Metastable behavior in Markov processes with internal states

Abstract. A perturbation framework is developed to analyze metastable behavior in stochastic processes with random internal and external states. The process is assumed to be under weak noise conditions, and the case where the deterministic limit is bistable is considered. A general analytical approximation is derived for the stationary probability density and the mean switching time between metastable states, which includes the pre exponential factor. The results are illustrated with a model of gene expression that displays bistable switching. In this model, the external state represents the number of protein molecules produced by a hypothetical gene. Once produced, a protein is eventually degraded. The internal state represents the activated or unactivated state of the gene; in the activated state the gene produces protein more rapidly than the unactivated state. The gene is activated by a dimer of the protein it produces so that the activation rate depends on the current protein level. This is a well studied model, and several model reductions and diffusion approximation methods are available to analyze its behavior. However, it is unclear if these methods accurately approximate long-time metastable behavior (i.e., mean switching time between metastable states of the bistable system). Diffusion approximations are generally known to fail in this regard.

1. Introduction

A common feature found in many stochastic models of biological processes is a distinction between internal and external states (Van Kampen, 1979). There are numerous examples of such Markov processes used as models for biological phenomena (Othmer et al., 1988; Bicout, 1997; Kepler and Elston, 2001; Friedman and Craciun, 2005; Newby and Bressloff, 2010). Examples of an internal state include the number of open ion channels in the membrane of a neuron that affect its membrane voltage (Keener and Newby, 2011) and the on/off state of a gene that affects its protein production rate (Newby, 2012). The distinction between internal and external states should not be confused with the concept of intrinsic and extrinsic noise (see (Thattai and van Oudenaarden, 2001) for an example related to gene
expression). A system with internal degrees of freedom is a classical idea in physics and applied mathematics, and the extension of this concept to Markov processes with internal states is well known in the literature (Hill and Hughes, 1985; Kramli and Szasz, 1983; Landman et al., 1977).

Consider the following two stochastic processes: the discrete internal state, $S(t)$, and the external state, $X(t)$. We consider two possibilities: $X(t) \in \mathbb{Z}$ and $X(t) \in \mathbb{R}$ (i.e., a discrete jump process and a continuous process). If $X(t)$ is independent of $S(t)$, there is one source of noise affecting $X(t)$, and we assume it is scaled by $1/\alpha_e$, where $\alpha_e \gg 1$, so that in the limit $\alpha_e \to \infty$, $X(t)$ is a deterministic process. Now consider the combined process where $S(t)$ and $X(t)$ are coupled. In this case, there is a second source of noise affecting $X(t)$ through its dependence on $S(t)$. We assume that there is a second large parameter, $\alpha_i$, such that in the limit $\alpha_i \to \infty$, the frequency of jumps in $S(t)$ becomes infinite. In this limit, $X(t)$ depends only on the average value of $S(t)$, effectively eliminating the second noise source. In the limit $\alpha_i \to \infty$, $\alpha_e \to \infty$, the combined process $(S(t), X(t)) \to \bar{x}(t)$, where $\bar{x}(t) \in \mathbb{R}$ is deterministic.

Under weak noise conditions, meaning close to the deterministic limit with $\alpha_i \gg 1$ and $\alpha_e \gg 1$, the dynamics of the deterministic system strongly influence the dynamics of the stochastic process. In particular, we are interested in the case where the deterministic system has multiple stable solutions depending on the initial conditions. On short timescales, a trajectory of the stochastic process fluctuates about the deterministic trajectory that has the same initial conditions. However, metastable behavior in the stochastic process is not seen in the deterministic system because it depends on a small amount of noise present in the system to cause a transition from one of the stable deterministic solutions to the other. Metastable transitions occur on a long timescale.

Metastable behavior is important because it represents fluctuation-induced phenomena not present in the deterministic system. The standard example of metastable behavior is Brownian motion in a double well potential. On short timescales the particle is most likely found near one of the two minima, and on long timescales the particle can transition over the energy barrier that separates each well. Metastable transitions by nonlinear Markov processes with an internal and external state are more difficult to analyze than diffusion in a potential well, and exact analytical solutions are rarely possible. Moreover, using Monte Carlo simulations to generate exact trajectories that display metastable behavior requires too much processor time to be practical. It is therefore necessary to develop approximation methods.

One approximation method is to reduce the complexity of the model by eliminating a noise source. Noise in the internal state is eliminated in the adiabatic limit, $\alpha_i \to \infty$, where $S(t)$ is averaged out of $(S(t), X(t))$ to obtain a Markov process that approximates $X(t)$. In other words, although $X(t)$ is not Markovian due to its dependence on $S(t)$, it may be approximately Markovian. Eliminating noise in the external state with $\alpha_e \to \infty$ results in a velocity jump process where the external state evolves deterministically.
Metastable behavior in Markov processes with internal states

in between random jumps in the internal state. However, the timescale for a metastable transition is very sensitive to both the type of noise and the noise strength, and eliminating a noise source can lead to large errors.

Another way to reduce the complexity of the model is with a diffusion approximation obtained using a quasi-steady-state (QSS) reduction (Gardiner, 1983; Thomas et al., 2012). This is very similar to the adiabatic limit, but uses a perturbation approach so that higher order terms can be included that account for noise in the internal state. The QSS reduction also approximates \((S(t), X(t))\) with a single continuous Markov process for \(X(t)\), but includes effects from both noise sources. This approximation reduces the problem to diffusion in a double well potential. The underlying assumption behind the QSS reduction is that \(S(t)\) is well approximated by a random variable chosen from its steady-state distribution conditioned on a fixed value of \(X(t)\). While the QSS diffusion approximation is a useful tool in most circumstances, it is not accurate for characterizing metastability.

To describe metastable behavior, it is necessary to approximate both the effective potential and the timescales for metastable transitions. For a 1D continuous Markov process on the state space \(x \in \mathbb{R}\), the potential is straightforward to define, and if a stationary solution exists, it must have zero probability flux everywhere. Given its usefulness at describing the qualitative features, we would like to know if we can define an effective potential in general. For higher dimensional continuous Markov processes, the potential is no longer well defined when the curl of the drift velocity field is nonzero, and it is possible for the stationary density to exhibit a nonzero probability flux. This is closely related to detailed balance conditions and thermodynamic equilibrium. Developing a systematic formalism to describe nonequilibrium stationary behavior is particularly relevant in biology. It turns out that an effective potential can still be defined using perturbation theory (Schuss, 2010; Ludwig, 1975; Matkowsky et al., 1983; Talkner, 1987; Nach et al., 1990; Maier and Stein, 1997; Hanggi et al., 1984; Dykman et al., 1994) and large deviation theory (Freidlin and Wentzell, 1998; Metzner et al., 2009; Heymann and Vanden-Eijnden, 2008). These tools can also be used to approximate the timescale associated with metastable transitions. The methods presented here fit within the perturbation framework.

The theory of large deviations (Freidlin and Wentzell, 1998; Shwartz and Weiss, 1995; Feng and Kurtz, 2006) is the mathematical foundation for the techniques used to study metastable transitions (rare events). Here, we focus on perturbation-theory-based techniques (Schuss, 2010), which we refer to as the quasi-stationary analysis (QSA). Large deviation theory provides rigorous results and error estimates, but does not provide a means of explicitly calculating the pre exponential factor (see Section 3.1), which is part of the leading order transition time and stationary density approximations. The QSA is formal but systematic and generally more practical for applications. The QSA was developed to analyze the differential Chapman–Kolmogorov (CK) equation, which describes the process by its probability density function. For a continuous Markov process, the QSA is
well-developed for the Fokker–Planck equation (Schuss, 2010; Ludwig, 1975; Matkowsky et al., 1983; Talkner, 1987; Naeh et al., 1990; Maier and Stein, 1997). The QSA has also been applied to the Master equation to analyze certain birth-death processes (Hanggi et al., 1984; Dykman et al., 1994; Hinch and Chapman, 2005; Doering et al., 2005; Vellela and Qian, 2007; Doering et al., 2007; Escudero and Kamenev, 2009; Bressloff, 2010; Assaf et al., 2011). (The Fokker–Planck and Master equation are instances of the more general CK equation (Gardiner, 1983).) However, for weak noise problems where adiabatic elimination (i.e., stochastic averaging) of one noise source is necessary to reach the deterministic system, no one has developed these methods (as far as we are aware) to study metastable behavior without first applying a QSS-type diffusion approximation or adiabatic reduction, which has been shown to result in significant errors (Freidlin and Wentzell, 1998; Newby and Bressloff, 2010). Recently, the QSA has been developed for the velocity-jump process (also sometimes called a piecewise deterministic process or hybrid process) (Keener and Newby, 2011; Newby and Keener, 2011; Newby, 2012), which is the simplest example of a process with internal states. In this paper, we further develop the QSA for the case where $S(t)$ and $X(t)$ are both intrinsically stochastic.

There are several advantages to the QSA. First, if the process includes a discrete state, the QSA provides an approximation that accounts for all moments of the jump propagator (infinitesimal generator), whereas the diffusion approximations include only the first two moments (e.g., a diffusion approximation of a discrete jump Markov process by truncation of a Kramers–Moyal (KM) expansion). Second, it provides a uniformly accurate approximation of the stationary probability density function. Third, physically meaningful quantities, such as the effective potential and metastable transition rates, can be generalized to processes that do not assume detailed balance. Finally, the QSA can be applied to higher dimensional (by which we mean the deterministic limit $\bar{x} \in \mathbb{R}^d, d \geq 1$) nonequilibrium processes.

The main goal of the paper is to develop the QSA for a general class of Markov processes that have a discrete internal state, and we illustrate the analysis using a simple example problem. The example problem is ideal because we can derive several approximations that serve as examples to which we apply the general QSA. Since each is an approximation of a single model, we can compare the effects of metastability in different types of Markov processes. In particular, we are interested in approximating two quantities: the timescales for metastable transitions and the effective potential. The analysis of the example problem should inform our understanding about when reduction techniques, such as a diffusion approximation, fail to approximate these two quantities and why. Previous work has shown that diffusion approximations lead to errors in both the $\alpha_i \to \infty$ (Hanggi et al., 1984; Walczak et al., 2005) and $\alpha_e \to \infty$ limits (Newby, 2012). But what happens when both noise sources are present? When is one noise source more significant than the other? If the QSS reduction fails, why does it fail? Is it due to large deviation errors like the system-size expansion, or is it
because the QSS assumption is invalid? Does it fail for the same reasons in each limit?

The paper is organized as follows. First, in Section 2, we describe in detail two related versions of \((S(t), X(t))\): one where \(X(t)\) is a discrete birth-death process and one where \(X(t)\) is a continuous Markov process. The quasi-stationary analysis is presented in Section 3. Then, in Section 4, we introduce the example problem along with various approximations and model reductions. After we apply the QSA to the example problem, results are presented in Section 5.

2. Two Markov processes with a discrete internal state

Consider the following coupled stochastic processes. Let \(S(t)\) be the discrete internal state on a finite state space having \(M\) states, and let \(X(t)\) be the external state. The main conceptual difference between an internal and external state is that the dynamics of \(S(t)\) is fast compared to \(X(t)\), and in the deterministic limit, the effect of \(S(t)\) on \(X(t)\) is in some sense “averaged out” so that only \(X(t)\) is observable.

To make the QSA as general as possible, we consider both a continuous and a discrete external state. In the case of a discrete external state, we assume conditions under which a continuous approximation is valid. The joint probability density function (probability mass function if \(X(t)\) is discrete) can be written as

\[
p(s, x, t) \Delta x = \Pr[S(t) = s, X(t) \in (x, x + \Delta x)] = \Pr[S(t) = s|X(t) \in (x, x + \Delta x)] \Pr[X(t) \in (x, x + \Delta x)].
\]

Define the conditional internal state distribution to be

\[
w(s, t|x) \equiv \Pr[s = S(t)|X(t) = x], \tag{2.1}\]

and the marginal external state density function to be

\[
u(x, t) \equiv \Pr[X(t) \in (x, x + \Delta x)]/\Delta x. \tag{2.2}\]

It is convenient to use vector notation for the probability density with

\[
p(x, t) \equiv (p(0, x, t), p(1, x, t), \ldots, p(M - 1, x, t))^T. \tag{2.3}\]

In general, we sometimes represent a given function \(f(s)\) as the vector \(f \in \mathbb{R}^M\) where the \(s\)th component of \(f\) is \(f(s)\). Diagonal matrices are written as \(\Sigma_f\), where the diagonal entries are given by the elements of the vector \(f\), and occasionally we may use the notation \(\Sigma_{f(s)}\), defined as \(\Sigma_{f(s)} = \Sigma_f\).

We write the stationary versions of (2.1)-(2.3) as

\[
\lim_{t \to \infty} p(x, t) = \hat{p}(x) = \hat{w}(x)\hat{u}(x), \tag{2.4}\]

where \(\hat{w}(x) \in \mathbb{R}^M\) and

\[
\sum_s \hat{w}(s|x) = 1. \tag{2.5}\]
For a fixed external state $X(t) = x$, the process $S(t)$ is described by a Master equation

$$\frac{\partial w}{\partial t} = \alpha_i A(x) w(t|x), \quad (2.6)$$

where $A$ is a transition rate matrix and $\alpha_i \gg 1$ is a large parameter. The matrix $A$ is a member of a family of matrices called $W$-matrices, which have the following properties. First, the columns sum to zero, which means that the matrix is singular and the vector, $\mathbf{1} \equiv (1, \cdots, 1)^T$, is the left eigenvector corresponding to a zero eigenvalue. For a transition rate matrix to be a $W$-matrix, it must have negative diagonal elements, nonnegative off-diagonal elements, and it must be irreducible. One can show, using the Perron-Frobenius theorem, that the nullspace of a $W$-matrix is one dimensional and that the right nullvector has strictly positive elements. Hence, there exists a unique $\rho > 0$ such that

$$A\rho = 0, \quad \sum_s \rho(s|x) = 1. \quad (2.7)$$

For a fixed external state, $\rho(s|x)$ is the steady state distribution of the internal state, and we refer to it as the quasi-steady-state distribution.

We call the coupled process, $(S(t), X(t))$, with $X(t)$ continuous the semi-continuous process. In this case the external state is given by the Ito stochastic differential equation,

$$dX(t) = -v(S(t), X(t))dt + \sqrt{b(S(t), X(t))} dW(t), \quad (2.8)$$

where $dW(t)$ is a Wiener process, $v(s, x)$ is the drift, $b(s, x)$ is the scaled diffusivity, and $\alpha_e \gg 1$ is a large parameter. The coupled process, $(S(t), X(t))$, is described by the CK equation

$$\frac{\partial}{\partial t} p(x, t) = \alpha_i A(x) p + \sum g(s) p, \quad (2.9)$$

where the operator, $g$, is defined by

$$g(s)p(s, x, t) \equiv -\frac{\partial}{\partial x} (v(s, x)p) + \frac{1}{2\alpha_e} \frac{\partial^2}{\partial x^2} (b(s, x)p). \quad (2.10)$$

The coupled process with $X(t)$ discrete is referred to as the discrete process. The external state is defined in terms of the birth/death process, $N(t) \in \mathbb{Z}_+$, satisfying

$$N(t) = N(0) + Y_+(\int_0^t \hat{W}_+(N(\tau)|S(\tau)) d\tau) - Y_-(\int_0^t \hat{W}_-(N(\tau)|S(\tau)) d\tau), \quad (2.11)$$

where $\hat{W}_+$ and $\hat{W}_-$ are the birth and death rates, respectively, and $Y_+(t)$ are unit Poisson processes. We assume that the rates can be written as
\[ W_\pm(n|s) = \alpha_e W_\pm(n/\alpha_e|s) \] where \( \alpha_e \gg 1 \) is a large parameter. Let \( X(t) = N(t)/\alpha_e \). Then, (2.11) can be written as
\[
X(t) = X(0) + \frac{1}{\alpha_e} Y_+(\alpha_e \int_0^t W_+(X(\tau)|S(\tau))d\tau) \\
- \frac{1}{\alpha_e} Y_-(\alpha_e \int_0^t W_-(X(\tau)|S(\tau))d\tau). \tag{2.12}
\]

The CK equation describing the discrete process is
\[
\frac{\partial}{\partial t} p(x,t) = \alpha_e A(x)p + \sum d(s)p(s|x,t), \tag{2.13}
\]
where the operator, \( d \), is defined by
\[
d(s)p(s|v_\pm(x)) \equiv \alpha_e \left[ (\alpha_x - 1)W_+(x|s)p + (\alpha_x - 1)W_-(x|s)p \right]. \tag{2.14}
\]
The jump operator,
\[ e^{\pm \alpha_x} f(x) \equiv f(x \pm 1) = \sum_{n=0}^{\infty} \frac{(-1)^n}{\gamma^n n!} f^{(n)}(x), \tag{2.15} \]
can be written in terms of a Taylor series expansion, which formally yields the Kramers–Moyal expansion of (2.13).

Note that if \( v(s,x) \) and \( b(s,x) \) are chosen appropriately, the semi-continuous process (2.8) is a diffusion approximation of the birth death process (2.12). Without loss of generality, we assume that the two external state processes are related by
\[ v(s,x) = W_+(x|s) - W_-(x|s), \quad b(s,x) = \frac{1}{2} (W_+(x|s) + W_-(x|s)). \tag{2.16} \]

Metastable behavior requires the stochastic processes to be under weak noise conditions. Generally speaking, the QSA is an asymptotic analysis where a small variable, call it \( \epsilon \), controls the global noise strength. That is, in the limit \( \epsilon \to 0 \), the stochastic process converges to a deterministic system. As discussed in the Introduction, the purpose of the two large parameters \( \alpha_i \) and \( \alpha_e \) is to place the stochastic process in weak noise conditions. In the limit \( \alpha_i \to \infty \), noise from the internal state is eliminated. Likewise, in the limit \( \alpha_e \to \infty \), noise from the external state is eliminated. Hence, both limits must be taken to reach a deterministic system. In order to carry out a systematic asymptotic analysis with a single small parameter, we define \( \epsilon = 1/\alpha_i = 1/(\phi \alpha_e) \). Hence, the limit \( \epsilon \to 0 \) is equivalent to taking the limit \( \alpha_i \to \infty, \alpha_e \to \infty \) with the ratio \( \phi = \alpha_i/\alpha_e \) fixed.

We assume for either process that the deterministic limit,
\[ \dot{x} = \bar{v}(x) \equiv \sum_s \rho(s|x)v(s,x), \tag{2.17} \]
is bistable. That is, there are three fixed points satisfying \( \tilde{v}(x) = 0 \), label them \( x_- < x < x_+ \), with \( \tilde{v}(x) > 0 \) for \( x < x_- \), \( \tilde{v}(x) < 0 \) for \( x_- < x < x_+ \), \( \tilde{v}(x) > 0 \) for \( x_+ < x < x_+ \), and \( \tilde{v}(x) < 0 \) for \( x > x_+ \). Then, \( x_\pm \) are stable fixed points and \( x_* \) is unstable.

To ensure a well-defined process, we assume for some interval \( (x_a, x_b) \), with \( x_a < x_- < x_+ < x_b \), that \( v, b \), and \( W_\pm \) are smooth functions of \( x \).

Assume further that \( W_\pm(x|s) > 0 \), for all \( 0 \leq s \leq M - 1 \) and \( x \in (x_a, x_b) \).

3. Quasi-stationary analysis

We now present a systematic perturbation method to analyze metastable, or long-time, behavior of the discrete and semi-continuous processes. Suppose we have a CK equation of the form

\[
\frac{\partial}{\partial t} p(s, x, t) = -\mathcal{L}_\epsilon p, \tag{3.1}
\]

where \( \mathcal{L}_\epsilon \) is a compact linear operator acting on functions of \((s, x)\). Note that one can easily generalize this theory to the case \( x \in \Omega \subset \mathbb{R}^N \) (see (Newby, 2012)). For illustration, take \( \mathcal{L}_\epsilon \) to have the form

\[
\mathcal{L}_\epsilon p = \frac{\partial}{\partial x} (v(s, x)p) - \epsilon \frac{\partial^2}{\partial x^2} (b(s, x)p) - \frac{1}{\epsilon} \sum_{s'} (A(s, s'|x)p(s', x, t) - A(s', s|x)p(s, x, t)), \tag{3.2}
\]

where \( \epsilon \ll 1 \) is a small parameter. (Note that we have absorbed \( \varphi \) into the definition of \( b(s, x) \).)

Assume that \( \mathcal{L}_\epsilon \) has a complete set of eigenfunctions, \( \{\phi_j(s, x)\} \) and adjoint eigenfunctions \( \{\xi_j(s, x)\} \). If the initial condition is \( p(s, x, 0) = \delta(x - x_0)\delta_{s,s_0} \), the solution can be written

\[
p(s, x, t) = \sum_{j=0}^{\infty} \xi_j(s_0, x_0)\phi_j(s, x)e^{-\lambda_j t}, \tag{3.3}
\]

where we assume that all of the eigenvalues, \( \lambda_j \), are nonnegative. Since we are interested in metastable behavior, assume that in the limit \( \epsilon \to 0 \), \( X(t) \) converges to a bistable deterministic process. Label the two stable fixed points \( x_\pm \) and the unstable fixed point \( x_* \) and assume \( x_- < x_+ < x_+ \).

The random process will look very different if the external state starts at \( x_0 < x_- \) or \( x_0 > x_+ \). For the sake of illustration assume that \( x_0 = x_- \). On intermediate time scales, the solution will converge to a stationary density around \( x_- \) that, figuratively speaking, does not see the other stable fixed point—or said another way, the solution does not see beyond \( x_+ \). Slowly, over a long timescale, the solution converges to the full stationary density as

\( \text{1 The last constraint can be relaxed somewhat provided the process converges to a unique stationary density.} \)
probability slowly leaks out past $x_*$ toward $x_+$. The timescale for this long-
time convergence is exponentially large (i.e., $O(e^{1/\epsilon})$). Since a stationary
solution exists, the smallest eigenvalue $\lambda_0$, called the principal eigenvalue,
is $\lambda_0 = 0$, and the stationary density is the eigenfunction $\phi_0(s,x)$; that is,
we normalize the principal eigenfunction so that $\langle \phi_0, 1 \rangle = 1$ with respect to
the inner product defined by

$$ \langle f(s,x), g(s,x) \rangle \equiv \int_{x \in \Omega} f(s,x)g(s,x)dx. \quad (3.4) $$

The separation of time scales in the problem can be exploited to approxi-
mate the solution. To understand how this works consider the process where
a boundary condition is placed at $x_*$ so that the process truly does not see
beyond the unstable fixed point. We want to consider two different boundary
conditions: reflecting and absorbing. To distinguish between each case, we
write the principal eigenvalue and eigenfunction (dropping the subscript) as
$\lambda^{(a)}$, $\phi^{(a)}$ and $\lambda^{(r)}$, $\phi^{(r)}$ for absorbing and reflecting boundary conditions,
respectively. If we place a reflecting boundary at $x_*$ the principal eigenvalue
$\lambda^{(r)} = 0$, but the eigenfunction $\phi^{(r)}$ is now restricted to $x \in \Omega_- = (-\infty, x_*)$
(or $x \in \Omega_+ = (x_*, \infty)$ if we instead assume that $x_0 > x_*$). We call this the
quasi-stationary density.

Now suppose that an absorbing boundary is imposed at $x_*$. In this case,
no stationary density exists, and the principal eigenvalue is perturbed by an
exponentially small amount, that is, $\lambda^{(a)} = O(e^{-C/\epsilon})$, for some $C > 0$.
The eigenfunction $\phi^{(a)}$ is also perturbed, but away from the boundary, $\phi^{(a)} \sim \phi^{(r)}$, which turns out to be straight forward to compute using a Wentzel–
Kramers–Brillouin (WKB) approximation method. Thus, if we can calculate
the eigenvalue and eigenfunction, we have an accurate approximation to the
absorbing boundary problem with

$$ p(s,x,t) \sim \phi^{(a)}(s,x)e^{-\lambda^{(a)}t}, \quad t\lambda^{(a)}_1 \gg 1, \quad (3.5) $$
or, since $\phi^{(a)} \sim \phi^{(r)}$,

$$ p(s,x,t) \sim \phi^{(r)}(s,x)e^{-\lambda^{(a)}t}, \quad t\lambda^{(a)}_1 \gg 1, \quad \epsilon \ll 1. \quad (3.6) $$

We discuss how to approximate $\lambda^{(a)}$ later in this section.

This approximation can be repeated for the initial condition $x_0 > x_+$,
and a different principal eigenvalue and quasi-stationary density are ob-
tained, call the eigenvalues $\lambda^{(a)}_\pm$ and quasi-stationary densities $\phi^{(r)}_\pm(s,x)$.
The full system, without any boundary condition imposed at $x_*$, can then
be approximated by

$$ p(s,x,t) \sim \begin{cases} 
q_-(t)\phi^{(r)}_-(s,x), & x < x_* \\
q_+(t)\phi^{(r)}_+(s,x), & x > x_*. 
\end{cases} \quad (3.7) $$
where \( q_{\pm}(t) \) satisfy the system of ordinary differential equations

\[
\frac{dq_-}{dt} = -\lambda_-^{(a)} q_- + \lambda_+^{(a)} q_+ \\
\frac{dq_+}{dt} = \lambda_-^{(a)} q_- - \lambda_+^{(a)} q_+ ,
\]

(3.8)

(3.9)

with \( q_-(0) = 1 \) and \( q_+(0) = 0 \) if \( x_0 < x_* \), or \( q_-(0) = 0 \) and \( q_+(0) = 1 \) if \( x_0 > x_* \).

A closely related problem is the mean escape time from a potential well. Define the escape time, \( \tau_{\pm} \), as the first time the process reaches \( x_* \) having started at \( x_0 = x_{\pm} \). Define the mean escape time as \( T_{\pm} = \langle \tau_{\pm} \rangle \). It follows from (3.6) that \( \tau_{\pm} \) can be approximated by an exponential random variable with mean \( T_{\pm} \sim 1/\lambda_{\pm}^{(a)} \).

To obtain an approximation of the principal eigenvalues for each well, \( \lambda_{\pm}^{(a)} \), we use a spectral projection method that makes use of the adjoint operator \( L^* \). (For simplicity, we drop the \( \pm \) notation as the following analysis applies for either potential well.) The spectral projection method was first developed for scalar-value PDE eigenvalue problems (Lee and Ward, 1995; Hinch and Chapman, 2005) and later generalized to a vector-valued PDE eigenvalue problem (Newby and Keener, 2011; Keener and Newby, 2011; Newby, 2012). The present treatment further generalizes the method.

Consider the adjoint eigenfunctions \( \{ \xi_j \} \), \( j = 0, 1, \cdots \), satisfying \( L^* \xi_j = \lambda_j \xi_j \), and take \( \langle \phi_i, \xi_j \rangle = \delta_{i,j} \) so that the two sets of eigenfunctions are biorthogonal. We use the same notation to distinguish between the two boundary conditions for the adjoint eigenfunction. If the boundary is reflecting, the first adjoint eigenfunction is \( \xi^{(r)}(x) \sim 1 \), and if the boundary is absorbing then \( \xi^{(a)} \sim \xi^{(r)} \) away from the boundary, but develops a boundary layer at \( x_* \).

Using integration by parts we have

\[
\langle \phi^{(r)}, \lambda^{(a)} \xi^{(a)} \rangle = \langle \phi^{(r)}, L^* \xi^{(a)} \rangle = \langle L^* \phi^{(r)}, \xi^{(a)} \rangle + J(\phi^{(r)}, \xi^{(a)}),
\]

(3.10)

where the boundary contribution,

\[
J(\phi^{(r)}, \xi^{(a)}) = \epsilon \sum_s b(s, x_*) \phi^{(r)}(s, x_*) \frac{d}{dx} \xi^{(a)}(s, x_*),
\]

(3.11)

is nonzero because \( \phi^{(r)} \) does not satisfy the absorbing boundary condition. Then, since \( L^* \phi^{(r)} = 0 \), the principal eigenvalue is

\[
\lambda^{(a)} = \frac{J(\phi^{(r)}, \xi^{(a)})}{\langle \phi^{(r)}, \xi^{(a)} \rangle}.
\]

(3.12)

The above identity can be used to approximate the principal eigenvalue as follows. Since away from the boundary \( x_* \), \( \xi^{(a)} \sim \xi^{(r)} = 1 \), we can make this substitution for the term in the denominator of (3.12) so that

\[
\lambda^{(a)} \sim \frac{J(\phi^{(r)}, \xi^{(a)})}{\langle \phi^{(r)}, 1 \rangle},
\]

(3.13)
Metastable behavior in Markov processes with internal states

with exponentially small error. Notice that the denominator is then well approximated by the normalization factor for the eigenfunction \( \phi^{(r)} \). We cannot make the same substitution for the term in the numerator, since (3.12) becomes a formula for \( \lambda^{(r)} = 0 \) instead of \( \lambda^{(a)} \). Of course, in some sense zero is actually a very good approximation because the error is \( O(e^{-C/\epsilon}) \), but to capture the metastable behavior we need to capture the small exponential. Note that we could have just as well used \( \phi^{(a)} \) and \( \xi^{(r)} \) in (3.10) instead of \( \phi^{(r)} \) and \( \xi^{(a)} \). We choose the latter because it simplifies the boundary layer analysis.

The recipe for approximating the solution requires approximations of the first eigenfunction and the first adjoint eigenfunction, where the latter satisfies the appropriate adjoint absorbing boundary condition. In the remainder of this section, we calculate asymptotic approximations for the two eigenfunctions and then use the results to obtain an asymptotic approximation of \( \lambda^{(a)} \). The main results are stated in Theorems 1-4.

3.1. WKB approximation of the eigenfunction \( \phi^{(r)}(s,x) \)

To simplify notation, we refer to \( \phi^{(r)}(s,x) \) as \( \phi(s,x) \), and consistent with the vector notation introduced in Section 2, we define the vector \( \phi(x) \) as having elements given by \( \phi(s,x) \), \( s = 0, 1, \cdots, M - 1 \). From the CK equations (2.9) and (2.13) it follows that the eigenfunction (up to terms exponentially small in \( \epsilon \)) satisfies

\[
\left[ A(x) + \frac{1}{\alpha_i} \Sigma w(s) \right] \phi(x) = 0, \quad (3.14)
\]

where \( w = g \) (\( w = d \)) for the semi-continuous (discrete) process with \( g \) and \( d \) defined by (2.14) and (2.10), respectively. We assume that the eigenfunction has the following WKB form

\[
\phi(x) \sim (r_0(x) + \epsilon r_1(x) + \cdots) \exp \left[ -\frac{1}{\epsilon} \Phi(x) \right], \quad (3.15)
\]

where \( \Phi \) is a scalar function and \( r_{0,1} \in \mathbb{R}^M \) (with \( r_0 \) positive). Substituting (3.15) into (3.14) and collecting leading order terms yields

\[
O(1) : \quad \left[ A(x) + \Sigma h(x,\Phi'(x)) \right] r_0(x) = 0, \quad (3.16)
\]

where,

\[
h^{\text{disc}}(s,x,p) = \frac{1}{\varphi} \left[ W_+(x|s)(e^{-\varphi p} - 1) + W_-(x|s)(e^{\varphi p} - 1) \right], \quad (3.17)
\]

for the discrete process and

\[
h^{\text{sc}}(s,x,p) = pv(s,x) + p^2 b(s,x) \quad (3.18)
\]

for the semi-continuous process. For notational convenience, we have set \( p = \varphi' \). We rewrite the remaining term in (3.16) as \( r_0(x) = k(x)\tilde{w}(x) \), where the approximation for the conditional internal state distribution (2.1)
is determined by calculating the nullspace of \( A + \Sigma_h \). Note that at fixed points, \( x_c = x_\pm, x_* \), we have that \( \hat{w}(x_c) = \rho(x_c) \), where \( \rho(x) \) is the quasi-steady-state distribution satisfying \( A\rho = 0 \). The scalar function \( k(x) \) is a normalization factor, often referred to as the pre exponential factor in the literature, and is determined at higher order. An equation for \( \Phi' \) is given by

\[
H(x, p) \equiv \det(A(x) + \Sigma_h(x, p)) = 0, \tag{3.19}
\]

where the function \( H(x, p) \) is called the Hamiltonian. Since we must have \( r_0 > 0 \), a suitable solution to (3.19) must result in a positive nullspace of \( A(x) + \Sigma_h(x, p) \).

To calculate \( k(x) \), substitute (3.15) into (3.14) and collect \( O(\epsilon) \) terms to get

\[
O(\epsilon) : \quad \left[ A + \Sigma_h(x, \Phi'(x)) \right] r_1 = \frac{dk}{dx} \Sigma_{hp} \hat{w} + k \Sigma_{hp} \frac{d\hat{w}}{dx} + k \left( \Sigma_{hp} + \frac{1}{2} \Phi''(x) \Sigma_{hp} \right) \hat{w}.
\]

While the expansion is straightforward for the semi-continuous process, it is somewhat more complicated for the discrete process. We leave the details to Appendix C. We can use a solvability condition to derive an equation for \( k(x) \) as follows. Define the left nullvector, \( l \), with

\[
l^T \left[ A + \Sigma_h(x, \Phi'(x)) \right] \hat{w} = 0.
\]

It follows from the Fredholm Alternative Theorem that \( r_1 \) exists if and only if \( k(x) \) satisfies

\[
\frac{dk}{dx} + \Psi'(x)k = 0, \tag{3.20}
\]

where

\[
\Psi'(x) = \frac{U^T(x)H_{px}(x, \Phi'(x)) + \frac{1}{2} \Phi''(x) U^T(x)H_{pp}(x, \Phi'(x))}{U^T(x)H_p(x, \Phi'(x))}, \quad x \neq x_\pm, x_*.
\]

(For more about evaluating the limit \( x \to x_c, x_c = x_\pm, x_* \), of \( \Phi''(x) \) and \( \Psi'(x) \), see Appendix D.) Hence,

\[
k(x) = \exp \left[ -\Psi(x) \right]. \tag{3.24}
\]
Theorem 1. Given a solution \( p = \Phi(x) \) to (3.19) and its integral \( \Phi(x) \), an asymptotic approximation of the solution to (3.14) is given by

\[
\phi(x) \sim \hat{w}(x) \exp \left[ -\frac{1}{\epsilon} \Phi(x) - \Psi(x) \right],
\]

where \( \Psi(x) \) is given by integration of (3.21), \( r_0 = k(x)\hat{w}(x) \) satisfies (3.16), and \( \hat{w}(x) \) satisfies (2.5).

3.2. Singular perturbation approximation of the adjoint eigenfunction \( \xi^{(a)}(s,x) \)

The WKB method used in the previous section provides only an approximation of the stationary density, not the timescale for metastable transitions (i.e., the principal eigenvalue \( \lambda^{(a)} \)). To get information about transition times we must calculate an approximation of the adjoint eigenfunction. As in Section 3.1, we simplify notation with \( \xi(s,x) = \xi^{(a)}(s,x) \) and define the vector \( \xi(x) \), having elements \( \xi(s,x) \), \( s = 0, 1, \cdots, M-1 \). The analysis for the semi-continuous and discrete processes are sufficiently different that we present each separately.

3.2.1. Semi-continuous process

Up to terms exponentially small in \( \epsilon \), the first adjoint eigenfunction satisfies

\[
\left[ A(x)^T + \epsilon \Sigma v(x) \frac{d}{dx} + \epsilon^2 \Sigma b(x) \frac{d^2}{dx^2} \right] \xi(x) = 0,
\]

along with the absorbing boundary condition,

\[
\xi(x_*) = 0.
\]

The outer solution, which does not satisfy the boundary condition, is exactly \( \xi_{\text{out}} \equiv 1 \).

To obtain an approximate solution that also satisfies boundary conditions, we must rescale \( x = x_* + \theta z \), for some \( \theta > 0 \). A reasonable first try is to take \( \theta = 1 \) so that \( x = x_* + \epsilon z \). Equation (3.26) becomes

\[
\left[ A(x^*)^T + \epsilon \Sigma v(x^*) \frac{d}{dz} + \epsilon^2 \Sigma b(x^*) \frac{d^2}{dz^2} \right] \xi_{bl}(z) = 0,
\]

where \( \xi_{bl}(z) \equiv \xi(x_* + \epsilon z) \). The solution is a linear combination of the subsolutions \( c_j \Upsilon_j e^{-\gamma_j z} \), \( j = 0, \cdots, 2M - 1 \) where

\[
(A(x^*))^T - \gamma_j \Sigma v(x^*) + \gamma_j^2 \Sigma b(x^*) \right) \Upsilon_j = 0,
\]

and \( c_j, j = 0, \cdots, 2M - 1 \) are unknown constants. We can specify the first solution as \( \Upsilon_0 = 1 \) and \( \gamma_0 = 0 \). A valid solution should be bounded in the limit \( z \to \infty \), which means that \( c_j = 0 \) if \( \gamma_j < 0 \); although we do not know a priori how many of the eigenvalues are negative. Note that the boundary condition (3.27) provides a system of \( M \) linear equations for the \( 2M \)
unknowns, $c_j$, which means that constraints to eliminate the remaining $M$
unknowns are required to close the system. One such constraint eliminates
an unknown (i.e., $c_0$) by matching to the outer solution, leaving $M - 1$
more constraints we must find. We assume that there are $M - 1$ negative
eigenvalues. For simplicity, we order the eigenvalues so that $\gamma_j < 0$ for
$j = M + 1, \ldots, 2M - 1$.

For the moment, consider the matrices in (3.29) as depending on $x$ so
that $\Upsilon_j$ and $\gamma_j$ are also functions of $x$. It is simple to show that $\Upsilon_0(x) = 1$
and $\gamma_0(x) = 0$ even if the matrices are evaluated away from $x_*$. However,
one of the solutions, label it $j = 1$, is $\gamma_1 = p(x) = \Phi'(x)$ from (3.16). Moreover, $\Upsilon_1(x) \to 1$ and $\gamma_1(x) \to 0$ as $x \to x_*$. In fact, we know that $\gamma_1(x)$ vanishes at all of the deterministic fixed points because $\Phi'(x_c) = 0$,
for $x_c = x_\pm, x_*$, and is nonzero otherwise. It follows that the zero eigenvalue
has a degenerate eigenspace, and the solution must include a secular term
involving the generalized eigenvector satisfying
\begin{equation}
A(x_*)^T \zeta = \Sigma v(x_*) 1 = v(x_*). \tag{3.30}
\end{equation}

One can show (Newby and Keener, 2011) that the deterministic fixed points
are the only points where the eigenspace associated with the zero eigenvalue
is degenerate. The solution to (3.28) is thus
\begin{equation}
\xi_{bl}(z) = c_0 1 + c_1 (\zeta - z 1) + \sum_{j=2}^{M} c_j \Upsilon_j e^{-\gamma_j z}. \tag{3.31}
\end{equation}

However, because of the secular term, the solution is unbounded in the
limit $z \to \infty$, and as a result, it cannot be matched to the outer solution.
Therefore, there is a transition layer that sits between the boundary layer
and the outer region.

To find the scaling for this transition layer, we change variables to $x = x_* + \epsilon \theta y$, for $0 < \theta < 1$, and define $\xi_\theta(y) \equiv \xi(x_* + \epsilon \theta y)$. Introduce the
asymptotic expansion
\begin{equation}
\xi_\theta(y) \sim \xi_\theta^{(0)}(y) + \epsilon \kappa \xi_\theta^{(1)}(y) + \epsilon^2 \kappa \xi_\theta^{(2)}(y), \tag{3.32}
\end{equation}
where $\kappa > 0$. Equation (3.26) becomes
\begin{equation}
\begin{bmatrix}
(A(x_*) + \epsilon^\theta y A'(x_*) + \cdots)^T \\
+ \epsilon^{1-\theta} \left( \Sigma v(x_*) + \epsilon^\theta y \Sigma' v(x_*) + \cdots \right) \frac{d}{dy} \\
+ \epsilon^{2(1-\theta)} \left( \Sigma b(x_*) + \cdots \right) \frac{d^2}{dy^2}
\end{bmatrix}
\times \left( \xi_\theta^{(0)}(y) + \epsilon \kappa \xi_\theta^{(1)}(y) + \epsilon^2 \kappa \xi_\theta^{(2)}(y) \right) = 0. \tag{3.33}
\end{equation}
Setting $\epsilon = 0$ in (3.33) yields
\begin{equation}
O(1) : \quad A(x_*)^T \xi_y^{(0)}(y) = 0,
\end{equation}
which implies that
\begin{equation}
\xi_y^{(0)}(y) = a_0(y)1,
\end{equation}
for some scalar function $a_0(y)$. The expansion (3.33) then becomes
\begin{equation}
\epsilon^k A(x_*)^T \xi_y^{(1)}(y) + \epsilon^{1-\theta} a_0(y) v(x_*) + O(\epsilon) + o(\epsilon^\kappa) + o(\epsilon^{1-\theta}) = 0,
\end{equation}
where we have used the fact that $\frac{\partial^n}{\partial x^n} A^T 1 = 0$ for all $n \geq 0$. Setting $\kappa = 1$ recovers the outer solution. The only remaining possibility is to set $\kappa = 1-\theta$, which yields
\begin{equation}
O(\epsilon^{1-\theta}) : \quad A(x_*)^T \xi_y^{(1)}(y) = -a'_0(y)v(x_*),
\end{equation}
and since $\rho(x_*)^T v(x_*) = 0$, the solution is
\begin{equation}
\xi_y^{(1)}(y) = -a'_0(y)\zeta,
\end{equation}
where $\zeta$ satisfies (3.30). Hence,
\begin{equation}
\xi_y(y) \sim a_0(y)1 - \epsilon^{1-\theta} a'_0(y)\zeta.
\end{equation}
The function $a_0(y)$ is determined at higher order; we find
\begin{equation}
\epsilon^{1-\theta} A(x_*)^T \xi_y^{(2)}(y) + \epsilon^\theta a'_0(\zeta v'(x_*) - zA'(x_*)^T \zeta)
\end{equation}
\begin{equation}
- \epsilon^{1-\theta} a''_0(\Sigma v(x_*)\zeta - b(x_*)) = 0.
\end{equation}
Setting $\theta = 1/2$ yields
\begin{equation}
O(\epsilon) : \quad A(x_*)^T \xi_y^{(2)}(y) = a''_0(y)\Sigma v(x_*)\zeta - a'_0(y)v'(x_*)_A(x_*)^T \zeta,
\end{equation}
and the resulting solvability condition is
\begin{equation}
a''_0(y) - y \left( \frac{(\rho(x_*)^T v(x_*))'}{\rho(x_*)^T (\Sigma v(x_*)\zeta - b(x_*))} \right) a'_0(y) = 0.
\end{equation}
Note that $A\rho = 0 \Rightarrow A'\rho = -A\rho'$. Furthermore, $\zeta^T A(x_*)\rho'(x_*) = v(x_*)^T \rho'(x_*)$. Hence, $\rho(x_*)^T v(x_*) - \rho(x_*)^T A'(x_*)\zeta = (\rho(x_*)^T v(x_*))' = \tilde{v}(x)'$, where $\tilde{x} = \tilde{v}(x)$ is the deterministic limit (2.17). One can show that (see Appendix A)
\begin{equation}
\frac{\tilde{v}(x)'}{\rho(x_*)^T (b(x_*) - \Sigma v(x_*)\zeta)} = -\Phi''(x_*).
\end{equation}
Assuming that $\Phi''(x_*) < 0$, the solution to (3.42) is
\begin{equation}
a'_0(y) = \tilde{c}_1 e^{\frac{1}{2} \Phi''(x_*)y^2},
\end{equation}
\begin{equation}
a_0(y) = \tilde{c}_0 + \tilde{c}_1 \int_0^y e^{\frac{1}{2} \Phi''(x_*)y'^2} dy',
\end{equation}
where $\tilde{c}_0, \tilde{c}_1$ are constants. Uniqueness of solution to (3.33) can be proven by considering the Euler-Lagrange equation and imposing the boundary condition $\xi = 0$ on the eigenfunction $\xi^0$. Note that $c_0 = \tilde{c}_0 + \tilde{c}_1 \int_0^y e^{\frac{1}{2} \Phi''(x_*)y'^2} dy'$, where $c_0$ is a constant that depends on $\xi^0$ and the boundary condition.

```plaintext

```
where $c_{0,1}$ are unknown constants. Note that since $x = x_*$ is a local maxima of $\Phi(x)$, we assume that $\Phi''(x_*) < 0$ so that $a_0(y) \to 0$ as $y \to \infty$. The solution (3.39) becomes

$$\xi_\theta(y) \sim \left( \hat{c}_0 + \hat{c}_1 \int_0^y e^{\hat{c}_1 \Phi''(x_*)y^2/\epsilon} \, dy' \right) 1 - e^{1/2 \hat{c}_1} e^{\hat{c}_1 \Phi''(x_*)y^2/\epsilon} \zeta,$$

(3.46)

which replaces the first two terms in (3.31) (i.e., $\hat{c}_0 1 + \hat{c}_1 (\zeta - z 1)$). Notice that the solution is now bounded in the limit $z \to \infty$, which allows us to match it to the outer solution; we require $\lim_{y \to \infty} \xi_\theta(z) = 1$ so that

$$\hat{c}_0 + \hat{c}_1 \int_0^\infty e^{\hat{c}_1 \Phi''(x_*)y^2} \, dy = \hat{c}_0 + \hat{c}_1 \frac{\pi}{2 |\Phi''(x_*)|} = 1,$$

(3.47)

and take

$$\hat{c}_0 = 1 - \hat{c}_1 \frac{\pi}{2 |\Phi''(x_*)|},$$

(3.48)

As $y \to 0$,

$$\xi_\theta(y) \sim (\hat{c}_0 + \hat{c}_1 y) 1 - e^{1/2 \hat{c}_1} \zeta.$$

(3.49)

This matches with (3.31) if $c_0 = \hat{c}_0$ and $c_1 = -e^{1/2 \hat{c}_1}$. The remaining unknown constants $c_j$, $j = 1, \cdots, M$, are determined using the absorbing boundary condition (3.27), resulting in the linear system of equations,

$$\hat{c}_1 \left( \frac{\pi}{2 |\Phi''(x_*)|} 1 + e^{1/2 \zeta} \right) - \sum_{j=2}^M c_j Y_j = 1.$$

(3.50)

**Theorem 2.** A uniform asymptotic approximation, valid throughout the boundary layer and transition regions, of the solution to (3.26) is given by

$$\xi(x) \sim \left[ 1 - \hat{c}_1 \left( \frac{\pi}{2 |\Phi''(x_*)|} - \int_0^{(x-x_*)/\epsilon^{1/2}} e^{\hat{c}_1 \Phi''(x_*)y^2} \, dy' \right) \right] 1$$

$$- e^{1/2 \hat{c}_1} e^{\hat{c}_1 \Phi''(x_*) (x-x_*)^2/\epsilon} \zeta + \sum_{j=2}^M c_j Y_j e^{-\gamma_j (x-x_*)/\epsilon},$$

(3.51)

where

$$\hat{c}_1 \sim \sqrt{\frac{2 |\Phi''(x_*)|}{\pi}} - e^{1/2} \sqrt{\frac{2 |\Phi''(x_*)|}{\pi}} \hat{c}_1^{(1)} + O(\epsilon),$$

(3.52)

$$c_j \sim -e^{1/2 \hat{c}_1^{(1)}} + O(\epsilon).$$

(3.53)

The constants $\hat{c}_1^{(1)}$ and $c_j^{(1)}$, $j = 2, \cdots, M$, satisfy

$$\hat{c}_1^{(1)} + \sum_{j=2}^M c_j^{(1)} Y_j = \sqrt{\frac{2 |\Phi''(x_*)|}{\pi}} \zeta.$$

(3.54)
3.2.2. Discrete process  
The adjoint eigenfunction for the discrete process satisfies
\[ (\alpha_i [A(x)]^T + \Sigma_{\alpha_i}) \xi(x) = 0, \]  
(3.55)
where
\[ d^*(s) = \alpha_i \left[ W_-(x|s)(e^{-\partial x} - 1) + W_+(x|s)(e^{\partial x} - 1) \right], \]  
(3.56)
with \( e^{\pm \partial x} \) defined by (2.15). The absorbing boundary condition is \( \xi(x_\ast) = 0 \). Once again, the outer solution is \( \xi_{\text{out}} = 1 \).

Motivated by the boundary layer analysis in Section 3.2.1, we rescale with \( x = x_\ast + \epsilon_0^\theta y \). We are interested in two cases: \( \theta = 1 \) and \( \theta = 1/2 \). In the former case, the scaling simply returns the process to a discrete variable since \( x = n_\epsilon \). Let \( \hat{n} = n - n_\ast \) and \( \xi_{\text{bl}}(\hat{n}) = \xi(x_\ast + \varphi \epsilon \hat{n}) \). Then to leading order
\[ \varphi [A(x_\ast)]^T \xi_{\text{bl}}(\hat{n}) + \Sigma_{W_-(x_\ast|s)} (\xi_{\text{bl}}(\hat{n} - 1) - \xi_{\text{bl}}(\hat{n})) + \Sigma_{W_+(x_\ast|s)} (\xi_{\text{bl}}(\hat{n} + 1) - \xi_{\text{bl}}(\hat{n})) = 0. \]  
(3.57)
Solutions have the form \( \xi_{\text{bl}}(\hat{n}) = \Gamma_j \mu_j^\theta \). Substituting this into (3.57) yields
\[ [\varphi \mu_j [A(x_\ast)]^T + \mu_j (\mu_j - 1) \Sigma_{W_+(x_\ast|s)} - (\mu_j - 1) \Sigma_{W_-(x_\ast|s)}] \Gamma_j = 0. \]  
(3.58)
As before (see (3.31)), one of the linearly independent solutions is
\[ \xi_{\text{bl}}(\hat{n}) = \zeta - \varphi \epsilon \hat{n}, \]  
(3.59)
where \( \zeta \) is given by (3.30). On the other hand, if \( \theta = 1/2 \) we recover (3.33), which means that we can replace (3.59) by (3.46). We assume that \( |\mu_j| < 1 \), \( j = 2, \cdots, M \). The boundary condition \( \xi(x_\ast) = 0 \), results in a linear system having the same form as (3.50).

**Theorem 3.** A uniform asymptotic approximation, valid throughout the boundary layer and transition regions, of the solution to (3.55) is given by
\[ \xi(x) \sim \left[ 1 - \hat{c}_1 \left( \sqrt{\frac{\pi}{2 |\Phi''(x_\ast)|}} - \int_0^{(x-x_\ast)/\epsilon_0^{1/2}} e^{\frac{1}{2} \Phi''(x_\ast) x^2} dx' \right) \right] 1 \]
\[ - \epsilon^{1/2} \hat{c}_1 e^{\frac{1}{2} \Phi''(x_\ast) (x-x_\ast)^2} \zeta + \sum_{j=2}^{M} c_j \Gamma_j \mu_j^{\frac{x-x_\ast}{\epsilon_0^{1/2}}}, \]  
(3.60)
where \( \Gamma_j \) and \( \mu_j \) satisfy (3.58). The constants \( \hat{c}_1 \) and \( c_j, j = 2, \cdots, M \), are given by (3.32) and (3.54), after substituting \( \Gamma_j \) for \( T_j \).
3.3. Principal eigenvalue

Now that we have approximations for the right and left eigenfunction, we can construct the approximation of the principal eigenvalue using the spectral projection method (see (3.13)) outlined in the introduction of this section.

**Theorem 4.** Let $\lambda^{(a)}_+$ and $\lambda^{(a)}_-$ be defined for the domain $x < x_*$ and $x > x_*$, respectively. Given the asymptotic approximation of the eigenfunction in Theorem 1 and the adjoint eigenfunction in Theorem 2 or 3, an asymptotic approximation of the principal eigenvalue is

$$
\lambda^{(a)}_+ \sim \left( \frac{B}{\pi} \sqrt{|\Phi'(x_*)|} \right) \frac{k(x_*)}{k(x_\pm)} e^{-\frac{1}{\epsilon}(\Phi(x_*) - \Phi(x_\pm))},
$$

where

$$
B = \sum s \rho(s|x_*) (b(s, x_*) - v(s, x_*)\zeta(s)),
$$

and

$$
\frac{k(x_*)}{k(x_\pm)} \equiv \exp \left[-(\Psi(x_*) - \Psi(x_\pm))\right],
$$

with $\Phi$, $\Psi$, $\rho$, and $\zeta$ defined by (3.19), (3.21), (2.7), and (3.30), respectively.

For both processes, the normalization constant (the denominator in (3.13)) is approximated using Laplace’s method with

$$
\langle \phi, \xi \rangle \sim \left(\frac{\Phi'(x_\pm)}{2\pi \epsilon}\right)^{-1/2}.
$$

The boundary contribution (the numerator in (3.13)) for each process is computed as follows.

First, for the semi-continuous process, substituting the eigenfunctions (3.25) and (3.51) into (3.11) yields

$$
J(\phi, \xi) \sim B \sqrt{\frac{2\epsilon |\Phi'(x_*)|}{\pi}} k(x_*) e^{-\frac{1}{\epsilon}\Phi(x_*)},
$$

where

$$
B \equiv \rho(x_*)^T b(x_*) - \sqrt{\frac{\pi}{2|\Phi'(x_*)|}} \sum_{j=2}^M c_j(1) \gamma_j \rho(x_*)^T \Sigma b(x_*) \gamma_j.
$$

From (3.29) we have that $\gamma_j \rho(x_*)^T \Sigma b(x_*) \gamma_j = \rho(x_*)^T \Sigma v(x_*) \gamma_j$. Then, using (3.54) and $\rho(x_*)^T v(x_*) = 0$, it follows that (3.66) can be rewritten as (3.62).

The discrete version of (3.12) can be obtained using a summation by parts argument. The resulting boundary contribution is

$$
J(\phi, \xi) = \frac{1}{2} \phi(x_*)^T \left( \Sigma_{W_+(x_*, s)} \xi(x_*) + \frac{1}{\alpha_c} \right) - \Sigma_{W_-(x_*, s)} \xi(x_*) - \frac{1}{\alpha_c}
$$

(3.67)
The first two terms in $\xi(x_{*} \pm 1/\alpha_{\varepsilon})$ (see (3.60)) can be expanded in $1/\alpha_{\varepsilon} \ll 1$ (the third term is the boundary layer solution). Substituting (3.60) and (3.25) into (3.67) (using (2.16)) shows that $J$ takes the form (3.65) with

$$B = \rho(x_{*})^{T} \left( b(x_{*}) - \frac{1}{2} \Sigma_{V(x_{*})} \xi \right)$$

$$+ \frac{1}{2\epsilon_{1}} \rho(x_{*})^{T} \left( \mu_{j}^{2} \Sigma_{W_{+}(x_{*})} - \Sigma_{W_{-}(x_{*})} \right) I_{j}. \quad (3.68)$$

From (3.58) we have that $\mu_{j}\rho(x_{*})^{T} \Sigma_{W_{+}(x_{*})} I_{j} = \rho(x_{*})^{T} \Sigma_{W_{-}(x_{*})} I_{j}$, so that

$$\frac{1}{\mu_{j}} \rho(x_{*})^{T} \left( \mu_{j}^{2} \Sigma_{W_{+}(x_{*})} - \Sigma_{W_{-}(x_{*})} \right) I_{j} = -\rho(x_{*})^{T} \Sigma_{V(x_{*})} I_{j}.$$  

Thus, we can rewrite (3.68) as (3.62).

### 4. Example: stochastic model of gene expression

Consider the following as an example of a discrete Markov process with an internal state. A population of proteins is modeled as a birth/death process, where the protein production rate depends on the internal state. The hypothetical gene responsible for producing the protein is said to be activated if an activator molecule is bound to the gene’s promotor. When the gene is activated, protein is produced at a higher rate than when it is unactivated. For simplicity we refer to “activated” and “unactivated as “on” and “off,” respectively. All parameters are presented in nondimensional form (see (Kepler and Elston, 2001) for the original dimensional version). The following state diagram, where $N_{n}$ is the state where $n$ proteins are present in the system, represents the external state transitions:

$$N_{0} \xrightarrow{\tau(S(t))} N_{1} \xleftarrow{\delta} N_{2} \cdots \xrightarrow{\tau(S(t))} N_{n} \xleftarrow{\delta} \cdots, \quad (4.1)$$

where we set $\delta = 1$. The two state stochastic process, $S(t)$, represents the on/off state of the gene: $S(t) = 1$ when the gene is on and $S(t) = 0$ when it is off. The production rate is a function of the gene state, with $\tau(0) = \sigma \alpha_{\varepsilon}$, $\tau(1) = \alpha_{\varepsilon}$. The nondimensional parameter $\sigma$ controls how much spontaneous protein production occurs when the gene is off, and we assume $0 < \sigma < 1$ so that protein production is higher when the gene is on. To get nonlinear phenomena, the internal state transitions must depend on the external state. Assume that the activator molecule is a dimer of the protein product so that the protein activates its own gene. A simple model of the gene is given by

$$\xrightarrow{\alpha_{\varepsilon} \frac{N(t)^{2}}{\alpha_{\varepsilon}^{2}}} \quad \text{(off)} \quad \xleftarrow{\alpha_{\varepsilon}^{\beta}} \quad \text{(on)}, \quad (4.2)$$
where $N(t)$ is the number of protein copies. The transition rate matrix and quasi-steady-state distribution are given by

$$A(x) \equiv \begin{bmatrix} -x^2 & \beta \\ x^2 & -\beta \end{bmatrix}, \quad \rho(x) = \begin{bmatrix} \beta \\ \beta \end{bmatrix}, \quad (4.3)$$

respectively, where $x = n/\alpha_e$. The transitions between the two gene states are assumed to be fast by specifying that $1/\alpha_i \ll 1$. Writing $p_s(n,t) = p(s,n,t)$, the CK equation (2.13) is

$$\frac{\partial}{\partial t} p_0(n,t) = \left[ (E^+ - 1)n + \alpha_e \sigma (E^- - 1) \right] p_0 + \alpha_i \left( \frac{n^2}{\alpha_e} p_0 + \beta p_1 \right) \quad (4.4a)$$

$$\frac{\partial}{\partial t} p_1(n,t) = \left[ (E^+ - 1)n + \alpha_e (E^- - 1) \right] p_1 + \alpha_i \left( \frac{n^2}{\alpha_e} p_0 - \beta p_1 \right), \quad (4.4b)$$

where the jump operators $E^\pm$ are defined by $E^\pm f(n) = f(n \pm 1)$.

A semi-continuous process (2.8) is given by applying a diffusion approximation to (4.4). The mean number of proteins when the gene is on is $\alpha_e$. When $\alpha_e \gg 1$, we can rescale to a continuous variable $X(t) = N(t)/\alpha_e$. It is straightforward to show that the drift in each state is $v(s,x)$, where

$$v(0,x) = \sigma - x, \quad v(1,x) = 1 - x, \quad (4.5)$$

and the diffusivity is $\epsilon b(s,x)$, where

$$b(0,x) = \frac{\varphi}{2} (\sigma + x), \quad b(1,x) = \frac{\varphi}{2} (1 + x). \quad (4.6)$$

Recall that in Section 2 we defined the small parameter $\epsilon = 1/\alpha_i = 1/(\varphi \alpha_e)$. The corresponding CK equation (2.9) is

$$\frac{\partial}{\partial t} p_0(x,t) = -\frac{\partial}{\partial x} [(\sigma - x)p_0] + \frac{1}{2\alpha_e} \frac{\partial^2}{\partial x^2}[(\sigma + x)p_0] - \alpha_i \left( x^2 p_0 - \beta p_1 \right) \quad (4.7a)$$

$$\frac{\partial}{\partial t} p_1(x,t) = -\frac{\partial}{\partial x} [(1 - x)p_1] + \frac{1}{2\alpha_e} \frac{\partial^2}{\partial x^2}[(1 + x)p_1] + \alpha_i \left( x^2 p_0 - \beta p_1 \right). \quad (4.7b)$$

### 4.1. Quasi-stationary analysis of the example problem

We now apply the QSA from Section 3 to the example problem. From (3.19), the equation for $\Phi'$ can be expressed as $H(x, \Phi'(x)) = 0$. In particular, for the discrete process we have

$$H_{\text{disc}}(x,p) = \frac{x^2}{\varphi^2} (e^{-\varphi p} - 1)^2 + \frac{\sigma}{\varphi^2} (e^{\varphi p} - 1)^2 + \frac{x(\sigma + 1)}{\varphi^2} (e^{-\varphi p} - 1)(e^{\varphi p} - 1)$$

$$- \frac{x}{\varphi} (\beta + x^2)(e^{-\varphi p} - 1) - \frac{1}{\varphi} (\beta \sigma + x^2)(e^{\varphi p} - 1), \quad (4.8)$$
and for the semi-continuous process
\[ \mathcal{H}_\text{sc}(x, p) = b(0, x)b(1, x)p^4 + (b(0, x)v(1, x) + b(1, x)v(0, x))p^3 \\
+ (v(0, x)v(1, x) - (\beta b(0, x) + x^2(b(1, x))))p^2 \\
- (\beta v(0, x) + x^2v(1, x))p. \] (4.9)

For the discrete problem, \( \mathcal{H}_{\text{disc}}(x, p) \) can be transformed to a cubic polynomial in \( q = e^{\varphi p} \). Then, the solutions are given by the positive real roots of
\[ \frac{\varphi^2q^2}{(q - 1)} \mathcal{H}_{\text{disc}}(x, \ln(q)) = \sigma q^3 - (x + \varphi x^2 + \sigma(1 + x + \varphi \beta))q^2 \\n+ x(1 + x + \varphi(\beta + x^2) + \sigma)q - x^2. \] (4.10)

All of the roots are real, but only one satisfies \( q = 1 \) at the deterministic fixed points. Likewise, there is a single suitable root of \( \mathcal{H}_{\text{sc}} = 0 \). Once \( \Phi' \) is calculated, the potential function \( \Phi(x) \) is computed numerically by quadrature.

The pre exponential factor is calculated using \( \Psi' \) from (3.21) with
\[ \begin{align*}
    w(x) &\equiv \left[ -\frac{h(s, x, \Phi'(x))}{h(0, x, \Phi'(x))} - \frac{h(1, x, \Phi'(x))}{h(0, x, \Phi'(x))} \right] , \\
    u(x) &\equiv 1 - \left[ \frac{h(1, x, \Phi'(x))}{h(0, x, \Phi'(x))} \right] ,
\end{align*} \] (4.11)

where
\[ \begin{align*}
    h_{\text{disc}}(s, x, p) &\equiv \frac{1}{\varphi} (e^{\varphi p} - 1) \left( v(s, 0) - xe^{-\varphi p} \right) , \\
    h_{\text{sc}}(s, x, p) &\equiv pv(s, x) + p^2b(s, x).
\end{align*} \] (4.12) (4.13)

The asymptotic approximation of the quasi-stationary density is then given by Theorem 1, and the eigenvalue approximation is given by Theorem 4, with
\[ \zeta = \frac{1 - x_*}{2\beta} \frac{\beta + 1}{\beta - 1} , \quad B = \varphi x_* + \frac{(x_* - \sigma)(1 - x_*)}{\beta + x_*^2}. \] (4.14)

4.2. QSS diffusion approximation

If jumps in \( S(t) \) are much more frequent than jumps in \( X(t) \), then \( S(t) \) is approximately stationary (i.e., distributed according to the quasi-steady-state distribution conditioned on a fixed value of \( X(t) \)). The combined process \( (S(t), X(t)) \), can be approximated by averaging out \( S(t) \) to obtain a Markov process that approximates \( X(t) \). In other words, although \( X(t) \) is not Markovian due to its dependence on \( S(t) \), it may be approximately Markovian.

2 In practice, we find that the best way of numerically integrating \( \Phi'(x) \) and \( \Psi'(x) \) is to use Chebychev approximation methods (we use the GNU Scientific Library).
A projection method results in a scalar Fokker–Planck equation for the marginal external-state probability density function

\[ u(x, t) = p_0(x, t) + p_1(x, t). \]  

(4.15)

For a general discussion of the QSS projection method see (Gardiner, 1983; Thomas et al., 2012). For brevity we only quote the result here (see (Kepler and Elston, 2001) for further details). The result is

\[ \frac{\partial u}{\partial t} = -\frac{\partial}{\partial x} \left( \langle v(x) \rangle u \right) + \frac{1}{\alpha_i} \frac{\partial^2}{\partial x^2} \left( \langle b(x) \rangle u \right) + \frac{1}{\alpha_i} \frac{\partial}{\partial x} \left( D(x) \frac{\partial u}{\partial x} \right), \]  

(4.16)

where

\[ \langle v(x) \rangle = \frac{\beta(\sigma - x)}{\beta + x^2} + \frac{x^2(1 - x)}{\beta + x^2}, \]  

(4.17)

\[ \langle b(x) \rangle = \frac{\alpha_i}{2\alpha_e} \left( \frac{\beta(\sigma + x)}{\beta + x^2} + \frac{x^2(1 + x)}{\beta + x^2} \right), \]  

(4.18)

\[ D(x) = \frac{\beta((\sigma - x) - \langle v(x) \rangle)(\sigma - x)}{(\beta + x^2)^2} + \frac{x^2((1 - x) - \langle v(x) \rangle)(1 - x)}{(\beta + x^2)^2}. \]  

(4.19)

Protein fluctuations are captured by \( \langle b(x) \rangle \) and gene fluctuations by \( D(x) \). Define the combined diffusivity

\[ B(x) = \frac{1}{\alpha_i} \left( \langle b(x) \rangle + D(x) \right). \]  

(4.20)

To make comparisons to other approximations, we define

\[ \Phi'(x) = \frac{-\langle v(x) \rangle}{\langle b(x) \rangle + D(x)}, \quad \Psi'(x) = \frac{\frac{d}{dx} \langle b(x) \rangle}{\langle b(x) \rangle + D(x)}. \]  

(4.21)

The mean escape time to reach \( x_\ast \) having starting at one of the stable fixed points, \( x_{\pm} \), can be approximated (Gardiner, 1983) by \( T_{\pm} \sim 1/\lambda_{\pm}^{(0)} \), where

\[ \lambda_{\pm}^{(0)} = \left( \frac{B(x_\ast)}{\pi} \right)^{\frac{1}{2}} \left[ R_{\pm}(x_\ast) \right] e^{-\Phi(x_\ast)} \exp \left[ -\frac{1}{\epsilon \Phi(x_\ast)} \right]. \]  

(4.22)

### 4.3. Limiting processes

The full model is given by a discrete process that is valid for any value of \( \alpha_i > 0 \) and \( \alpha_e > 0 \). The semi-continuous process is an approximation of the discrete process if \( \alpha_e \gg 1 \), and it is valid for any value of \( \alpha_i > 0 \). If we assume that \( \alpha_i \) is also a large parameter then further reduction is possible using a QSS diffusion approximation, call it the QSS process, presented in Section 4.2. All three versions contain terms that depend on \( \alpha_i \) and \( \alpha_e \), and if these parameters are assumed to be large, all three should account for contributions of noise in the internal and external state.
Fig. 1. Bifurcation diagram for the deterministic dynamics.

Of course, further model reduction is possible by removing one source of noise: either $\alpha_i \to \infty$ or $\alpha_e \to \infty$. The former is known in the literature as the adiabatic limit (see (Kepler and Elston, 2001; Walczak et al., 2005)), and the later we call the quasi-deterministic (QD) limit. If both limits are taken, a deterministic dynamical system is obtained. Note that all three versions of the example problem—the discrete, semi-continuous, and QSS processes—converge to the same deterministic limit (4.23). However, as we show is in this section, the three approximations do not necessarily converge in the adiabatic limit ($\alpha_e \to \infty$) or the QD limit ($\alpha_i \to \infty$). In the rest of this section we explore each limit in turn.

4.3.1. Deterministic limit $\alpha_i \to \infty$ and $\alpha_e \to \infty$ If we take the limit $\alpha_e \to \infty$ and $\alpha_i \to \infty$, the resulting deterministic system (2.17) is

$$\dot{x} = \frac{\beta(\sigma - x) + x^2(1 - x)}{\beta + x^2}.$$  \hfill (4.23)

Assuming that $\sigma \ll 1$, the system is described as follows. For $\beta_- < \beta < \beta_+$, where $\beta_- \sim 4\sigma + O(\sigma^2)$ and $\beta_+ \sim \frac{1}{4} + \frac{\sigma}{2} + O(\sigma^2)$, the system is bistable, with an unstable fixed point at

$$x_+ \sim \frac{1}{2}(1 - \sqrt{1 - 4\beta}) + O(\sigma)$$  \hfill (4.24)

and two stable fixed points at

$$x_- \sim \sigma + O(\sigma^2), \quad x_+ \sim \frac{1}{2}(1 + \sqrt{1 - 4\beta}) + O(\sigma)$$  \hfill (4.25)

This is the regime of interest as we wish to characterize the transition times between the two stable fixed points when the system is stochastic with weak fluctuations.
4.3.2. Quasi-deterministic limit $\alpha_e \to \infty$ A velocity jump process can be obtained from the discrete or semi-continuous process by taking the limit $\alpha_e \to \infty$ (both processes converge to the same velocity jump process). This limit is discussed in Kepler and Elston (2001) and later a metastable analysis was introduced in Newby (2012). In this limit, the CK equation converges to

$$\frac{\partial}{\partial t} p_0(x,t) = -\frac{\partial}{\partial x} \left[ (\sigma - x)p_0 - \alpha_i \left( x^2 p_0 - \beta p_1 \right) \right]$$

(4.26a)

$$\frac{\partial}{\partial t} p_1(x,t) = -\frac{\partial}{\partial x} \left[ (1 - x)p_1 + \alpha_i \left( x^2 p_0 - \beta p_1 \right) \right].$$

(4.26b)

The QSS approximation (4.16) does not converge to (4.26); instead the Fokker–Planck equation (4.16) becomes

$$\frac{\partial u}{\partial t} = -\frac{\partial}{\partial x} \langle v(x) \rangle u + \frac{1}{\alpha_i} \frac{\partial}{\partial x} \left( D(x) \frac{\partial u}{\partial x} \right),$$

(4.27)

where $\langle v(x) \rangle$ and $D(x)$ are given by (4.17) and (4.19), respectively.

A similar analysis can be carried out on the CK equation (4.26) for this process (see (Newby and Keener, 2011; Keener and Newby, 2011; Newby, 2012) for details), and Theorem 1 and 4 hold. The result is a fully analytical approximation. For $\sigma < x < 1$ we have

$$\Phi(x) = -\beta \ln(1-x) - \sigma^2 \ln(x - \sigma) - \frac{1}{2} (x - \sigma)^2 - 2\sigma (x - \sigma),$$

(4.28)

$$\hat{w}(x) = \left[ \frac{1-x}{1-\sigma} \right], \quad k(x) = \frac{1}{(x - \sigma)(1 - x)}, \quad B = \frac{(x_+ - \sigma)(1 - x_+)}{\beta + x_+^2}. \quad (4.29)$$

4.3.3. Adiabatic limit $\alpha_i \to \infty$ The CK equation for the semi-continuous process (4.7) is asymptotic to (4.16) as $\alpha_i \to \infty$. That is, the semi-continuous process converges to a fully continuous process. On the other hand, the discrete process converges to a birth/death process as $\alpha_i \to \infty$, which can be derived using a reduction procedure. The reduction procedure is based on a projection method very similar to the QSS reduction in Section 4.2. We leave the details to Appendix B and state the result. The limiting master equation is

$$\frac{du}{dt} = \alpha_e K u,$$

(4.30)

where $u(n,t) \equiv \sum_{s=0}^{\infty} p(s,n,t)$, and

$$K \equiv (E^+ - 1) \frac{n}{\alpha_e} + (E^- - 1) f(\frac{n}{\alpha_e}), \quad f(x) = \frac{\sigma \beta + x^2}{\beta + x^2}. \quad (4.31)$$

Note that at deterministic fixed points, $x_c$, we have that $f(x_c) = x_c$.

In the adiabatic limit, we have that $\hat{w}(x) = \rho(x)$ (see (2.11)). The QSA is well known for the reduced process, and Theorem 1 holds. We quote the
result here and refer the reader to (Schuss, 2010; Doering et al., 2007); that is,
\( k(x) = \frac{1}{\sqrt{x f(x)}}, \) \( B = x_*, \) and
\[
\Phi(x) = x \left( \ln \left( \frac{x}{f(x)} \right) - 1 \right) + 2 \sqrt{\beta} \left( \tan^{-1} \left( \frac{x}{\sqrt{\beta}} \right) - \sqrt{\sigma} \tan^{-1} \left( \frac{x}{\sqrt{\sigma \beta}} \right) \right).
\] (4.32)

The eigenvalue approximation is
\[
\lambda^{(n)}_\pm \sim \left( \frac{B}{\pi} \sqrt{\Phi''(x_*) \Phi''(x_\pm)} \right) k(x_*) k(x_\pm) \exp \left[ -\alpha_e (\Phi(x_*) - \Phi(x_\pm)) \right]. \quad (4.33)
\]

5. Results

In this section, we compare the approximations of the stability landscape, defined as \(-\epsilon \ln(\hat{u}(x))\), (see (2.4)) and of the mean time of a metastable transition from the minimum of one well to the other. The shape of the stability landscape can be described as a double-well potential, and in Fig. 2 it is shown for \( \sigma = 0.015 \) and two different values of the bifurcation parameter, \( \beta \), located within the region of deterministic bistability (see Section 4.3.1). The stability landscape is shown in two columns of plots, each using different parameter values. In the left column \( \beta = 0.24 \), which is near the bifurcation point that eliminates the right stability well, and in the right column \( \beta = 0.11 \), which is near the bifurcation eliminating the left stability well. Each row shows a different value of \( \epsilon \) with \( \varphi = \alpha_i / \alpha_e = 1 \) so that both noise sources are present. Approximations of the stability landscape are given by \( \Phi(x) + \epsilon \Psi(x) \), where \( \Phi \) and \( \Psi \) are defined in Section 3.1. Note that the WKB approximation of the discrete process breaks down as \( x \to 0 \) due to small copy number, requiring a boundary correction (see Appendix E). Each approximation—the QSA discrete and semi-continuous approximations, and the QSS diffusion approximation—is compared to a numerical approximation obtained by SVD decomposition in the top two rows for which \( 0 < \epsilon \ll 1 \). In the bottom row we take the limit \( \epsilon \to 0 \). Note that the SVD approximation cannot be computed for this case. First, we observe that the QSA approximation of the discrete and semi-continuous process are so close that they are indistinguishable for every parameter set. (Indeed, we find this to be the case for all of the results presented in this section). On the other hand, the QSS diffusion approximation shows significant inaccuracies, particularly in the left stability well. The most significant aspect of the stability landscape that affects metastable transitions is the height of each well in the \( \epsilon \to 0 \) limit. Although the QSS diffusion approximation does show some error in right stability well, including the height when \( \epsilon = 0 \), these differences are much less significant than the differences in the left well region. Even for the left well, the QSS diffusion approximation is not always inaccurate. Indeed, all of the approximations closely agree when \( \epsilon = 0.005 \) and \( \beta = 0.24 \) (first column, second row of Fig. 2). However, for other values of \( \epsilon \) (top and bottom row) this is clearly not the case.
Fig. 2. The stability landscape, $-\epsilon \ln(\tilde{u}(x))$, for $\varphi = 1$ and $\sigma = 0.015$. Each row shows results for a different value of $\epsilon$, and each column shows a different value of the bifurcation parameter $\beta$. The light blue curve is the quasi-steady-state approximation, the green curve is the semi-continuous QSA approximation, and the red curve (which cannot be seen beneath the green curve because both approximations are very close) is the discrete QSA approximation. For nonzero $\epsilon$ (the top four panes), the first and second order approximation $\Phi(x) + \epsilon \Psi(x)$ is compared to the value of $-\epsilon \ln(p_x)$ obtained by a numerical SVD decomposition, shown as “x” symbols. For $\epsilon = 0$ (the bottom two panes), the leading order approximation $\Phi(x)$ is shown. Note that the SVD solution can not be computed in the $\epsilon \to 0$ limit.

To examine the differences in the approximations more closely, we plot the absolute error in the stability landscape and the error in the conditional internal state distribution in Fig. 3 for the parameter values used in the left column of Fig. 2 (i.e., $\sigma = 0.015$, $\varphi = 1$, and $\beta = 0.11$). The conditional internal state distribution $\tilde{w}(x)$ is (4.11) for the discrete and semi-continuous QSA approximations and (4.3) for the QSS approximation. These are again compared to a numerical approximation obtained
Fig. 3. Absolute error. A comparison of each approximation to the numerical SVD result, for $\beta = 0.11$, $\sigma = 0.015$ shown in Fig. 2 (right column). The top row shows the error in stability landscape $-\epsilon \ln(\hat{u}(x))$ for $\epsilon = 0.01$ and $\epsilon = 0.005$. The bottom row shows the error in the internal state distribution for the same values of $\epsilon$. The colors for each curve are the same as in Fig. 2.

The approximation of the mean time for a metastable transition between wells is shown in Fig. 4. The mean escape time approximations, defined as $T_\pm \sim 1/\lambda_\pm^{(a)}$ (see (3.61)), are compared to exact Monte-Carlo (MC) simulations (using the Gillespie algorithm) for parameter values used in Fig. 2. The mean escape time is plotted on a log scale as a function of $1/\epsilon$ because $\ln(\lambda_\pm)$ is a linear function of this quantity, with a slope determined
Fig. 4. Mean exit time approximations compared to Monte-Carlo results. (a) Exit from the left well for the same parameters as used in the first column of Fig. 2. (b) Exit from the right well for the same parameters as used in the second column of Fig. 2.

by the height of the potential well in the $\epsilon \to 0$ limit (see Fig. 2 bottom row). Escape from the left well (for $\beta = 0.11$, Fig. 2 left column) is shown on the left, where the discrete and semi-continuous QSA approximations are in good agreement with MC simulations. The three approximations converge near $\epsilon = 0.005$ consistent with Fig. 2 (first column, second row).

A somewhat unexpected result is obtained for escape from the right well (corresponding to the right column of Fig. 2). All three approximations are very close, and the QSS approximation is actually more accurate for smaller values of $1/\epsilon$. The difference in the slope of each approximation is slight (see Fig. 2 right column, bottom row) and the error in the QSS approximation should grow as $1/\epsilon \to \infty$. We cannot offer a definitive explanation for the accuracy of the QSS approximation for escape from the right well. One explanation is that the QSS approximation is valid for larger values of $x$, which seems reasonable since it relies on fast transitions between internal states and the rate of transitioning from the inactive to the active internal state is proportional to $x^2$. However, this is inconsistent with the error in the conditional internal state distribution shown in Fig. 3 (bottom row), which is the key assumption underlying the QSS approximation.

Finally, we compare the mean escape time in the adiabatic limit $\alpha_i \to \infty$ and in the QD limit $\alpha_e \to \infty$. In Fig. 5, the mean time for escape from the left well is shown for $\sigma = 0.04$ and $\beta = 0.23$. In contrast to previous results, we do not fix $\varphi = \alpha_i/\alpha_e = 1$. Fig. 5 (right) illustrates that the discrete and semi-continuous approximation converge in the QD limit, and as expected, the QSS approximation error is significant. In the adiabatic limit (Fig. 5 left) the discrete and semi-continuous QSA approximations show close agreement for all values of $\alpha_i$, and as expected, all three approximations converge as $\alpha_i \to \infty$. Even though the discrete and semi-continuous QSA approximations do not converge in the adiabatic limit, the difference is very small. This suggests that a diffusion approximation for the external state—recall that we used such a procedure to derive the semi-continuous process from the full discrete process—may be valid in certain situations,
Metastable behavior in Markov processes with internal states

Fig. 5. The mean exit time for escape from the left well to the right well, with $\beta = 0.23$, $\sigma = 0.04$. Three different approximations (solid curves) are compared to Monte-Carlo simulation results (symbols). The discrete (red) and semi-continuous (green) QSA approximations are indistinguishable. Also shown is the QSS approximation (light blue). (a) The mean exit time as a function of $\alpha_c$ for fixed $\alpha_i = 333$. (b) The mean exit time as a function of $\alpha_i$ for fixed $\alpha_e = 200$.

which is interesting since diffusion approximations generally break down for metastable behavior due to large deviation errors. It is possible that the good agreement that we see for the example problem is due to the linear nature of the birth-death process governing transitions in the external state (i.e., that it is due to the simplicity of the example problem). Since the general QSA procedure presented here does not depend on this assumption, it would be interesting to see how this type of diffusion approximation behaves for a more complicated process.

A. Curvature prefactor

The purpose of this section is to show that the part of the eigenvalue estimate that contains information about the curvature of the stability well at the stable and unstable fixed point is unaffected by the QSS diffusion approximation. This is a reflection of the fact that diffusion approximations, in general, are accurate in a neighborhood of a determinstic fixed point. The eigenvalue approximation (3.61) contains a prefactor term of the form $\sqrt{|\Phi''(x^*_c)|}$ $\Phi''(x^*_\pm)$. We would like to show that, when evaluated at a fixed point, $x_c$, the second derivative of $\Phi$ for the discrete, semi-continuous, and QSS processes are all identical. We can express the second derivative in terms of $\mathcal{H}$, defined by (3.19), as follows.

Differentiating $\mathcal{H}(x, \Phi'(x)) = 0$ with respect to $x$ yields

$$\frac{d}{dx} \mathcal{H}(x, \Phi'(x)) = \mathcal{H}_x(x, \Phi'(x)) + \Phi''(x) \mathcal{H}_p(x, \Phi'(x)) = 0,$$  \hspace{1cm} (A.1)

and it follows that

$$\Phi''(x) = -\frac{\mathcal{H}_p(x, \Phi'(x))}{\mathcal{H}_p(x, \Phi'(x))}.$$ \hspace{1cm} (A.2)

However, we have that

$$\mathcal{H}_p(x_c, 0) = \mathcal{H}_x(x_c, 0) = \mathcal{H}_{xx}(x_c, 0) = 0.$$ \hspace{1cm} (A.3)
A formula valid at fixed points can be obtained as follows. Differentiating $\mathcal{H}(x, \Phi'(x)) = 0$ twice with respect to $x$ yields

$$\frac{d^2}{dx^2} \mathcal{H}(x, \Phi'(x)) = \mathcal{H}_{xx} + \Phi''(x) \mathcal{H}_{xp} + \Phi''(x) (\mathcal{H}_{px} + \mathcal{H}_{pp}) + \Phi''(x) \mathcal{H}_{p} = 0,$$

(A.4)

and it follows from (A.3) that

$$\Phi''(x_c) = -2 \frac{d^2}{dx^2} \mathcal{H}(x_c, 0).$$

(A.5)

At a fixed point, we have that $p = 0$. Expand $\mathcal{H}(x, p)$ in a Taylor series around $p = 0$. To second order in $p$, the expansion is consistent with a diffusion approximation, which always corresponds to a Hamiltonian that is quadratic in $p$ with

$$\mathcal{H}_{\text{diff}}(x, p) = a(x)p + g(x)p^2,$$

(A.6)

where $a(x)$ is the drift and $g(x)$ is the scaled diffusivity. For a QSS diffusion approximation of the processes described in Section 2, one can show that

$$a(x) = \rho(x)^T \mathbf{v}(x), \quad g(x) = \rho(x)^T \mathbf{b}(x) - \rho(x)^T (\Sigma_{\mathbf{v}(x)} - a(x) I) [A^T(x)]^T \mathbf{v}(x).$$

(A.7)

It follows that at a fixed point $a(x_c) = 0$ and $g(x_c) = \rho(x_c)^T \mathbf{b}(x) - \Sigma_{\mathbf{v}(x_c)} \xi$). Substituting (A.6) into (A.5) yields $\Phi''(x_c) = -\frac{a'(x_c)}{g(x_c)}$.

B. Adiabatic limit of the discrete process

Consider the Master equation for the probability distribution function $p_j(n, t) \equiv p(j, n, t|j_0, n_0, t_0)$. In matrix/operator form, the CK equation is

$$\frac{dp}{dt} = L_1 p + \frac{1}{\epsilon} L_2 p,$$

(B.1)

where $p(n, t) = (p_1(n, t), p_2(n, t), \ldots, p_M(n, t))^T$; $L_1 = \Sigma_{\mathbf{D}}$ is a diagonal matrix of linear operators acting on $\mathbf{n}$, each of which has a $W$-matrix representation; $L_2$ is an $M \times M W$-matrix governing the transitions between internal states, with transition rates that may depend on $\mathbf{n}$. Define the projection operator $\mathcal{P} \equiv \rho I^T$, where $L_2 \rho(n) = 0$, with $\rho(n) > 0$ and $\sum_{j=1}^M \rho_j(n) = 1$; and $\mathbf{1} \equiv (1, 1, \ldots, 1)^T$. We assume the solution has the following form

$$p(n, t) = \mathcal{P} p(n, t) + (I - \mathcal{P}) p(n, t) = u(n, t) \rho(n) + \epsilon w(n, t),$$

(B.2)

where

$$u(n, t) \equiv \mathbf{1}^T p(n, t), \quad \mathbf{1}^T w(n, t) = 0.$$

(B.3)

Applying the projection operator to both sides of (B.1) yields

$$\frac{du}{dt} \rho = \mathcal{P} L_1 (u \rho + \epsilon w).$$

(B.4)
On the other hand, applying the orthogonal projection yields
\[ \epsilon \frac{dw}{dt} - \epsilon (I - P) L_1 w = (I - P) L_1 (u \rho) + L_2 w. \] (B.5)

After setting \( \epsilon = 0 \) in the above equation we get
\[ w(n, t) \sim -L^{-1}_2 (I - P) L_1 (u(n, t) \rho(n)). \] (B.6)

Substituting (B.6) into (B.4) yields the scalar-valued operator equation for \( u(n, t) \)
\[ \frac{du}{dt} = 1^T L_1 (u \rho) - \epsilon 1^T L_1 L_2^{-1} (I - P) L_1 (u \rho). \] (B.7)

One can rewrite (B.7) in matrix form to obtain a linear system of ODEs for the vector \( u(t) \) with elements \( u_n(t) \equiv u(n, t) \)
\[ \frac{du}{dt} = W u, \] (B.8)

where \( W \equiv \sum_{j=1}^M D_j \rho_j \). In general, the reduced equation represents a Markov process only at leading order.

C. WKB/KM expansion

Consider the action of the operator \( e^{\partial x} \) on \( g(x)e^{-\alpha \hat{\phi}(x)} \) where \( g(x) \) is scalar function and \( \hat{\Phi}(x) = \varphi \Phi(x) \). We have that
\[ e^{\pm \partial x} \left( g(x)e^{-\alpha \hat{\phi}(x)} \right) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \frac{d^n}{dx^n} \left[ g(x)e^{-\alpha \hat{\phi}(x)} \right] \]
\[ = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \sum_{k=0}^{n} \binom{n}{k} g^{(n-k)}(x) \frac{d^k}{dx^k} e^{-\alpha \hat{\phi}(x)} \]
\[ = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \sum_{k=0}^{n} \binom{n}{k} g^{(n-k)}(x) \left( 1 + B_k(-\alpha \hat{\Phi}(x)) \right) e^{-\alpha \hat{\phi}(x)}, \] (C.1)

where \( B_k \) is the \( k \)th complete Bell polynomial,
\[ B_n(f(x)) \equiv \det \begin{bmatrix} f' & (n-1) f'' & (n-2) f''' & \cdots & f^{(n)} \\ -1 & f' & f'' & \cdots & f^{(n-1)} \\ 0 & -1 & f' & \cdots & f^{(n-2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 & f' \end{bmatrix}, \] (C.2)

and \( B_0 = 0 \). One can show that
\[ B_k(-\alpha \hat{\Phi}(x)) = \alpha_k^k (-\hat{\Phi}'(x))^k - \alpha_k^{k-1} \frac{k}{2} (k-1) \hat{\Phi}''(x) (-\hat{\Phi}'(x))^{k-2} + O(\alpha_k^{-2}). \] (C.3)
Expanding (C.1) in terms of $1/\alpha_c$ yields
\[ e^{\pm \alpha_c x} \left( g(x)e^{-\alpha_c \tilde{\phi}(x)} \right) = e^{\mp \tilde{\phi}(x)} \left[ g(x) - \frac{1}{\alpha_c} \left( g'(x) + \frac{1}{2}g(x)\tilde{\phi}''(x) \right) + O(\alpha_c^{-2}) \right] e^{-\alpha_c \tilde{\phi}(x)}. \]  
(C.4)

D. Evaluating $\lim_{x \to x_c} \Psi'(x)$ for $x_c = x_\pm, x_*$

Using L'Hôpital's rule, we find that
\[
\Psi'(x_c) = \left[ H_{pxx} + \frac{1}{2} \Phi''(x_c)(3H_{ppx} + \Phi''(x_c)H_{ppp}) + \frac{1}{2} \Phi'''(x_c)H_{pp} 
\right. 
\left. + l'(x_c)^T H_{px}(x_c, 0) + \frac{1}{2} \Phi''(x_c)l'(x_c)^T H_{pp}(x_c, 0) \right] \left[ l'(x_c)^T H_p(x_c, 0) + H_{px} + \Phi''(x_c)H_{pp} \right],
\]  
(D.1)

where $H(x, p)$ is defined by (3.22) and partial derivatives of $H(x, p) \equiv 1^T H(x, p)$ are evaluated at $x = x_c$ and $p = 0$, as for example,
\[ H_{xp} \equiv 1^T \frac{\partial^2}{\partial x \partial p} H(x_c, 0). \]  
(D.2)

We also have that $\Phi''(x_c)$ is given by (A.5), and
\[ \Phi'''(x_c) = -2 \frac{H_{pxx}(x_c, 0) + \frac{1}{2} \Phi''(x_c)H_{ppp}(x_c, 0)}{H_{pp}(x_c, 0)}. \]  
(D.3)

Note that $H(x, p) \neq \mathcal{H}(x, p)$, where $\mathcal{H}(x, p)$ is the Hamiltonian (3.19).

E. $x \to 0$ limit of the quasi-stationary density

The WKB approximation (3.25) of the discrete process breaks down in the limit $x \to 0$, due to small copy number effects (i.e., fluctuations are on the same order). This fact is not relevant if one is interested only in approximating the mean exit time. However, we also approximate the effective potential. Although $\Phi(x)$ is bounded in the limit $x \to 0$, $\Psi(x)$ has a logarithmic singularity. To correct this, we use the discrete master equation (4.4) to calculate $\phi(0)$, with $\phi(0) = -(\alpha_i A(0) - \alpha_e \Sigma_{i(0)})^{-1} \phi(\frac{1}{\alpha_c})$. 
References

Assaf, M., Roberts, E., Luthey-Schulten, Z.: Determining the stability of genetic switches: Explicitly accounting for mRNA noise. Phys. Rev. Lett. 106(24), 248,102 (2011). DOI 10.1103/PhysRevLett.106.248102

Biccott, D.J.: Green’s functions and first passage time distributions for dynamic instability of microtubules. Phys. Rev. E 56, 6656–6667 (1997). DOI 10.1103/PhysRevE.56.6656

Bressloff, P.C.: Metastable states and quasicycles in a stochastic wilson-cowan model of neuronal population dynamics. Phys. Rev. E 82(5), 051,903 (2010). DOI 10.1103/PhysRevE.82.051903

Bressloff, P.C.: Metastable states and quasicycles in a stochastic wilson-cowan model of neuronal population dynamics. Phys. Rev. E 82(5), 051,903 (2010). DOI 10.1103/PhysRevE.82.051903

Doering, C., Sargsyan, K., Sander, L.: Extinction times for birth-death processes: Exact results, continuum asymptotics, and the failure of the fokker-planck approximation. Multiscale Model. Simul. 3(2), 283–299 (2005). DOI 10.1137/030602800

Doering, C.R., Sargsyan, K.V., Sander, L.M., Vanden-Eijnden, E.: Asymptotics of rare events in birth–death processes bypassing the exact solutions. Journal of Physics: Condensed Matter 19(6), 065,145 (2007)

Dykman, M.I., Mori, E., Ross, J., Hunt, P.M.: Large fluctuations and optimal paths in chemical kinetics. J. Chem. Phys. 100(8), 5735–5750 (1994). DOI 10.1063/1.467139

Escudero, C., Kamenev, A.: Switching rates of multistep reactions. Phys. Rev. E 79, 041,149 (2009). DOI 10.1103/PhysRevE.79.041149

Feng, J., Kurtz, T.G.: Large deviations for stochastic processes, Mathematical surveys and monographs, vol. v. 131. American Mathematical Society (2006)

Freidlin, M.I., Wentzell, A.D.: Random Perturbations of Dynamical Systems. 2nd edition edn. Springer-Verlag, New York (1998)

Friedman, A., Craciun, G.: A model of intracellular transport of particles in an axon. J. Math. Biol. 51(2), 217–246 (2005). DOI 10.1007/s00285-004-0285-3

Gardiner, C.W.: Handbook of stochastic methods for physics, chemistry, and the natural sciences, vol. v. 13. Springer-Verlag, Berlin (1983)

Hanggi, P., Grabert, H., Talkner, P., Thomas, H.: Bistable systems: Master equation versus fokker-planck modeling. Phys. Rev. A 29(1), 371–378 (1984). DOI 10.1103/PhysRevA.29.371

Heymann, M., Vanden-Eijnden, E.: The geometric minimum action method: A least action principle on the space of curves. Communications on Pure and Applied Mathematics 61(8), 1052–1117 (2008). DOI 10.1002/cpa.20238

Hill, J.M., Hughes, B.D.: On the general random walk formulation for diffusion in media with diffusivities. The ANZIAM Journal 27, 73–87 (1985). DOI 10.1017/S033427000000477X

Hinch, R., Chapman, S.J.: Exponentially slow transitions on a markov chain: the frequency of calcium sparks. Eur. J. Appl. Math. 16(Part 4), 427–446 (2005)
Keener, J.P., Newby, J.M.: Perturbation analysis of spontaneous action potential initiation by stochastic ion channels. Phys. Rev. E 84(1), 011,918 (2011). DOI 10.1103/PhysRevE.84.011918

Kepler, T.B., Elston, T.C.: Stochasticity in transcriptional regulation: Origins, consequences, and mathematical representations. Biophys. J. 81(6), 3116–3136 (2001)

Kramli, A., Szasz, D.: Random walks with internal degrees of freedom. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 63, 85–95 (1983). DOI 10.1007/BF00534179

Landman, U., Montroll, E.W., Shlesinger, M.F.: Random walks and generalized master equations with internal degrees of freedom. Proc. Natl. Acad. Sci. U.S.A. 74(2), 430–433 (1977)

Lee, J.Y., Ward, M.J.: On the asymptotic and numerical-analyses of exponentially ill-conditioned singularly perturbed boundary-value-problems. Stud. Appl. Math. 94(3), 271–326 (1995)

Ludwig, D.: Persistence of dynamical systems under random perturbations. SIAM Review 17(4), pp. 605–640 (1975)

Maier, R.S., Stein, D.L.: Limiting exit location distributions in the stochastic exit problem. SIAM J. Appl. Math. 57(3), 752–790 (1997). DOI 10.1137/S0036139994271753

Matkowsky, B.J., Schuss, Z., Tier, C.: Diffusion across characteristic boundaries with critical points. SIAM J. Appl. Math. 43(4), 673–695 (1983). DOI 10.1137/0143046

Metzner, P., Schütte, C., Vandewalle, E.: Transition path theory for markov jump processes. Multiscale Model. Simul. 7(3), 1192–1219 (2009). DOI 10.1103/070699500

Naeh, T., Klosek, M.M., Matkowsky, B.J., Schuss, Z.: A direct approach to the exit problem. SIAM J. Appl. Math. 50(2), pp. 595–627 (1990)

Newby, J., Bressloff, P.C.: Local synaptic signaling enhances the stochastic transport of motor-driven cargo in neurons. Physical Biol. 7(3), 036,004 (2010). DOI 10.1088/1478-3975/7/3/036004

Newby, J.M.: Isolating intrinsic noise sources in a stochastic genetic switch. Physical Biology 9(2), 026,002 (2012)

Newby, J.M., Keener, J.P.: An asymptotic analysis of the spatially inhomogeneous velocity-jump process. Multiscale Model. Simul. 9(2), 735–765 (2011). DOI 10.1137/10080676X

Othmer, H., Dunbar, S., Alt, W.: Models of dispersal in biological-systems. J. Math. Biol. 26(3), 263–298 (1988)

Schuss, Z.: Theory and applications of stochastic processes: an analytical approach, Applied mathematical sciences, vol. v. 170. Springer, New York (2010)

Shwartz, A., Weiss, A.: Large deviations for performance analysis: queues, communications, and computing. Stochastic modeling series, 1st ed edn. Chapman & Hall (1995)

Talkner, P.: Mean first passage time and the lifetime of a metastable state. Zeitschrift für Physik B Condensed Matter 68, 201–207 (1987). DOI
Thattai, M., van Oudenaarden, A.: Intrinsic noise in gene regulatory networks. Proc. Natl. Acad. Sci. U.S.A. 98(15), 8614–8619 (2001). DOI 10.1073/pnas.151588598

Thomas, P., Grima, R., Straube, A.V.: Rigorous elimination of fast stochastic variables from the linear noise approximation using projection operators. Phys. Rev. E 86, 041110 (2012). DOI 10.1103/PhysRevE.86.041110

Van Kampen, N.G.: Composite stochastic-processes. Physica A 96(3), 435–453 (1979)

Vellela, M., Qian, H.: A quasistationary analysis of a stochastic chemical reaction: Keizer’s paradox. Bull. Math. Biol. 69, 1727–1746 (2007). DOI 10.1007/s11538-006-9188-3

Walczak, A.M., Onuchic, J.N., Wolynes, P.G.: Absolute rate theories of epigenetic stability. Proc. Natl. Acad. Sci. U.S.A. 102(52), 18,926–18,931 (2005). DOI 10.1073/pnas.0509547102