (26)$$

The IAMM needs only a few minor adjustments to compute these paths numerically. Primarily, the presence of the delay and advance terms will add additional entries into our linear system Eq. (10). Our method uses a non-uniform time step, and so  $x(t_k - \tau)$  may not coincide exactly with one of our points  $x_k$  at time  $t_k$ . To overcome this, we can use Lagrange interpolation on the closest four points  $x_{j-2}, x_{j-1}, x_j, x_{j+1}$  such that  $t_{j-2} < t_{j-1} < t_k - \tau < t_j < t_{j+1}$ . Since we keep the time domain fixed, the  $j$  needed for each  $t_k$  can be easily computed before the iterative scheme is started, and fewer or more terms can be used in the Lagrange interpolation scheme depending upon desired accuracy. Further, if  $t_k - \tau < T_\epsilon$  or  $t_k + \tau > T_\epsilon$ , i.e. the delay or advance terms fall outside of our numerical domain, then we can set  $x_k = x_A$  or  $x_k = x_B$  respectively.

To demonstrate the effectiveness of the IAMM in solving these stochastic delay problems, we show the optimal path in Fig. 9 and compare the scaling of the action versus a stochastic simulation of the stochastic delay difference equation in Fig. 10, and note the good agreement.

## 6. Discussion

We have considered the problem of finding the trajectory in stochastic dynamical systems that optimizes the probability of switching between two states, or causes one or more components to go extinct. In computing such a trajectory, called the optimal path, we needed to consider a numerical technique which could solve a Hamiltonian system with asymptotic boundary conditions in time.

We have developed a numerical method, which we call the iterative action minimizing method (IAMM), for finding the optimal path of transition between two steady states in stochastic dynamical systems. This method is ideal for systems which can be written as two-point boundary value problems governed by Hamiltonian systems. The dynamical structure of the solution to

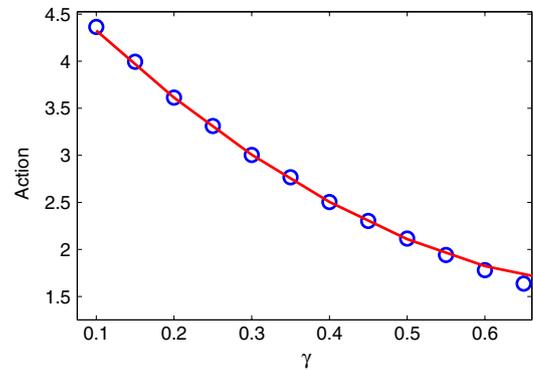


Fig. 10. Scaling of the action along the optimal path as a function of  $\gamma$ . Here the solid line indicates the IAMM prediction, and the circles represent a Monte-Carlo simulation of 1000 runs.

the Hamiltonian may be quantified by dynamical systems ideas such as local FTLE analysis [39]. The solution using the Hamiltonian formulation also need not be a gradient system, which is the case for many of the molecular-dynamics formulations. It also has been used to quantify the relaxation times explicitly when looking at high dimensional multi-time scale formulations. These rates are necessary to know when trying to solve for optimal paths such as in non-Gaussian switching problems, eg., [14].

We have validated the IAMM by presenting a variety of problems of interest, and have compared the numerical results with either analytic results or Monte-Carlo simulations of full stochastic systems. As demonstrated here, the IAMM method is robust enough to be applicable to a variety of different types of problems, including continuous SDE systems, such as the Duffing equation, discrete epidemic models of finite population size, such as the SIS model, and stochastic delay differential equations, in which the deterministic problem is infinite dimensional. The methodology is straightforward enough to generalize to higher dimensions, in contrast to other commonly used methods, such as the shooting method, which is a major advantage of the IAMM.

The primary limitations of this method are scaling issues in very high state space dimensions, and the finesse required in picking an initial guess that guarantees convergence, both of which are typical of iterative methods of quasi-Newton type. We note that a poor initial guess may lead to an ill-conditioned linear system [36] in some cases, though we have not observed this in our examples. Here we employ a preconditioned fast linear solver for sparse systems in the case of ill-conditioned systems. In the limit of small noise or large system size, however, due to the robustness and ease of generalization to complex and high dimensional dynamical systems, the method offers a considerable advantage over simulating large systems, or systems which require many Monte Carlo runs to generate statistics of the transitions paths. As a result, we expect this method will be useful in efficiently solving a large variety of optimal transition problems in the field of stochastic dynamical systems. In addition, since the action yields transition rates, the theory may be connected directly with experiments.

In the case where there exist multiple paths satisfying both the Hamiltonian equations of motion and the boundary conditions, the technique will find them given initial data that is sufficiently close to the paths. The optimal path in that case would be the one with the minimum action. The optimal path is the one that is the most likely path observed when running stochastic simulations.

When there exists a continuum of optimal paths, then it is likely that the numerical technique will need to be adjusted to account for solutions which are non-isolated. Singularities may develop in the Newton scheme, but quasi-Newton methods will help alleviate the issue of convergence, which will be less than a quadratic error rate [46].

## Acknowledgments

The authors gratefully acknowledge the Office of Naval Research for their support under N0001412WX20083, and support of the NRL Base Research Program N0001412WX30002. Brandon Lindley is currently an NRC Postdoctoral Fellow. We thank Lora Billings for providing the Monte Carlo data used in Figs. 7 and 10, and Eric Forgoston for a preliminary reading of this manuscript.

## References

- [1] C.W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*, Springer-Verlag, 2004.
- [2] N.G. Van Kampen, *Stochastic Processes in Physics and Chemistry*, Elsevier, 2007.
- [3] M.I. Freidlin, A.D. Wentzell, *Random Perturbations of Dynamical Systems*, Springer-Verlag, 1984.
- [4] L.J. Lapidus, D. Enzer, G. Gabrielse, Stochastic phase switching of a parametrically driven electron in a penning trap, *Phys. Rev. Lett.* 83 (1999) 899–902.
- [5] I. Siddiqi, R. Vijay, F. Pierre, C.M. Wilson, L. Frunzio, M. Metcalfe, C. Rigetti, R.J. Schoelkopf, M.H. Devoret, D. Vion, D. Esteve, Direct observation of dynamical bifurcation between two driven oscillation states of a josephson junction, *Phys. Rev. Lett.* 94 (2005) 027005.
- [6] J.S. Aldridge, A.N. Cleland, Noise-enabled precision measurements of a duffing nanomechanical resonator, *Phys. Rev. Lett.* 94 (2005) 156403.
- [7] K. Kim, M.S. Heo, K.H. Lee, H.J. Ha, K. Jang, H.R. Noh, W. Jhe, Noise-induced transition of atoms between dynamic phase-space attractors in a parametrically excited atomic trap, *Phys. Rev. A* 72 (2005) 053402.
- [8] R. Gommers, P. Douglas, S. Bergamini, M. Goonasekera, P.H. Jones, F. Renzoni, Resonant activation in a nonadiabatically driven optical lattice, *Phys. Rev. Lett.* 94 (2005) 143001.
- [9] C. Stambaugh, H.B. Chan, Noise activated switching in a driven, nonlinear micromechanical oscillator, *Phys. Rev. B* 73 (2006) 172302.
- [10] B. Abdo, E. Segev, O. Shtempluck, E. Buks, Escape rate of metastable states in a driven nbn superconducting microwave resonator, *J. Appl. Phys.* 101 (2007) 083909.
- [11] A. Lupaşcu, S. Saito, T. Picot, P.C. De Groot, C.J.P.M. Harmans, J.E. Mooij, Quantum non-demolition measurement of a superconducting two-level system, *Nat. Phys.* 3 (2007) 119–123.
- [12] I. Katz, A. Retzker, R. Straub, R. Lifshitz, Signatures for a classical to quantum transition of a driven nonlinear nanomechanical resonator, *Phys. Rev. Lett.* 99 (2007) 040404.
- [13] I. Serban, F.K. Wilhelm, Dynamical tunneling in macroscopic systems, *Phys. Rev. Lett.* 99 (2007) 137001.
- [14] L. Billings, I. Schwartz, M. McCrary, A. Korotkov, M. Dykman, Switching exponent scaling near bifurcation points for non-gaussian noise, *Phys. Rev. Lett.* 104 (2010) 140601.
- [15] F. de Castro, B. Bolker, Mechanisms of disease-induced extinction, *Ecol. Lett.* 8 (2005) 117–126.
- [16] A.L. Lloyd, J. Zhang, A.M. Root, Stochasticity and heterogeneity in host-vector models, *J. R. Soc. Interface* 4 (2007) 851–863.
- [17] I. Schwartz, E. Forgoston, S. Bianco, L. Shaw, Converging towards the optimal path to extinction, *J. R. Soc.* 8 (2011) 1699–1707.
- [18] E. Forgoston, S. Bianco, L.B. Shaw, I.B. Schwartz, Maximal sensitive dependence and the optimal path to epidemic extinction, *Bull. Math. Biol.* 73 (2011) 495–514.
- [19] B.A. Melbourne, A. Hastings, Extinction risk depends strongly on factors contributing to stochasticity, *Nature* 454 (2008) 100–103.
- [20] D. Alonso, A.J. McKane, M. Pascual, Stochastic amplification in epidemics, *J. R. Soc. Interface* 4 (2006) 575–582.
- [21] L. Stone, R. Olinky, A. Huppert, Seasonal dynamics of recurrent epidemics, *Nature* 446 (2007) 533–536.
- [22] M.S. Bartlett, Some evolutionary stochastic processes, *J. Roy. Stat. Soc. B Met.* 11 (1949) 211–229.
- [23] L.J.S. Allen, A.M. Burgin, Comparison of deterministic and stochastic SIS and SIR models in discrete time, *Math. Biosci.* 163 (2000) 1–33.
- [24] R. Kubo, K. Matsuo, K. Kitahara, Fluctuation and relaxation of macrovariables, *J. Stat. Phys.* 9 (1973) 51–96.
- [25] I.B. Schwartz, L. Billings, M. Dykman, A. Landsman, Predicting extinction rates in stochastic epidemic models, *J. Stat. Mech.—Theory E.* (2009) P01005.
- [26] M.I. Dykman, I.B. Schwartz, A.S. Landsman, Disease extinction in the presence of random vaccination, *Phys. Rev. Lett.* 101 (2008) 078101.
- [27] H. Keller, *Numerical Methods for Two-point Boundary-value Problems*, Dover Publications, 1992.
- [28] A. Kamenev, B. Meerson, Extinction of an infectious disease: a large fluctuation in a nonequilibrium system, *Phys. Rev. E* 77 (2008) 061107.
- [29] O. Gottesman, B. Meerson, Multiple extinction routes in stochastic population models, *Phys. Rev. E* 85 (2012) 021140.
- [30] V. Elgart, A. Kamenev, Rare event statistics in reaction–diffusion systems, *Phys. Rev. E* 70 (2004) 041106.
- [31] A.I. Chernykh, M.G. Stepanov, Large negative velocity gradients in burgers turbulence, *Phys. Rev. E* 64 (2001) 026306.
- [32] M. Heymann, E. Vanden-Eijnden, The geometric minimum action method: a least action principle on the space of curves, *Commun. Pure Appl. Math.* 61 (2008) 1052–1117.
- [33] W.E.R. Weiqing, E. Vanden-Eijnden, String method for the study of rare events, *Phys. Rev. B* 66 (2002) 052301.
- [34] H. Schwetlick, J. Zimmer, Calculation of long time classical trajectories: algorithmic treatment and applications for molecular systems, *J. Phys. Chem.* 130 (2009) 124106.
- [35] H. Schwetlick, J. Zimmer, A convergent string method: existence and approximation for the Hamiltonian boundary-value problem, Pre-print, <http://www.md-net.org.uk/preprints/jacobi7.pdk>.
- [36] W.E.R. Weiqing, E. Vanden-Eijnden, Minimum action method for the study of rare events, *Commun. Pure Appl. Math.* 57 (2004) 0001–0020.
- [37] W.H. Fleming, *Deterministic and Stochastic Optimal Control*, Springer-Verlag, 1975.
- [38] M.I. Dykman, Large fluctuations and fluctuational transitions in systems driven by colored gaussian-noise: a high-frequency noise, *Phys. Rev. A* 42 (1990) 2020–2029.
- [39] E. Forgoston, I.B. Schwartz, Escape rates in a stochastic environment with multiple scales, *SIAM J. Appl. Dyn. Syst.* 8 (2009) 1190–1217.
- [40] A. Kessler, L.B. Shaw, I.B. Schwartz, On the construction of optimal paths to extinction, U.S. naval Research Laboratory Report No. 6790-12-9374, 2012.
- [41] G. Haller, Finding finite-time invariant manifolds in two-dimensional velocity fields, *Chaos* 10 (2000) 99–108.
- [42] G. Haller, Distinguished material surfaces and coherent structures in three-dimensional fluid flows, *Physica D* 149 (2001) 248–277.
- [43] G. Haller, Lagrangian coherent structures from approximate velocity data, *Phys. Fluids* 14 (2002) 1851–1861.
- [44] S.C. Shadden, F. Lekien, J.E. Marsden, Definition and properties of Lagrangian coherent structures from finite-time Lyapunov exponents in two-dimensional aperiodic flows, *Physica D* 212 (2005) 271–304.
- [45] I.B. Schwartz, T. Carr, L. Billings, M. Dykman, Noise induced switching in delayed systems, 2012. [arXiv:1207.7278v1](https://arxiv.org/abs/1207.7278v1).
- [46] J.M. Ortega, W.C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, SIAM, Philadelphia, 2000.