

LARGE DEVIATIONS FOR GAUSSIAN DIFFUSIONS WITH DELAY

ROBERT AZENCOTT, BRETT GEIGER AND WILLIAM OTT

ABSTRACT. Dynamical systems driven by nonlinear delay SDEs with small noise can exhibit important rare events on long timescales. When there is no delay, classical large deviations theory quantifies rare events such as escapes from nominally stable fixed points. Near such fixed points one can approximate nonlinear delay SDEs by linear delay SDEs. Here, we develop a fully explicit large deviations framework for (necessarily Gaussian) processes X_t driven by linear delay SDEs with small diffusion coefficients. Our approach enables fast numerical computation of the action functional controlling rare events for X_t and of the most likely paths transiting from $X_0 = p$ to $X_T = q$. Via linear noise local approximations, we can then compute most likely routes of escape from metastable states for nonlinear delay SDEs. We apply our methodology to the detailed dynamics of a genetic regulatory circuit, namely the co-repressive toggle switch, which may be described by a nonlinear chemical Langevin SDE with delay.

1. INTRODUCTION

Dynamical processes are often influenced by small random fluctuations acting on a variety of spatiotemporal scales. Small noise can dramatically affect the underlying deterministic dynamics by transforming stable states into metastable states and giving positive probability to “rare events” of high interest, such as excursions away from nominally stable states or transitions between metastable states. These rare events play important functional roles in a wide range of applied settings, including genetic circuits [15], molecular dynamics, turbulent flows [8] and other systems with multiple timescales [7].

The main goal of this paper is to present an *explicit* computational and theoretical large deviations analysis of rare events for Gaussian diffusion processes with *delays*. We are motivated in part by the importance of delay for the dynamics of genetic regulatory circuits. Indeed, we apply our approach to a *bistable genetic switch* driven by a delay stochastic differential equation (delay SDE) of Langevin type.

Consider a family of random processes $X^\varepsilon(t)$ indexed by a small parameter $\varepsilon > 0$ and driven by the following generic small-noise SDE with drift b , diffusion $\sqrt{\varepsilon}\sigma$, and *no delays*:

$$dX^\varepsilon(t) = b(X^\varepsilon(t)) dt + \sqrt{\varepsilon}\sigma(X^\varepsilon(t)) dW(t).$$

Large deviations theory for SDEs of this form was developed by Freidlin and Wentzell [16]. Freidlin-Wentzell theory estimates the probability that the process $X^\varepsilon(t)$ lies within a small tube around any given continuous path $\psi \in C([0, T], \mathbb{R}^d)$ in terms of the *action* $S_T(\psi)$ of ψ :

$$\mathbb{P}_x \left\{ \sup_{0 \leq t \leq T} |X^\varepsilon(t) - \psi(t)| \leq \delta \right\} \approx \exp \left(-\varepsilon^{-1} S_T(\psi) \right).$$

Here \mathbb{P}_x denotes probability conditioned on $X^\varepsilon(0) = x$ and we assume that $\psi(0) = x$.

The Freidlin-Wentzell action functional $0 \leq S_T(\psi) \leq \infty$ was originally defined for uniformly bounded coefficients b, σ and uniformly elliptic $\sigma\sigma^*$ by an explicit time integral involving $b(\psi_t)$,

UNIVERSITY OF HOUSTON

Date: September 29, 2016.

2010 *Mathematics Subject Classification*. 60F10, 60G15, 60H10.

Key words and phrases. Gaussian process, diffusion, delay, optimal transition path, chemical Langevin equation, linear noise approximation, co-repressive toggle switch.

$\sigma\sigma^*(\psi_t)^{-1}$, and ψ'_t . These remarkable results were widely extended by S. Varadhan [2] to arbitrary sets of trajectories and by R. Azencott [2] to hypoelliptic diffusions with unbounded coefficients. Numerous extensions and applications to broad classes of stochastic processes have been published by D. Strook, R. Ellis, A. Dembo, O. Zeitouni, G. Dupuis, and many others. For SDEs with delays, large deviations principles have been established or reasonably justified under a variety of hypotheses [5, 11, 17, 26, 35, 36, 39].

For fixed time T and points p, q in the state space, the path $\hat{\psi}$ that minimizes $S_T(\psi)$ (under the constraints $\psi(0) = p$ and $\psi(T) = q$) is the most likely transition path starting at p and reaching q at time T . A second minimization over T provides the most likely transition path from p to q and the energy $V(p, q)$ associated with this optimal path. Often called the quasi-potential, V is central to the quantification of large deviations on long timescales [16].

A computational framework has been developed for the application of Freidlin-Wentzell theory. This framework includes the minimum action method [14], an extension called the geometric minimum action method that synthesizes the minimum action method and the string method [22], as well as variants of these approaches (see *e.g.* [27, 28]).

For nonlinear delay SDEs, it is possible to compute a linear noise approximation [9] that is valid in a neighborhood of a given metastable state. Since linear noise approximations are Gaussian diffusions with delays, we have deliberately focused the present paper on Gaussian diffusions with delays. For such diffusions, we rigorously develop and implement a *fully explicit* large deviation framework enabling fast numerical computation of optimal transition paths and the quasi-potential. Our methodology does not require the numerical solution of Hamilton-Jacobi equations, a significant positive since Hamilton-Jacobi equations are computationally costly in even moderately high spatial dimension.

We thus center our study on the Itô delay SDE

$$(1) \quad \begin{cases} dX_t^\varepsilon = (a + BX_t^\varepsilon + CX_{t-\tau}^\varepsilon) dt + \varepsilon \Sigma dW_t, \\ X_t^\varepsilon = \gamma(t) \text{ for } t \in [-\tau, 0]. \end{cases}$$

Here $X_t^\varepsilon \in \mathbb{R}^d$, t denotes time, $\tau \geq 0$ is the delay, $a \in \mathbb{R}^d$, B and C are real $d \times d$ matrices, W_t denotes standard n -dimensional Brownian motion, $\Sigma \in \mathbb{R}^{d \times n}$ denotes the diffusion matrix, and $\varepsilon > 0$ is a small noise parameter. The initial history of the process is given by the Lipschitz continuous curve $\gamma : [-\tau, 0] \rightarrow \mathbb{R}^d$. We work with fixed delay to simplify the presentation – all of our results apply just as well to multiple delays and to random delay distributed over a finite time interval.

The Gaussian diffusion (1) arises via linear noise approximation of nonlinear delay SDEs near metastable states in the following way. Suppose the nonlinear delay SDE

$$dx_t = f(x(t), x(t - \tau)) dt + \varepsilon g(x(t), x(t - \tau)) dW_t$$

has a metastable state z ; that is, z is a stable fixed point of the deterministic limit ODE

$$dx_t = f(x(t), x(t - \tau)) dt.$$

Writing $x(t) = z + \xi(t)$ and expanding f and g around z yields the linear noise approximation

$$d\xi_t = [D_1 f(z, z)\xi(t) + D_2 f(z, z)\xi(t - \tau)] dt + \varepsilon g(z, z) dW_t.$$

This is (1) with $a = \mathbf{0}$, $B = D_1 f(z, z)$, $C = D_2 f(z, z)$, and $\Sigma = g(z, z)$.

We demonstrate the utility of our approach by computing optimal escape trajectories for the co-repressive toggle switch, a bistable genetic circuit driven by a nonlinear delay Langevin equation.

The paper is organized as follows. In Section 2, we review the theory of large deviations for Gaussian processes and present optimal transition path theory for (1). We detail our numerical implementation of this theory in Section 3. Section 4 discusses the general idea of linear noise approximation. We present our computational study of a bistable genetic toggle switch in Section 5.

2. THEORY

In this section we develop a rigorous large deviations framework for (1).

2.1. Outline. For brevity, we will often omit the superscript ε , writing X_t instead of X_t^ε . We first show that the process X_t driven by (1) is in fact a Gaussian process in Section 2.2. This is expected since (1) is linear, but not immediately obvious because of the presence of delay. Since X_t is a Gaussian process, it is completely determined by its mean $m(t) = \mathbb{E}[X_t]$ and covariance matrices $\mathbb{E}[X_s X_t^*] - m(s)m(t)^*$. Here $*$ denotes matrix transpose. We derive and analytically solve delay ODEs verified by the mean and lagged covariance matrix of X_t in Sections 2.3–2.7.

To develop the large deviations theory for Gaussian diffusions with delay, we first summarize the general theory of large deviations for Gaussian measures in Section 2.8. The main known result (Theorem 2.3) is a large deviations principle for centered Gaussian measures (defined in Section 2.8) on Hilbert spaces. We will then use this result to identify the action functional (Definition 2.2) for Gaussian diffusions with delay.

For Gaussian diffusions with delay, the action functional is linked to the Cramer transform of Gaussian probability measures on path spaces. Since the large deviations principle of interest here concerns centered Gaussian measures, we center X_t by writing $X_t = m(t) + \varepsilon Z_t$. Call ν the probability distribution of the process Z_t on the space $C_{\mathbf{0}}([0, T])$ of continuous paths f starting at $f(0) = \mathbf{0}$. The Cramer transform $\lambda(f)$ of ν is defined for all such paths f by Definition 2.2. Proposition 2.5 expresses $\lambda(f)$ in terms of the integral operator determined by the covariance function $\rho(s, t)$ of Z_t . This expression will be suitably transformed to derive explicit computational schemes.

To complete this program, we will explicitly derive the most likely transition path between two points p and q by first minimizing the Cramer transform over paths that begin at p and reach q at time T and then minimizing over T .

2.2. Solution of (1) is a Gaussian process. We show that (1) defines a Gaussian process by first discretizing and then taking an L^2 -limit.

Proposition 2.1. *The delay SDE (1) has a unique strong solution X_t , which is a Gaussian process.*

Proof of Proposition 2.1. The existence of a unique strong solution X_t is classical (see e.g. [32]). To see that X_t is Gaussian, we consider Euler-Maruyama discretizations [23]. For positive integers N , let $\Delta = \tau/N$ denote time step size. The Euler-Maruyama approximate solution $Y_t^{(\Delta)}$ to (1) is defined first at nonnegative integer multiples of Δ by

$$Y_{(k+1)\Delta}^{(\Delta)} = Y_{k\Delta}^{(\Delta)} + (a + BY_{k\Delta}^{(\Delta)} + CY_{k\Delta-\tau}^{(\Delta)})\Delta + \varepsilon\Sigma(W_{(k+1)\Delta} - W_{k\Delta})$$

and then on $[0, T]$ by interpolation. Much is known about the convergence of EM schemes for dSDEs (see e.g. [3, 33]). In particular, we have

$$\lim_{\Delta \rightarrow 0} \mathbb{E} \left[\sup_{0 \leq t \leq T} |Y_t^{(\Delta)} - X_t|^2 \right] = 0$$

by Theorem 2.1 of [33]. Since $Y_t^{(\Delta)}$ is Gaussian by construction, this L^2 -convergence implies that X_t is Gaussian as well. \blacksquare

2.3. Delay ODE for the mean of X_t . Writing (1) in integral form, we have

$$(2) \quad X_t = X_0 + \int_0^t (a + BX_z + CX_{z-\tau}) dz + \varepsilon\Sigma W_t.$$

Taking the expectation of (2) and applying Fubini gives

$$m(t) = m(0) + \int_0^t (a + Bm(z) + Cm(z - \tau)) dz,$$

or, in differential form, a delay ODE for $m(t)$:

$$(3) \quad \begin{cases} m'(t) = a + Bm(t) + Cm(t - \tau) \\ m(t) = \gamma(t) \text{ for } t \in [-\tau, 0]. \end{cases}$$

2.4. The centered Gaussian process Z_t . The process X_t is clearly not centered in general. The centered process Z_t defined by $X_t = m(t) + \varepsilon Z_t$ is a centered Gaussian diffusion with delay. Since X_t verifies (1) and $m(t)$ verifies (3), elementary algebra shows that Z_t verifies the delay SDE

$$(4) \quad \begin{cases} dZ_t = (BZ_t + CZ_{t-\tau}) dt + \Sigma dW_t, \\ Z_t = 0 \text{ for } t \in [-\tau, 0] \end{cases}$$

Note that this delay ODE *does not depend on* ε . This is a crucial point further on because our key large deviations estimates will be stated in path space for the centered Gaussian process εZ_t . So for the rest of the paper our large deviations computations will essentially involve the deterministic mean path of X_t and the covariance function $\rho(s, t)$ of the process Z_t .

2.5. Delay ODEs for the covariances of Z_t . We now find delay ODEs for the covariance of Z_t . Denote A^* the matrix transpose of A , and let

$$\rho(s, t) = \mathbb{E}[Z_s Z_t^*]$$

be the covariance matrix of Z_s and Z_t . Since the history of Z_t anterior to $t = 0$ is deterministic, $\rho(s, t) = 0$ when either s or t are in $[-\tau, 0]$. Fix $t \in [0, T]$, and let s vary. We have

$$\mathbb{E}[Z_s Z_t^*] = \int_0^s (B\mathbb{E}[Z_u Z_t^*] + C\mathbb{E}[Z_{u-\tau} Z_t^*]) du + \Sigma \mathbb{E}[W_s Z_t^*].$$

We thus obtain

$$(5) \quad \rho(s, t) = \int_0^s (B\rho(u, t) + C\rho(u - \tau, t)) du + \Sigma \mathbb{E}[W_s Z_t^*].$$

Let $G(s, t) = \mathbb{E}[W_s Z_t^*]$. Differentiating $\rho(s, t)$ with respect to s gives

$$(6) \quad \frac{\partial \rho}{\partial s}(s, t) = B\rho(s, t) + C\rho(s - \tau, t) + \Sigma \frac{\partial G}{\partial s}(s, t),$$

which is a first-order delay ODE in s for each fixed t . To close (6), we compute a differential equation for $\frac{\partial G}{\partial s}(s, t)$. Proceeding as just done for (5), one checks that the function $G(s, t)$ satisfies the ODE

$$(7) \quad \frac{\partial G}{\partial t}(s, t) = \begin{cases} G(s, t)B^* + G(s, t - \tau)C^* + \Sigma^* & (t \leq s); \\ G(s, t)B^* + G(s, t - \tau)C^* & (t > s), \end{cases}$$

where $G(s, t) = 0$ for $t \in [-\tau, 0]$. Let $H(x)$ denote the Heaviside function

$$H(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}.$$

We can rewrite (7) as

$$(8) \quad \frac{\partial G}{\partial t}(s, t) = G(s, t)B^* + G(s, t - \tau)C^* + \Sigma^* H(s - t).$$

Note that the partial derivative of the Heaviside distribution $H(s - t)$ is classically given by

$$\frac{\partial H}{\partial s}(s - t) = \delta(s - t),$$

where the distribution $x \rightarrow \delta(x)$ is the Dirac point mass concentrated at $x = 0$. By definition of $G(s, t)$ and by (8), the function $G(s, t)$ is continuous for all s and t and differentiable in s and t for $s \neq t$. We will denote, for $s \neq t$,

$$(9) \quad F(s, t) := \frac{\partial G}{\partial s}(s, t)$$

so that F verifies the initial condition $F(s, t) = 0$ for $s \neq t$ and $t \in [-\tau, 0]$.

Differentiating (8) in s for $s \neq t$ and switching the order of partial derivatives yields a linear delay ODE in $t > 0$ for $F(s, t)$, namely

$$(10) \quad \frac{\partial F}{\partial t}(s, t) = F(s, t)B^* + F(s, t - \tau)C^* + \Sigma^* \delta(s - t)$$

with initial condition $F(s, t) = 0$ for all $t \in [-\tau, 0]$.

Once $F(s, t)$ is determined, the covariance $\rho(s, t)$ for each fixed $t \in [0, T]$ will be computed by solving the delay ODE

$$(11) \quad \frac{\partial \rho}{\partial s}(s, t) = B\rho(s, t) + C\rho(s - \tau, t) + \Sigma F(s, t).$$

We now describe how to successively solve the delay ODEs driving $m(t)$, $F(s, t)$, and $\rho(s, t)$.

2.6. Analytical solution of the delay ODE verified by the mean. First-order delay ODEs can be analytically solved by a natural stepwise approach, sometimes called “method of steps,” a terminology which we will avoid since it has a different meaning in classical numerical analysis. The basic idea is to convert each one of our delay ODE into a finite sequence of nonhomogeneous ODEs in which the delay terms successively become known terms. Consider first the delay ODE (3) for $m(t)$ with $t \in [-\tau, T]$. The delay term $Cm(t - \tau)$ is unknown for $t \in (\tau, T]$ but is known for $t \in [0, \tau]$. So we can solve the delay ODE (analytically or numerically) on the interval $[0, \tau]$ as a linear non-homogeneous first-order ODE. Then, for $t \in [\tau, 2\tau]$, the delay ODE turns again into a linear non-homogeneous first-order ODE where the delay term has actually just been computed. One can thus successively solve the delay ODE on intervals $[k\tau, (k + 1)\tau]$ to get a full step-by-step solution on all of $[0, T]$.

We first describe the explicit solution of the mean $m(t)$ on $[-\tau, T]$. For $x \in \mathbb{R}$, denote $[x]$ as the greatest integer less than or equal to x . Partition the interval $[-\tau, T]$ into closed subintervals of the form $[(k - 1)\tau, k\tau]$ where $k = 0, 1, 2, \dots, \lfloor \frac{T}{\tau} \rfloor$ with final partition interval $[\lfloor \frac{T}{\tau} \rfloor \tau, T]$. Let $m_k(t)$ denote the solution to the DDE on the interval $[(k - 1)\tau, k\tau]$ and $m_{\lfloor \frac{T}{\tau} \rfloor + 1}(t)$ denote the solution of $m(t)$ on the final partition interval. When $k = 0$, we have $m_0(t) = \gamma(t)$. Now, when $k = 1$, the initial condition gives the following ODE for $m_1(t)$, the solution of $m(t)$ on $[0, \tau]$:

$$\begin{cases} m_1'(t) = a + Bm_1(t) + C\gamma(t - \tau) \\ m_1(0) = \gamma(0). \end{cases}$$

We can write the solution $m_1(t)$ as

$$m_1(t) = e^{tB} \int_0^t e^{-uB} (a + C\gamma(u - \tau)) du + e^{tB} \gamma(0).$$

Call $m_k(t)$ the solution $m(t)$ on the interval $J_k = [(k - 1)\tau, k\tau]$ for $k = 0, 1, 2, \dots, \lfloor \frac{T}{\tau} \rfloor$. Given $m_{k-1}(t)$ for t in J_{k-1} , we can similarly compute $m_k(t)$ by

$$m_k(t) = e^{(t-(k-1)\tau)B} \int_{(k-1)\tau}^t e^{-uB} (a + Cm_{k-1}(u - \tau)) du + e^{(t-(k-1)\tau)B} m_{k-1}((k-1)\tau).$$

Finally, piecing together all the $m_k(t)$ yields the full solution $m(t)$ on all of $[-\tau, T]$. Note that many characteristics of the initial segment γ , such as continuity, differentiability, discontinuities, etc., will essentially propagate through to the solution $m(t)$ at each step. More precisely, if γ is

of class C^q on $[-\tau, 0]$ for some integer $q \geq 0$, then $m(t)$ will be of class $q + 1$ for all t except possibly at integer multiples of τ . Since we assume here that γ is Lipschitz continuous, $m(t)$ will be differentiable except possibly at integer multiples of τ .

2.7. Analytical solutions of the delay ODEs verified by $F(s, t)$ and $\rho(s, t)$. We can extend the preceding method to the ODE in s verified by $F(s, t)$ for each fixed s and then to the ODE in t verified by $\rho(s, t)$. We first focus on $F(s, t)$. Fix $s \in [0, T]$. Due to the delay ODE (10), the distribution ϕ_s defined on R^+ by

$$\phi_s(t) = F(s, t) + \Sigma^* H(s - t)$$

clearly verifies the delay ODE

$$(12) \quad \frac{\partial \phi_s}{\partial t}(t) - \phi_s(t)B^* - \phi_s(t - \tau)C^* = -\Sigma^* H(s - t)B^* - \Sigma^* H(s - t + \tau)C^*$$

with initial condition

$$\phi_s(t) = \Sigma^* H(s - t) \text{ for all } t \in [-\tau, 0].$$

Note that for each fixed $s > 0$ this initial condition is a bounded and continuous function of $t \in [-\tau, 0]$. For each fixed s , the right hand side of equation (12) is the function θ_s defined for $t \geq 0$ by

$$\theta_s(t) = -\Sigma^* H(s - t)B^* - \Sigma^* H(s - t + \tau)C^*$$

which is uniformly bounded in t , and is continuous in t except for the two points $t = s$ and $t = s + \tau$. As was done above for $m(t)$, one can perform the iterative analysis of the delay ODE (12) on successive time intervals $J_k = [(k - 1)\tau, k\tau]$. Since both the initial condition and the right-hand side θ_s are known, the k^{th} step of this iterative construction amounts to solving a first order linear ODE with constant coefficients and known right-hand side. So this construction is essentially stepwise explicit and proves by recurrence on k that the distribution $\phi_s(t)$ is actually a bounded function of t which is differentiable except maybe at the points $t = k\tau$ and $t = s + k\tau$.

For each $s \geq 0$, once the full solution ϕ_s has been constructed for $t \in [-\tau, T]$ as just outlined, we immediately obtain $F(s, t) = \phi_s(t) - \Sigma^* H(s - t)$.

At this stage $F(s, t)$ is theoretically known for all $s \geq 0$ and $t \in [-\tau, T]$ and can be plugged into the delay ODE in s verified by $\rho(s, t)$ for each fixed t , with initial conditions $\rho(s, t) = 0$ for $(s, t) \in [-\tau, 0] \times [-\tau, 0]$. For each fixed $t \in [0, T]$, this delay ODE for $s \rightarrow \rho(s, t)$ can again be solved iteratively on the successive time intervals J_k .

The preceding approaches can easily be numerically implemented to derive explicit solution to the three types of delay ODEs involved. Each reduction to a succession of roughly T/τ linear ODEs enables the use of classical numerical schemes to compute $m(t)$ and $\rho(s, t)$. We have used the approach of [6], which implements the step-wise analysis presented above, along with standard numerical ODE methods. This numerical implementation is described explicitly in Section 3.

The key role played below by $m(t)$ and $\rho(s, t)$ is that these two functions essentially determine the rate functional of large deviations theory for the Gaussian diffusion with delay X_t .

2.8. General large deviations framework. We present, without proof, a brief overview of large deviations theory of Gaussian measures and processes (refer to Chapter 6 in [2] for proofs of theorems). We will then apply these principles to Gaussian diffusions with delay. The following notations and definitions will be used throughout this section.

- H is any separable Hilbert space, with scalar product denoted $\langle t, x \rangle := t(x)$ for $t, x \in H$.
- μ is any probability on the Borel σ -algebra $\mathcal{B}(H)$.
- For $t \in H$, the image probability $t(\mu)$ is defined on \mathbb{R} by $[t(\mu)](A) := \mu(t^{-1}(A))$ for all Borel subsets A of \mathbb{R} .
- μ is called *centered* iff $t(\mu)$ is centered for all $t \in H$.

- μ is called *Gaussian* iff for all $t \in H$, the image probability $t(\mu)$ is a Gaussian distribution on \mathbb{R} .

Large deviations concepts demand all $t(\mu)$ to have at least finite first-order moment. But most applicable results require all $t(\mu)$ to have some finite exponential moments since they depend on the Laplace transform $\hat{\mu}(t)$ of μ , defined as follows for $t \in H$,

$$\hat{\mu}(t) = \int_H e^{\langle t, x \rangle} d\mu(x).$$

For a full treatment of large deviations concepts for probabilities on general infinite dimensional Frechet vector spaces, refer to Chapters 2 and 3 of [2]. Here we only consider Borel probabilities μ on separable Hilbert spaces H . Later on below, H will be an L^2 -space of process paths and μ will be Gaussian. Probabilities of rare events under μ can be estimated via a key non-negative functional defined for $x \in H$: the Cramer transform $\lambda(x)$ of μ . The following definition of $\lambda(x)$ is actually Theorem 3.2.1 in [2].

Definition 2.2. The *Cramer transform* λ of μ , also called the large deviations *rate functional* of μ , is defined for $x \in H$ by

$$\lambda(x) = \sup_{t \in H} [\langle t, x \rangle - \log \hat{\mu}(t)].$$

Note that $0 \leq \lambda(x) \leq +\infty$. The *Cramer set functional* $\Lambda(A)$ is then defined for all $A \subseteq H$ by

$$\Lambda(A) = \inf_{x \in A} \lambda(x).$$

The set functional $\Lambda(A)$ quantifies the probabilities of rare events by the following key large deviations inequalities initially formalized by S. Varadhan.

Theorem 2.3. *Let μ be a probability measure on a separable Hilbert space H . Let Z be an H -valued random variable with probability distribution μ . Let Λ be the Cramer set functional of μ . For every Borel subset A of H one has*

$$(13) \quad -\Lambda(A^\circ) \leq \liminf_{\varepsilon \rightarrow 0} \varepsilon^2 \log \mathbb{P}(\varepsilon Z \in A) \leq \limsup_{\varepsilon \rightarrow 0} \varepsilon^2 \log \mathbb{P}(\varepsilon Z \in A) \leq -\Lambda(\bar{A}).$$

where A° and \bar{A} are resp. the interior and the closure of A .

Whenever $\Lambda(A^\circ) = \Lambda(\bar{A})$, which is necessarily the case when \bar{A} is the closure of A° , then the limits in (13) exist and

$$-\Lambda(A) = \lim_{\varepsilon \rightarrow 0} \varepsilon^2 \log \mathbb{P}(\varepsilon Z \in A).$$

In particular for small ε , one has the rough estimate by

$$\log \mathbb{P}(\varepsilon Z \in A) \approx -\frac{\Lambda(A)}{\varepsilon^2}.$$

In our applications below, Z is the random path of a centered Gaussian diffusion with delay so that the probability distribution ν of Z will be a centered Gaussian probability on the Hilbert space $H = L^2[0, T]$. So we now focus on Gaussian probabilities on Hilbert spaces.

2.9. Gaussian probabilities on Hilbert spaces. Let H be a separable Hilbert space, and let ν be a centered Gaussian probability on the Borel subsets of H . With no loss of generality, we assume that the only closed vector subspace F of H such that $\nu(F) = 1$ is H itself. The *covariance kernel* $\text{Cov}(s, t)$ of ν is defined for all $s, t \in H$ by

$$\text{Cov}(s, t) = \int_H \langle s, x \rangle \langle t, x \rangle d\nu(x) = \langle s, \Gamma t \rangle = \langle \Gamma s, t \rangle,$$

where the linear operator $\Gamma : H \rightarrow H$ is known to be a bounded, positive, self-adjoint operator with finite trace. The positive operator $\sqrt{\Gamma}$ then exists and is unique. The following theorem gives

a fairly concrete form for $\lambda(x)$. This result can be applied in path space to centered Gaussian processes, once the covariance operator Γ has been computed.

Theorem 2.4. *Let ν be a centered Gaussian probability on a separable Hilbert space H . Let $\Gamma : H \rightarrow H$ be the self-adjoint covariance operator of ν . Let U be the orthogonal complement in H of the null space $\ker \Gamma = \ker \sqrt{\Gamma}$. The restriction S of $\sqrt{\Gamma}$ to U is injective and maps U onto $\sqrt{\Gamma}(H)$. Then the Cramer transform λ of ν is given by*

$$\lambda(x) = \begin{cases} \frac{1}{2} \|S^{-1}x\|^2 & x \in \sqrt{\Gamma}(H) \\ \infty & \text{otherwise} \end{cases}.$$

2.10. Application to Gaussian processes. Let $Z_s : \Omega \rightarrow \mathbb{R}$ with $s \in [0, T]$ be a centered Gaussian stochastic process with almost surely continuous trajectories and continuous covariance function

$$\rho(s, t) = \int_{\Omega} Z_s(\omega) Z_t(\omega) d\mathbb{P}(\omega).$$

Call $C([0, T])$ the Banach space of continuous functions on $[0, T]$ endowed with its Borel σ -algebra. One can trivially construct a version of Z_s with surely continuous trajectories. This defines a $C([0, T])$ -valued random path Z , where $Z(\omega)$ is the path $s \rightarrow Z_s(\omega)$ with $0 \leq s \leq T$. The probability distribution ν of the random path Z is then a centered Gaussian probability on the Borel sets of $C([0, T])$. We now state the main large deviations result used below, which is essentially an application of Theorem 2.4 to the separable Hilbert space $L^2[0, T]$ and appears as Proposition 6.3.7 in [2].

Proposition 2.5. *Consider a centered continuous Gaussian process $Z_s : \Omega \rightarrow \mathbb{R}$ defined for s in $[0, T]$ with continuous covariance function $\rho(s, t)$. The linear operator $R : L^2[0, T] \rightarrow L^2[0, T]$ defined by*

$$Rf(s) = \int_0^T \rho(s, u) f(u) du$$

takes values in $C([0, T])$. Moreover, R is self-adjoint, positive, compact, and has finite trace. Let $U = (\ker(R))^\perp \subset L^2([0, T])$ and let S be the restriction of \sqrt{R} to U . Then $S : U \rightarrow L^2[0, T]$ is injective and maps U onto $\sqrt{R}(L^2[0, T])$. On the path space $C[0, T]$, the probability distribution ν induced by the process Z_t has Cramer transform λ defined for $f \in C[0, T]$ by

$$\lambda(f) = \begin{cases} \frac{1}{2} \|S^{-1}f\|_{L^2[0, T]}^2, & \text{if } f \in \sqrt{R}(L^2[0, T]); \\ \infty, & \text{otherwise.} \end{cases}$$

Note that by duality, R also acts on the space of all bounded Radon measures π on $[0, T]$, via the natural formula

$$R\pi(s) = \int_0^T \rho(s, u) d\pi(u).$$

Since the integral operator R is positive and self-adjoint, the square-root operator \sqrt{R} exists and is also an integral operator of the form

$$\sqrt{R}f(s) = \int_0^T k(s, u) f(u) du$$

where $k(s, u)$ is uniquely defined by the relation $\int_0^T k(s, u) k(u, v) du = \rho(s, v)$ for all $s, v \in [0, T]$.

The value $\lambda(f)$ of the Cramer transform can be viewed as the ‘‘energy’’ of the path f . In particular, for a small multiple εW_t of the Brownian motion W_t , the Cramer transform is indeed the kinetic energy $\lambda(f) = \frac{1}{2} \|f'\|_{L^2[0, T]}^2$ (Proposition 6.3.8 in [2]). Further on, we will apply Proposition 2.5 to the centered Gaussian process $\varepsilon Z_t = X_t - m_t$ associated to the Gaussian diffusion with delay X_t . Indeed, since the mean trajectory m_t is deterministic, probability estimates for the random paths of Z_t immediately translate into probability estimates for the random paths of X_t .

2.11. Large deviation rate functional for Gaussian diffusions with delay. Recall that $X_t = m(t) + \varepsilon Z_t$ denotes Gaussian diffusion with delay under study. In this section we minimize the Cramer transform associated with Z_t . For points $p, q \in \mathbb{R}^d$ and time $T > 0$, define

$$\text{Path}(p, q; T) = \left\{ f \in C([0, T], \mathbb{R}^d) : f(0) = p, f(T) = q \right\}.$$

Random paths of X_t lie in $\text{Path}(p, q; T)$. To study Z_t , we must shift this space of paths by the mean $m(t)$. Define

$$\begin{aligned} C_{\mathbf{0}}([0, T], \mathbb{R}^d) &= \left\{ f \in C([0, T], \mathbb{R}^d) : f(0) = \mathbf{0} \right\}, \\ \text{Path}(\mathbf{0}, q - m(T); T) &= \left\{ f \in C_{\mathbf{0}}([0, T], \mathbb{R}^d) : f(T) = q - m(T) \right\}. \end{aligned}$$

We now minimize the Cramer transform λ associated with Z_t over $\text{Path}(\mathbf{0}, q - m(T); T)$. The Cramer transform is linked to the covariance operator R of Z_t by

$$\lambda(f) = \frac{1}{2} \langle R^{-1} f, f \rangle_{L^2[0, T]}$$

for paths $f \in \sqrt{R}(L^2[0, T])$. Since $\text{Path}(\mathbf{0}, q - m(T); T)$ is determined by linear constraints on f , we minimize the quadratic form $\lambda(f)$ for $f \in \text{Path}(\mathbf{0}, q - m(T); T)$ using Lagrange multiplier theory [29].

For $f \in C_{\mathbf{0}}([0, T], \mathbb{R}^d)$, define the Lagrangian

$$\mathcal{L}_{f, \mu} := \frac{1}{2} \langle R^{-1} f, f \rangle + \mu \cdot (f(T) - (q - m(T))),$$

where \cdot is the usual dot product in \mathbb{R}^d and $\mu \in \mathbb{R}^d$ is the Lagrange multiplier vector. Setting the derivative $D\mathcal{L}_{f, \mu}(\varphi)$ for $\varphi \in C_{\mathbf{0}}([0, T], \mathbb{R}^d)$ equal to zero, we have

$$D\mathcal{L}_{f, \mu}(\varphi) = \lim_{\Delta \rightarrow 0} \frac{\mathcal{L}_{f+\Delta\varphi, \mu} - \mathcal{L}_{f, \mu}}{\Delta} = \langle R^{-1} f, \varphi \rangle + \mu \cdot \varphi(T) = 0,$$

yielding the condition

$$\langle R^{-1} f, \varphi \rangle = -\mu \cdot \varphi(T) = -\mu \cdot \delta_T(\varphi).$$

Here δ_T denotes the Dirac mass at time T . The minimizing path g is therefore given by

$$(14) \quad g = R(-\mu \cdot \delta_T).$$

The right side of (14) may be expressed in terms of the covariance function ρ of Z_t :

$$(15) \quad R(-\mu \cdot \delta_T)(s) = - \int_0^T \rho(s, u) \mu \, d\delta_T(u) = -\rho(s, T)\mu.$$

Since $g(T) = q - m(T)$, (15) implies

$$q - m(T) = -\rho(T, T)\mu,$$

so the Lagrange multiplier is given by

$$-\mu = \rho(T, T)^{-1}(q - m(T)).$$

The trajectory that minimizes the Cramer transform therefore takes the form

$$g^T(s) = \rho(s, T)[\rho(T, T)^{-1}(q - m(T))] \quad (0 \leq s \leq T),$$

and has Cramer transform

$$(16) \quad \lambda(g^T) = \frac{1}{2} [\rho(T, T)^{-1}(q - m(T))] \cdot [q - m(T)].$$

Notice that $g^T = \sqrt{R}(\hat{g})$, where

$$\hat{g}(s) = k(s, T)[\rho(T, T)^{-1}(q - m(T))]$$

and k is the kernel defined by

$$(\sqrt{R}f)(s) = \int_0^T k(s, u)f(u) du.$$

For the Gaussian diffusion with delay $X_t = m(t) + \varepsilon Z_t$, the most likely path $h^T(s)$ realizing $X_0 = p$ and $X_T = q$ is hence given by

$$(17) \quad h^T(s) = m(s) + g^T(s) = m(s) + \rho(s, T)[\rho(T, T)^{-1}(q - m(T))] \quad (0 \leq s \leq T)$$

and has energy

$$(18) \quad \lambda(h^T) = \frac{1}{2}[\rho(T, T)^{-1}(q - m(T))] \cdot [q - m(T)].$$

Below we interpret *most likely path* in a precise probabilistic sense. Minimizing $\lambda(h^T)$ over T produces the most likely time T of transition from $X_0 = p$ to $X_T = q$, the most likely transition path, and the associated energy.

2.12. Probabilistic interpretation. For any path from p to q over time interval $[0, T]$ in \mathbb{R}^d , the large deviations principle in Theorem 2.3 yields quantitative information about the probability that the process X_t remains within a small tube of the given path over $[0, T]$. To apply Theorem 2.3, we first shift the given tube by subtracting the mean $m(t)$ of X_t so that we may work with the centered process εZ_t . Once the large deviations principle has been applied to εZ_t , we then add $m(t)$ to recover information about X_t . Crucially, both the Cramer transform associated with Z_t and the path that minimizes it are independent of ε . The application of Theorem 2.3 proceeds as follows.

First assume $q \neq m(T)$. For any path $f \in \text{Path}(\mathbf{0}, q - m(T); T)$ and small radius $r > 0$, define the tube

$$\text{Tube}(f, r) = \left\{ \varphi \in C_0([0, T], \mathbb{R}^d) : \sup_{0 \leq t \leq T} |\varphi(t) - f(t)| \leq r \right\}.$$

Since $\text{Tube}(f, r)$ is the closure of its interior in $C_0([0, T], \mathbb{R}^d)$, Theorem 2.3 gives

$$\lim_{\varepsilon \rightarrow 0} \varepsilon^2 \log(\mathbb{P}(\varepsilon Z^T \in \text{Tube}(f, r))) = -\Lambda(\text{Tube}(f, r)),$$

where Z^T denotes the set of paths generated by Z_t over $[0, T]$. In particular, for the path g^T that minimizes the Cramer transform, we have

$$\lim_{\varepsilon \rightarrow 0} \varepsilon^2 \log(\mathbb{P}(\varepsilon Z^T \in \text{Tube}(g^T, r))) = -\Lambda(\text{Tube}(g^T, r)) = -\lambda(g^T).$$

In this sense, $\text{Tube}(g^T, r)$ is the most likely route of passage for εZ_t , and shifting $\text{Tube}(g^T, r)$ by $m(t)$ yields the most likely route of passage from p to q over $[0, T]$ for X_t .

Note that $q = m(T)$ is a special case. Here, X_t will remain within a small tube around $m(t)$ with probability converging to one as $\varepsilon \rightarrow 0$. That is, the most likely transition path from p to q over $[0, T]$ is simply the path of the mean in this case.

To effectively compute the most likely transition path from p to q for X_t , we have implemented a numerical scheme in three steps:

- Solve several ODEs with delay to compute the mean path $m(t)$ of X_t and the covariance function $\rho(s, t)$ of Z_t .
- For fixed T, p, q , compute the most likely transition path $h = h^T$ from $X_0 = p$ to $X_T = q$, and its energy $\lambda(h^T)$, as given by (17) and (18).
- Compute the optimal transition time T_{opt} by minimizing $\lambda(h^T)$ over all times $T > 0$.

3. NUMERICAL IMPLEMENTATION

3.1. Numerical solution of three delay ODEs. Each delay ODE of interest here is iteratively solved on the time intervals $J_k = [(k-1)\tau, k\tau]$ for $k = 1, 2, \dots, (1 + \lceil T/\tau \rceil)$. For each k , this amounts to solving numerically a linear ODE with known right-hand side. For this, we use a backward Euler scheme, which is known to be stable for equations of this form [6, 19]. To compute $m(t)$, we discretize $[0, T]$ into subintervals of equal length $\Delta t = \tau/N$. Backward Euler is given by

$$m(t) - m(t - \Delta t) = [a + Bm(t) + Cm(t - \tau)]\Delta t.$$

which yields the recursive equation

$$m(t) = (I - \Delta t B)^{-1}m(t - \Delta t) + \Delta t(I - \Delta t B)^{-1}[a + Cm(t - \tau)].$$

The initial history of the mean is used to numerically compute the solution $m(t) = m_1(t)$ on the initial interval J_1 . To numerically generate the solution $m(t) = m_k(t)$ on J_k , we proceed by iteration on k , using the discretized expressions just stated above. This yields a full numerical approximation of $m(t)$ on $[0, T]$. We apply a completely similar strategy to compute for each s the function $t \rightarrow \phi_s(t)$ as defined by (12). However both s and t will be constrained to belong to the finite grid

$$\text{Grid}(N) = \{j\tau/N : j = 1, \dots, M \text{ and } M = N(1 + \lceil T/\tau \rceil)\}.$$

After the computation of $\phi_s(t)$, we generate the $F(s, t)$ values for s and t in $\text{Grid}(N)$ by the explicit formula $F(s, t) = \phi_s(t) - \Sigma H(s - t)$ where $H(s - t)$ is a Heaviside function.

We then proceed to compute $\rho(s, t)$ for s and t in $\text{Grid}(N)$. For each fixed t in $\text{Grid}(N)$, the Backward Euler discretization of the delay ODE verified by the function $s \rightarrow \rho(s, t)$ yields the recursive relation

$$(19) \quad \rho(s, t) = (I - \Delta s B)^{-1}\rho(s - \Delta s, t) + \Delta s(I - \Delta s B)^{-1}[C\rho(s - \tau, t) + \Sigma F(s, t)]$$

where $\Delta s = \tau/N$. The initial values $\rho(s, t) = 0$ for $s \in [-\tau, 0]$ and the recursive relation (19) enable the computation of $\rho(s, t)$ for $s \in J_1$. Keeping t fixed, one then uses (19) as above and the values of $\rho(\cdot, t)$ on J_k to compute the values of $\rho(\cdot, t)$ on J_{k+1} . Repeating this operation for each t in $\text{Grid}(N)$ finally provides $\rho(s, t)$ for s and t in $\text{Grid}(N)$.

3.2. Numerical minimization of the Cramer transform. Fix $T > 0$. For the Gaussian diffusion with delay X_t , the most likely transition path h from $X_0 = p$ to $X_T = q$ and its energy $\lambda_h(T)$ have been explicitly expressed in terms of the functions $m(t)$ and $\rho(s, t)$ (see (17) and (18)). Plugging into these two formulas the values of $m(t)$ and $\rho(s, t)$ numerically computed for s and t in $\text{Grid}(N)$ immediately provides numerical approximations of $h(s)$ for s in $\text{Grid}(N)$ and of $\lambda_h(T)$ for a fixed terminal time T .

To compute the most likely time at which X_t will reach q , whenever this rare event is realized, we have to minimize $u(T) = \lambda_h(T)$ in T . So we select a large terminal time T_{large} , and we numerically minimize the function $u(T)$ on the interval $[0, T_{\text{large}}]$. If on that time interval $u(T)$ exhibits an actual minimum at T_{opt} , this gives us an approximate most likely transition time T_{opt} . Otherwise, we set $T_{\text{opt}} = \infty$.

3.3. Exit path from nominally stable stationary states. As $\varepsilon \rightarrow 0$, the limit dynamics of X_t is a deterministic dynamic system x_t driven by an obvious first-order ODE with delay. Let p be a stable stationary state of x_t , and let V be a small neighborhood of p . Determining for small ε the most likely path followed by X_t to exit from V when $X_0 = p$ is a problem of practical interest in many contexts. Our numerical computation of the most likely transition path from $X_0 = p$ to $X_T = q$ with q on the boundary of V will enable us to numerically solve these types of exit problems. We now illustrate this approach with the detailed study of a specific dynamical system from biochemistry.

4. LINEAR NOISE APPROXIMATIONS FOR EXCITABLE SYSTEMS

4.1. Excitable systems from biochemistry. We begin by explaining the importance of noise, delay, and metastability for the dynamics of genetic regulatory circuits. Such circuits may be described by delay SDEs [9, 20] and represent a significant class of systems to which our approach can be applied.

Cellular noise and transcriptional delay shape the dynamics of genetic regulatory circuits. Stochasticity within cellular processes arises from a variety of sources. Sequences of chemical reactions at low molecule numbers produce an intrinsic form of noise. Multiple other types of variability affect dynamics across spatial and temporal scales. Examples include fluctuations in environmental conditions, metabolic processes, energy availability, et cetera. Noise functions constructively in both microbial and eukaryotic cells and on multiple timescales. It enables probabilistic differentiation strategies for cell populations, such as stochastic state-switching in bistable circuits and transient cellular differentiation in excitable circuits (e.g. [12, 15, 41]).

Certain circuit architectures such as toggle switches and excitable circuits enable noise-induced rare events. These architectures allow cellular populations to probabilistically switch states in response to environmental fluctuations [15].

Bistability is a central characteristic of biological switches. It is essential in the determination of cell fate in multicellular organisms [24], the regulation of cell cycle oscillations during mitosis [21], and the maintenance of epigenetic traits in microbes [38]. Metastable states can be created by positive feedback loops. Once a trajectory enters a metastable state, it will typically remain there for a considerable amount of time before noise induces a transition [15, 25]. This phenomenon has been studied in many contexts, including the lysis/lysogeny switch of bacteriophage λ [1, 43], bacterial persistence [4], and synthetically constructed positive feedback loops [18, 37].

Many biological systems exhibit excitability [13, 34, 41]. Excitable systems commonly feature a single metastable state bordered by a sizable, active region of phase space. When stochastic fluctuations cause a trajectory to exit the basin of attraction of this metastable state, the trajectory will make a large excursion before returning to the basin. Transient differentiation into a genetically competent state in *Bacillus subtilis*, for example, is enabled by an excitable circuit architecture. Positive feedback controls the threshold for competent event initiation, while a slower negative feedback loop controls the duration of competence events [10, 30, 31, 40, 41]. Rare events in such excitable systems manifest as bursts of activity.

4.2. General linear noise approximations (LNAs). We explain how Gaussian diffusions driven by delay SDEs such as (1) arise from linear noise approximations of nonlinear delay SDEs. Brett and Galla [9] introduced linear noise approximations for chemical Langevin equations modeling biochemical reaction networks. Consider the delay SDE

$$(20) \quad dx_t = f(x(t), x(t - \tau)) dt + \frac{1}{\sqrt{N}} g(x(t), x(t - \tau)) dW_t.$$

Here $f : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$, $g : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times n}$, W_t denotes standard n -dimensional Brownian motion, and $N > 0$ denotes system size (characteristic number of molecules in a biochemical system). Notice that we allow both the drift and the diffusion to depend on the past. Suppose $x^\infty(t)$ solves the deterministic limit of (20); that is, $x^\infty(t)$ solves

$$(21) \quad dx_t = f(x(t), x(t - \tau)) dt.$$

As we have indicated in our introduction, around a stable point z of the limit ODE as N tends to infinity, one can approximate such a system by a Gaussian diffusion with delay and small diffusion matrix $\frac{1}{\sqrt{N}} \Sigma$. Define $\xi(t)$ by

$$x(t) = x^\infty(t) + \xi(t).$$

Substituting this ansatz into (20) and performing Taylor expansions of f and g based at the deterministic trajectory yields the linear noise approximation

$$(22) \quad \begin{aligned} d\xi_t = & [D_1 f(x^\infty(t), x^\infty(t-\tau))\xi(t) + D_2 f(x^\infty(t), x^\infty(t-\tau))\xi(t-\tau)] dt \\ & + \frac{1}{\sqrt{N}} g(x^\infty(t), x^\infty(t-\tau)) dW_t. \end{aligned}$$

Here D_1 and D_2 denote differentiation with respect to the first and second sets of d arguments, respectively. If $x^\infty(t)$ happens to be a stable fixed point of (21), say $x^\infty(t) \equiv z$, then (22) becomes

$$d\xi_t = [D_1 f(z, z)\xi(t) + D_2 f(z, z)\xi(t-\tau)] dt + \frac{1}{\sqrt{N}} g(z, z) dW_t.$$

This is (1) with $a = \mathbf{0}$, $B = D_1 f(z, z)$, $C = D_2 f(z, z)$, $\Sigma = g(z, z)$, and $\varepsilon = \frac{1}{\sqrt{N}}$.

5. A BISTABLE BIOCHEMICAL SYSTEM

5.1. Chemical Langevin equation. The genetic toggle switch we study consists of two protein species, each of which represses the production of the other. We model the switch using the chemical Langevin equation

$$(23a) \quad dx = \left(\frac{\beta}{1 + y(t-\tau)^2/k} - \gamma x \right) dt + \frac{1}{\sqrt{N}} \left(\frac{\beta}{1 + y(t-\tau)^2/k} + \gamma x \right)^{\frac{1}{2}} dW_1$$

$$(23b) \quad dy = \left(\frac{\beta}{1 + x(t-\tau)^2/k} - \gamma y \right) dt + \frac{1}{\sqrt{N}} \left(\frac{\beta}{1 + x(t-\tau)^2/k} + \gamma y \right)^{\frac{1}{2}} dW_2,$$

where x and y denote the concentrations of the two protein species, β denotes maximal protein production rate, k is the protein level at which production is cut in half, γ is the dilution rate, N denotes system size, and W_1 and W_2 are independent standard Brownian motions. Notice that (23) is a symmetric system. In the deterministic limit as $N \rightarrow \infty$, the co-repressive toggle switch is described by the reaction rate equations

$$(24a) \quad dx = \left(\frac{\beta}{1 + y(t-\tau)^2/k} - \gamma x \right) dt$$

$$(24b) \quad dy = \left(\frac{\beta}{1 + x(t-\tau)^2/k} - \gamma y \right) dt.$$

System (24) has two stable stationary states, $(x_{\text{low}}, y_{\text{high}})$ and $(x_{\text{high}}, y_{\text{low}})$, as well as a saddle stationary state (x_s, y_s) . See [20, Figure 7] or [42, Figure 3A, inset] for a phase portrait of (24).

In the stochastic ($N < \infty$) regime, a typical trajectory of the co-repressive toggle switch will spend most of its time near the metastable states, occasionally hopping from one to the other [42, Figure 3A]. Such rare events raise interesting questions. For large N , is the co-repressive toggle switch well-approximated by a two-state Markov chain on long timescales? If so, what are the transition rates? To determine these rates, one would need to compute both a quasipotential and a formula of Eyring-Kramers type.

Here, we focus on the problem of optimal escape from neighborhoods of metastable states. We fix a neighborhood of $(x_{\text{low}}, y_{\text{high}})$ (Figure 1, black curve) and ask: What is the most likely route of escape from this neighborhood for (23)? In Section 5.2, we compute a linear noise approximation of (23) that is valid near $(x_{\text{low}}, y_{\text{high}})$. Since this linear noise approximation is a Gaussian diffusion with delay, the framework of the present paper applies to it. We use this framework to compute most likely routes of escape for the linear noise approximation and thereby obtain (approximate) most likely routes of escape for (23).

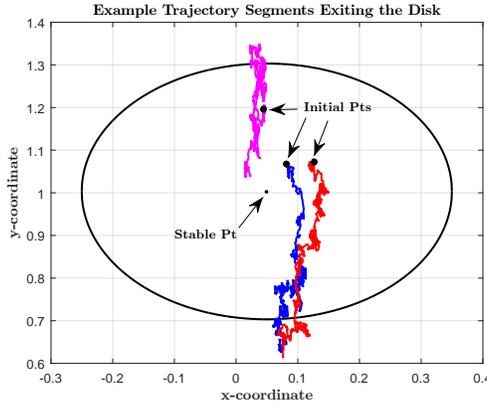


Figure 1. Sample trajectory segments of (23) in a neighborhood of the metastable state $(x_{\text{low}}, y_{\text{high}})$. We simulated 1000 trajectories over the time interval $[0, 5]$. We then chose three sample trajectories that exited the disk D and extracted a segment from each of them. The blue, red, and magenta trajectory segments begin near the metastable state (small black disk) at the coordinates $(0.0817, 1.0668)$, $(0.0673, 1.1233)$, and $(0.1272, 1.0733)$, respectively, and cover time intervals $[2.799, 3.399]$, $[3.099, 3.599]$, and $[1.699, 2.299]$. The history of each simulated trajectory over the time interval $[0, 5]$ is taken to be fixed at $(x_{\text{low}}, y_{\text{high}})$ over the time interval $[-\tau, 0]$. Trajectories have been generated using Euler-Maruyama with time step $\Delta t = \tau/1000 = 0.001$. Parameters: $\beta = 0.73$, $k = 0.05$, $\gamma = \ln(2)$, $\tau = 1$, $N = 30$.

5.2. Approximation by Gaussian Diffusions with delays. We study an approximation of (23) by Gaussian diffusions with delay that is valid in a neighborhood of $(x_{\text{low}}, y_{\text{high}}) =: (v, w)$. Writing

$$x(t) = v + \xi_1(t), \quad y(t) = w + \xi_2(t),$$

the Gaussian diffusion with delay is given by

$$(25) \quad \begin{aligned} d\xi_1(t) &= \left(-\gamma\xi_1(t) - \frac{2\beta w}{k[1+w^2/k]^2} \xi_2(t-\tau) \right) dt + \frac{1}{\sqrt{N}} \left(\frac{\beta}{1+w^2/k} + \gamma v \right)^{1/2} dW_1(t), \\ d\xi_2(t) &= \left(-\gamma\xi_2(t) - \frac{2\beta v}{k[1+v^2/k]^2} \xi_1(t-\tau) \right) dt + \frac{1}{\sqrt{N}} \left(\frac{\beta}{1+v^2/k} + \gamma w \right)^{1/2} dW_2(t). \end{aligned}$$

We are now in position to apply the large deviations framework of our paper to (25). Before doing so, we perform a preliminary numerical calculation and comment on the role of trajectory histories.

We numerically compute the stationary points of (24). We work with the parameter set $\beta = 0.73$, $k = 0.05$, $\gamma = \ln(2)$, and $\tau = 1$, a parameter set for which (24) has two stable stationary states and one saddle stationary state. We find these states by setting the drift expressions in (24) equal to zero along with $x(t-\tau) = x$ and $y(t-\tau) = y$. Approximate solutions can be found numerically using many well-known iterative methods. The two stable stationary states are approximately $(v, w) \approx (0.0498, 1.0033)$ and $(1.0033, 0.0498)$. The stationary saddle is approximately $(0.3306, 0.3306)$.

Notice that since the Gaussian diffusion (25) contains delay, one must specify a trajectory history over the time interval $[-\tau, 0]$ in order to properly initialize the equation. Trajectory history will influence the evolution of the mean of the Gaussian diffusion with delay and will therefore affect the computation of optimal large deviations trajectories. In general, this history may be deterministic or random. For our current study, we work with deterministic histories and take them to be constant on $[-\tau, 0]$. See Figure 2 for examples of the evolution of the mean of the Gaussian diffusion using various histories. Finally, note that although the process $\xi(t)$ is Gaussian, it will not be centered if the history is not identically zero. To be consistent with the notation of Section 2.11, we write

the process that locally approximates the delay chemical Langevin equation as $m(t) + \varepsilon Z_t$, where $m(t) = \mathbb{E}[\xi(t)]$, $\varepsilon = \frac{1}{\sqrt{N}}$, and Z_t satisfies (25) with no small parameter ($N = 1$) and history zero.

5.3. Optimal escape trajectories and exit points - analysis. We now apply our large deviations framework to the Gaussian diffusion that approximates the delay chemical Langevin equation (23) near (v, w) . We begin with an analytical view and then follow with numerical simulation.

We find the most likely exit path with constant initial history $m(0)$ that exits the disk

$$D = \left\{ (z_1, z_2) : (z_1 - v)^2 + (z_2 - w)^2 \leq R^2 \right\}.$$

(We choose $R = 0.3$ for the numerical computations in Section 5.4 so that the neighborhood of (v, w) has size of order one but remains bounded away from the separatrix.) To find this optimal path, we first find the path of least energy that exits D at a preselected point $q \in \partial D$ and at a preselected time T . We then optimize over T and q . For fixed exit time T and exit point $q \in \partial D$, the optimal escape path and associated energy are given by

$$\begin{aligned} h(s) &= \rho(s, T) \left[\rho(T, T)^{-1} (q - m(T)) \right] + m(s) \\ \lambda_h(T, q) &= \frac{1}{2} \left[\rho(T, T)^{-1} (q - m(T)) \right] \cdot (q - m(T)) \end{aligned}$$

using (17) and (18). Here, s ranges over $[0, T]$ and $\rho(s, t)$ is the covariance matrix of Z_t at times $s, t \in [0, T]$. Note that we are using the terms “exit time” and “escape path” loosely since we do not impose the *a priori* condition that h remain inside D until it reaches q at time T .

In order to optimize over q and T , we first fix T and optimize $\lambda_h(T, q)$ over points $q \in \partial D$. Notice that $\lambda_h(T, q)$ is a classical quadratic form on \mathbb{R}^2 for fixed T , so we apply standard minimization techniques to find the minimizer $\hat{q}(T)$ analytically. The minimization problem for fixed T is

$$(26) \quad \min_q \lambda_h(T, q) \quad \text{subject to} \quad (q_1 - v)^2 + (q_2 - w)^2 = R^2.$$

Using a Lagrange multiplier $\mu \in \mathbb{R}$, define the Lagrangian

$$L_\mu(q) := \lambda_h(T, q) - \mu((q_1 - v)^2 + (q_2 - w)^2 - R^2).$$

Calculating the gradient $\nabla_q(L_\mu(q))$ and setting the gradient equal to zero yields the equation

$$(27) \quad \rho(T, T)^{-1} (q - m(T)) = 2\mu(q - (v, w)^*).$$

Notice that if $m(T) = (v, w)^*$, then (27) becomes an eigenvalue problem for $\rho(T, T)^{-1}$. In this case, the optimal exit point $\hat{q}(T)$ is such that $\hat{q}(T) - (v, w)^*$ is the eigenvector of $\rho(T, T)^{-1}$ corresponding to the smallest eigenvalue, and the energy of the optimal path that exits D at time T is proportional to this smallest eigenvalue.

This observation has two implications. First, if the history of the linear noise process is taken to be $m(t) = (v, w)^*$ on $[-\tau, 0]$, then we will have $m(t) = (v, w)^*$ for all $t \geq 0$ as well. In this case, minimizing $\lambda_h(T, q)$ over T and q to find the optimal escape time T_{opt} and the optimal escape point $\hat{q}(T_{\text{opt}})$ amounts to minimizing the smallest eigenvalue of $\rho(T, T)^{-1}$ over T . Since (25) is essentially an Ornstein-Uhlenbeck process with delay, we expect the smallest eigenvalue of $\rho(T, T)^{-1}$ to decrease monotonically toward a limiting value as $T \rightarrow \infty$. See Figure 4 for numerical evidence. There exists no minimizer of $\lambda_h(T, q)$ in this case, as we would have $T_{\text{opt}} = \infty$.

Second, regardless of the initial history of the linear noise process, $m(T) \rightarrow (v, w)^*$ as $T \rightarrow \infty$ for the parameters we have selected. Consequently, (27) is approximately an eigenvalue problem for large values of T , so for such T the optimal exit point $\hat{q}(T)$ will be such that $\hat{q}(T) - (v, w)^*$ is close to the eigenvector of $\rho(T, T)^{-1}$ corresponding to the smallest eigenvalue.

5.4. Numerical results. We compute the optimal path of escape, the optimal exit time T_{opt} , and the optimal exit point $\hat{q}(T_{\text{opt}}) \in \partial D$ for the linear noise process (25) that approximates the toggle switch (23) in the disk D . Along the way, we discuss interesting related computations.

Parameter selection. We set $\beta = 0.73$, $k = 0.05$, $\gamma = \ln(2)$, and $\tau = 1$ for the toggle switch. System size for the linear noise approximation (25) is $N = 1000$. The history of the linear noise process is taken to be the constant position $(0.0453, 1.1323)$ over the time interval $[-\tau, 0]$. We choose $R = 0.3$ for the radius of D so that this neighborhood of (v, w) has size of order one but remains bounded away from the separatrix.

Optimization algorithm. To compute the optimal escape path, exit time, and exit point, we execute the following algorithm.

- Simulate the mean and covariance equations for a sufficiently large T_{large} using step sizes $\Delta t = \Delta s = \tau/500$.
- Discretize the boundary of the disk D using discretization $\Delta r = 0.006$ of $[-R, R]$.
- For each time $t_j = (j - 1)\Delta t \in [0, T_{\text{large}}]$ and each point q_k on the discretized boundary of the disk, compute the optimal trajectory that exits at time t_j through q_k as well as the energy $E_{j,k}$ of this trajectory.
- Minimize over the entries of the matrix E in order to find the optimal exit time and exit point (and hence the overall optimal path of escape).

Mean and covariance. We first compute the mean and covariance of the linear noise process. Figure 2 (blue curves) illustrates the evolution of the mean for our parameter set. As expected, the mean converges to the stationary state (v, w) (moved to $(0, 0)$ in Figure 2). It is important to choose T_{large} sufficiently large so that the covariance matrix $\rho(T_{\text{large}}, T_{\text{large}})$ has stabilized and the mean is close to the stationary state. Fig. 3 and Fig. 4 provide evidence that this stabilization occurs by time $T = 20$ for our parameter set. In particular, the variances of the two components of Z_t stabilize by time 20 (Fig. 3). Fig. 4 illustrates that the smallest eigenvalue of $\rho(t, t)^{-1}$ stabilizes as well.

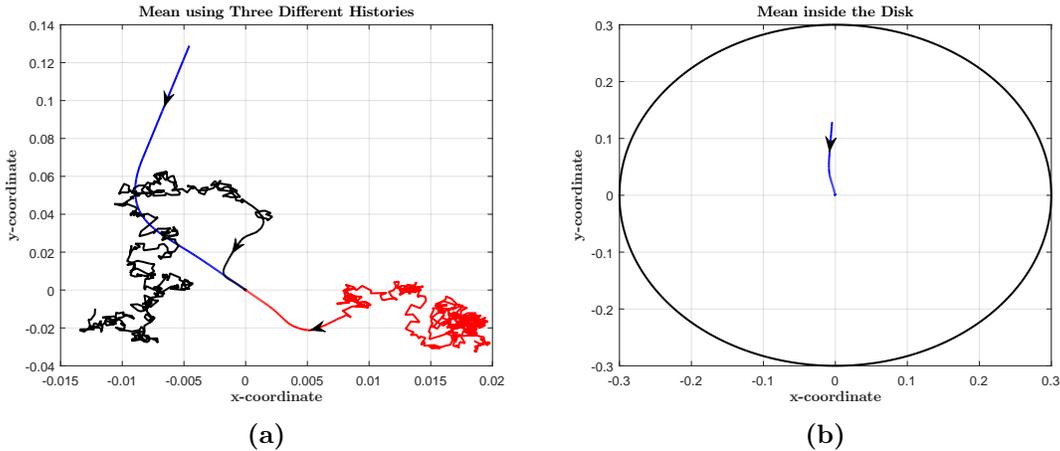


Figure 2. Evolution of the mean of the linear noise process. Here the stationary state (v, w) has been shifted to the origin. (2a) Blue curve: evolution of the mean using the constant history $(0.0453, 1.1323)$ (or $(-0.0046, 0.1289)$ in local coordinates). Red and black curves: evolution of the mean using trajectory segments of (23) for histories. In all three cases, the mean converges to the stationary state. (2b) Another view of the blue curve from Fig. 2a.

Numerical optimization results. We first examine the behavior of optimal paths and optimal path energies for fixed exit times. Fig. 5a illustrates the behavior of optimal path energy as a

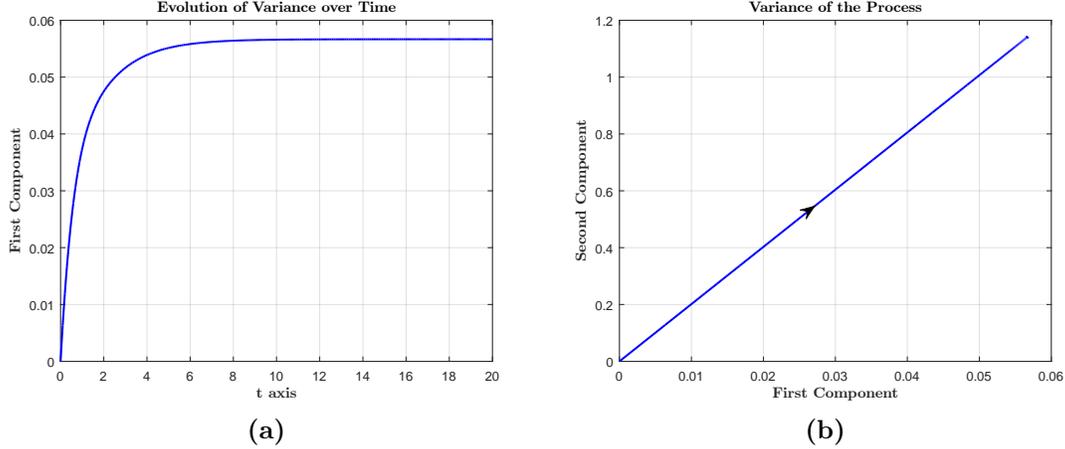


Figure 3. Evolution of the variances of the two components of Z_t over time. **(3a)** The variance of the first component stabilizes to approximately 0.0567 by time 20. **(3b)** A linear relationship exists between the evolutions of variances of the first and second components. By time 20, the variance of the second component has stabilized to approximately 1.1409.

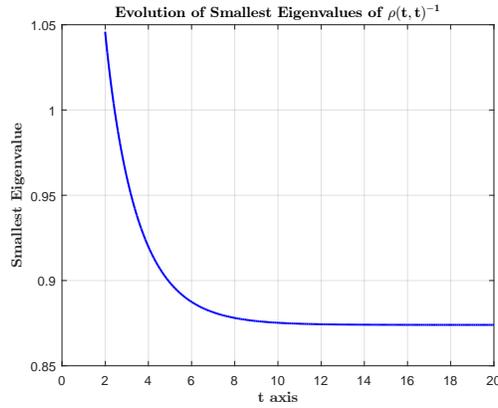


Figure 4. The smallest eigenvalue of the covariance matrix $\rho(t, t)^{-1}$ stabilizes to approximately 0.874 by time 20.

function of exit point over the upper half of ∂D for the fixed exit time $T = 10$. Note that optimal path energy is minimized near the top of ∂D . Fig. 5c depicts three different optimal escape paths for fixed escape time $T = 20$ and three different exit points. Notice that these trajectories follow the mean for some time before breaking away toward their respective exits. This behavior should not occur for the optimal exit time T_{opt} and the optimal exit point $\hat{q}(T_{\text{opt}})$. Fig. 5d (blue curve) illustrates the overall optimal escape trajectory. This trajectory exits at time $T_{\text{opt}} = 1.482$ and exit point $\hat{q}(T_{\text{opt}}) = (0.0384, 1.3031)$. Observe that the overall optimal escape trajectory diverges from the mean immediately.

Fig. 6 depicts overall optimal escape trajectories using three different constant initial histories. Notice that if the initial history is located in the lower half of D , then the overall optimal escape trajectory exits through the lower half of ∂D . This happens for the upper half of D as well. This behavior is natural, since moving ‘across’ the stationary state should not be energetically optimal. For the initial history corresponding to the blue curve in Fig. 6, the optimal escape path that exits through the bottom half of ∂D does so through $(0.0162, -0.2996)$ (in local coordinates) at exit time ∞ with energy 0.0394. This energy is strictly larger than that of the blue curve in Fig. 6 (0.0348).

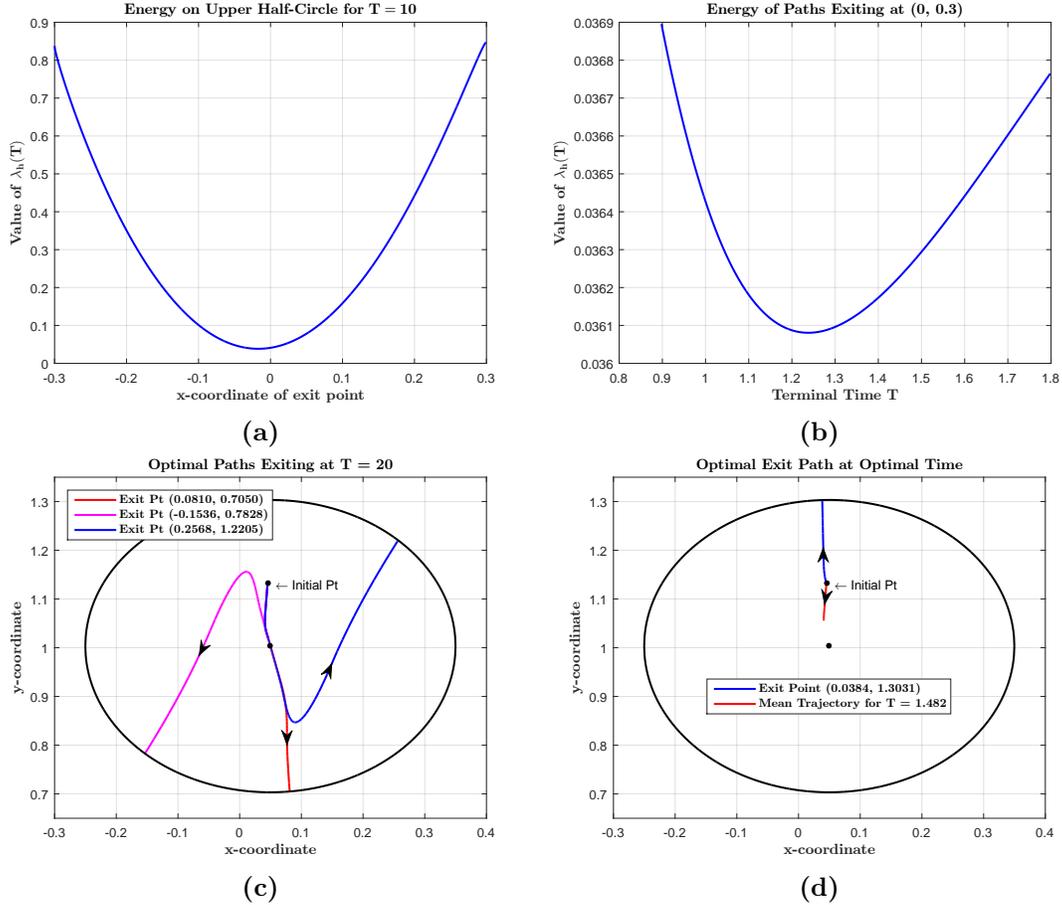


Figure 5. Optimal exit data. **(5a)** Energy of the optimal exit path at time $T = 10$ as a function of chosen exit point on the upper half of ∂D . The energy is minimized near the top of D . **(5b)** Energy of the optimal exit path as a function of exit time T for fixed exit point $(0, 0.3)$ (the top of D in local coordinates). **(5c)** Three different optimal escape paths for fixed escape time $T = 20$ and three different exit points. Notice that these trajectories follow the mean for some time before breaking away toward their respective exits. This behavior should not occur for the optimal exit time T_{opt} and the optimal exit point $\hat{q}(T_{\text{opt}})$. Energy values associated with the red, magenta, and blue trajectories are 0.0413, 0.4527, and 0.4661, respectively. **(5d)** Overall optimal escape trajectory. This trajectory exits at time $T_{\text{opt}} = 1.482$ and exit point $\hat{q}(T_{\text{opt}}) = (0.0384, 1.3031)$ with energy 0.0348.

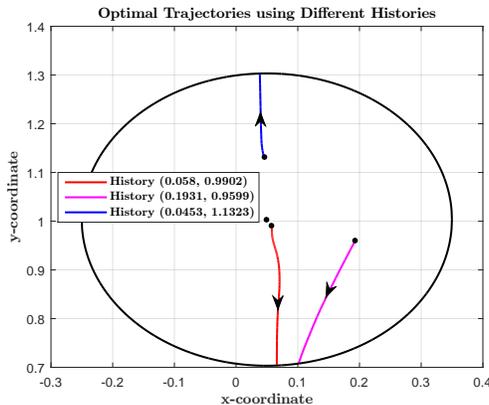


Figure 6. Optimal escape trajectories from D for the linear noise process using three different constant initial histories. Notice that if the initial history is located in the lower half of D , then the optimal escape trajectory exits through the lower half of ∂D . This happens for the upper half of D as well. The energy associated with the red, magenta, and blue trajectories is 0.0389, 0.0074, and 0.0348, respectively.

REFERENCES

- [1] E. AURELL AND K. SNEPPEN, *Epigenetics as a first exit problem*, Phys Rev Lett, 88 (2002), p. 048101.
- [2] R. AZENCOTT, M. I. FREIDLIN, AND S. R. S. VARADHAN, *Large deviations at Saint-Flour*, Probability at Saint-Flour, Springer, Heidelberg, 2013.
- [3] C. T. H. BAKER AND E. BUCKWAR, *Numerical analysis of explicit one-step methods for stochastic delay differential equations*, LMS J. Comput. Math., 3 (2000), pp. 315–335 (electronic).
- [4] N. Q. BALABAN, J. MERRIN, R. CHAIT, L. KOWALIK, AND S. LEIBLER, *Bacterial persistence as a phenotypic switch*, Science, 305 (2004), pp. 1622–1625.
- [5] J. BAO AND C. YUAN, *Large deviations for neutral functional SDEs with jumps*, Stochastics, 87 (2015), pp. 48–70.
- [6] A. BELLEN AND M. ZENARO, *Numerical methods for delay differential equations*, Numerical Mathematics and Scientific Computation, The Clarendon Press, Oxford University Press, New York, 2003.
- [7] F. BOUCHET, T. GRAFKE, T. TANGARIFE, AND E. VANDEN-EIJNDEN, *Large deviations in fast-slow systems*, J. Stat. Phys., 162 (2016), pp. 793–812.
- [8] F. BOUCHET, J. LAURIE, AND O. ZABORONSKI, *Langevin dynamics, large deviations and instantons for the quasi-geostrophic model and two-dimensional Euler equations*, J. Stat. Phys., 156 (2014), pp. 1066–1092.
- [9] T. BRETT AND T. GALLA, *Stochastic processes with distributed delays: Chemical langevin equation and linear-noise approximation*, Physical Review Letters, 110 (2013).
- [10] T. ÇAĞATAY, M. TURCOTTE, M. ELOWITZ, J. GARCIA-OJALVO, AND G. SÜEL, *Architecture-dependent noise discriminates functionally analogous differentiation circuits*, Cell, 139 (2009), pp. 512–522.
- [11] A. CHIARINI AND M. FISCHER, *On large deviations for small noise Itô processes*, Adv. in Appl. Probab., 46 (2014), pp. 1126–1147.
- [12] C. DAVIDSON AND M. SURETTE, *Individuality in bacteria*, Annual Review of Genetics, 42 (2008), pp. 253–268.
- [13] E. DUPIN, C. REAL, C. GLAVIEUX-PARDANAUD, P. VAIGOT, AND N. M. LE DOUARIN, *Reversal of developmental restrictions in neural crest lineages: Transition from schwann cells to glial-melanocytic precursors in vitro*, Proceedings of the National Academy of Sciences, 100 (2003), pp. 5229–5233.
- [14] W. E. W. REN, AND E. VANDEN-EIJNDEN, *Minimum action method for the study of rare events*, Comm. Pure Appl. Math., 57 (2004), pp. 637–656.
- [15] A. ELДАР AND M. ELOWITZ, *Functional roles for noise in genetic circuits*, Nature, 467 (2010), pp. 167–173.
- [16] M. I. FREIDLIN AND A. D. WENTZELL, *Random perturbations of dynamical systems*, vol. 260 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer, Heidelberg, third ed., 2012. Translated from the 1979 Russian original by Joseph Szücs.
- [17] S. GADAT, F. PANLOUP, AND C. PELLEGRINI, *Large deviation principle for invariant distributions of memory gradient diffusions*, Electron. J. Probab., 18 (2013), pp. no. 81, 34.
- [18] T. S. GARDNER, C. R. CANTOR, AND J. J. COLLINS, *Construction of a genetic toggle switch in Escherichia coli*, Nature, 403 (2000), pp. 339–342.

- [19] N. GUGLIELMI, *Delay dependent stability regions of Θ -methods for delay differential equations*, IMA J. Numer. Anal., 18 (1998), pp. 399–418.
- [20] C. GUPTA, J. M. LÓPEZ, R. AZENCOTT, M. R. BENNETT, K. JOSIĆ, AND W. OTT, *Modeling delay in genetic networks: From delay birth-death processes to delay stochastic differential equations*, J Chem Phys, 140 (2014), p. 204108.
- [21] E. HE, O. KAPUY, R. A. OLIVEIRA, F. UHLMANN, J. J. TYSON, AND B. NOVÁK, *Systems-level feedbacks make the anaphase switch irreversible*, Proc Natl Acad Sci USA, 108 (2011), pp. 10016–10021.
- [22] M. HEYMANN AND E. VANDEN-EIJNDEN, *The geometric minimum action method: a least action principle on the space of curves*, Comm. Pure Appl. Math., 61 (2008), pp. 1052–1117.
- [23] D. J. HIGHAM, *An algorithmic introduction to numerical simulation of stochastic differential equations*, SIAM Rev., 43 (2001), pp. 525–546 (electronic).
- [24] T. HONG, J. XING, L. LI, AND J. J. TYSON, *A simple theoretical framework for understanding heterogeneous differentiation of $cd4^+$ t cells*, BMC Syst Biol, 6 (2012), p. 66.
- [25] T. B. KEPLER AND T. C. ELSTON, *Stochasticity in transcriptional regulation: origins consequences, and mathematical representations*, Biophys J, 81 (2001), pp. 3116–3136.
- [26] H. J. KUSHNER, *Large deviations for two-time-scale diffusions, with delays*, Appl. Math. Optim., 62 (2010), pp. 295–322.
- [27] T. LI, X. LI, AND X. ZHOU, *Finding transition pathways on manifolds*, Multiscale Model. Simul., 14 (2016), pp. 173–206.
- [28] B. S. LINDLEY AND I. B. SCHWARTZ, *An iterative action minimizing method for computing optimal paths in stochastic dynamical systems*, Phys. D, 255 (2013), pp. 22–30.
- [29] D. G. LUENBERGER, *Optimization by vector space methods*, John Wiley & Sons, Inc., New York-London-Sydney, 1969.
- [30] H. MAAMAR AND D. DUBNAU, *Bistability in the bacillus subtilis k-state (competence) system requires a positive feedback loop*, Molecular Microbiology, 56 (2005), pp. 615–624.
- [31] H. MAAMAR, A. RAJ, AND D. DUBNAU, *Noise in gene expression determines cell fate in bacillus subtilis*, Science, 317 (2007), pp. 526–529.
- [32] X. MAO, *Stochastic differential equations and their applications*, Horwood Publishing Series in Mathematics & Applications, Horwood Publishing Limited, Chichester, 1997.
- [33] X. MAO AND S. SABANIS, *Numerical solutions of stochastic differential delay equations under local Lipschitz condition*, J. Comput. Appl. Math., 151 (2003), pp. 215–227.
- [34] J. C. MEEKS, E. L. CAMPBELL, M. L. SUMMERS, AND F. C. WONG, *Cellular differentiation in the cyanobacterium nostoc punctiforme*, Archives of Microbiology, 178 (2002), pp. 395–403.
- [35] C. MO AND J. LUO, *Large deviations for stochastic differential delay equations*, Nonlinear Anal., 80 (2013), pp. 202–210.
- [36] S.-E. A. MOHAMMED AND T. ZHANG, *Large deviations for stochastic systems with memory*, Discrete Contin. Dyn. Syst. Ser. B, 6 (2006), pp. 881–893 (electronic).
- [37] D. NEVOZHAY, R. M. ADAMS, E. VAN ITALLIE, M. R. BENNETT, AND G. BALÁZSI, *Mapping the environmental fitness landscape of a synthetic gene circuit*, PLoS Comp Biol, 8 (2012), p. e1002480.
- [38] E. M. OZBUDAK, M. THATTAI, H. N. LIM, B. I. SHRAIMAN, AND A. VAN OUDENAARDEN, *Multistability in the lactose utilization network of Escherichia coli*, Nature, 427 (2004), pp. 737–740.
- [39] A. A. PUHALSKII, *On some degenerate large deviation problems*, Electron. J. Probab., 9 (2004), pp. no. 28, 862–886.
- [40] G. SÜEL, R. KULKARNI, J. DWORKIN, J. GARCIA-OJALVO, AND M. ELOWITZ, *Tunability and noise dependence in differentiation dynamics*, Science, 315 (2007), pp. 1716–1719.
- [41] G. M. SUEL, J. GARCIA-OJALVO, L. M. LIBERMAN, AND M. B. ELOWITZ, *An excitable gene regulatory circuit induces transient cellular differentiation*, Nature, 440 (2006), pp. 545–550.
- [42] A. VELIZ-CUBA, C. GUPTA, M. R. BENNETT, K. JOSIĆ, AND W. OTT, *Effects of cell cycle noise on excitable gene circuits*. arXiv:1605.09328, 2016.
- [43] P. B. WARREN AND P. R. TEN WOLDE, *Chemical models of genetic toggle switches*, J Phys Chem B, 109 (2005), pp. 6812–6823.