Journal of Statistical Physics manuscript No.
(will be inserted by the editor)

R. E. Lee DeVille · Eric Vanden-Eijnden

A nontrivial scaling limit for multiscale Markov chains

Received: date / Accepted: date

Abstract We consider Markov chains with fast and slow variables and show that in a suitable scaling limit, the dynamics becomes deterministic, yet is far away from the standard mean field approximation. This new limit is an instance of self-induced stochastic resonance which arises due to matching between a rare event timescale on the one hand and the natural timescale separation in the underlying problem on the other. Here it is illustrated on a model of a molecular motor, where it is shown to explain the regularity of the motor gait observed in some experiments.

Keywords Markov chain · scaling limit · large deviations · self-induced stochastic resonance · molecular motors

1 Introduction

Consider a Markov chain whose generator is

$$(Lf)(x) = \sum_{j=1}^{J} \lambda_j(x)(f(x+e_j) - f(x)), \tag{1}$$

The authors were partially supported by the NSF through Grants DMS02-09959 and DMS02-39625, and by the ONR via grant N00014-04-1-0565.

R. E. Lee DeVille
Courant Institute of Mathematical Sciences, New York, NY 10012
Tel.: 212-998-3218
Fax: 212-995-4121
E-mail: deville@cims.nyu.edu

Eric Vanden-Eijnden
Courant Institute of Mathematical Sciences, New York, NY 10012
Tel.: 212-998-3154
Fax: 212-995-4121
E-mail: eve2@cims.nyu.edu

where $x, e_j \in \mathbb{Z}^n$, and let $x(t)$ be a sample path of the process with this generator. If we make the jumps smaller by a factor of $\varepsilon < 1$ and simultaneously speed up the evolution by a factor of ε^{-1}, we obtain a process $z_\varepsilon(t)$ whose dynamics is governed by the generator

$$(L_\varepsilon f)(z_\varepsilon) = \varepsilon^{-1} \sum_{j=1}^{J} \bar{\lambda}_j(z_\varepsilon)(f(z_\varepsilon + \varepsilon e_j) - f(z_\varepsilon)). \tag{2}$$

where $\bar{\lambda}_j(z_\varepsilon) = \lambda_j(z_\varepsilon/\varepsilon)$. This type of scaling is suggested by various applications, in particular chemical kinetics. The dynamics of z_ε is well-understood in two limits. First consider the Markov chain on a finite time interval $[0, T]$, $T > 0$ but fixed, and let $\varepsilon \to 0$. Then Kurtz' Theorem [17] asserts that the sample paths of the Markov chain converge uniformly to $z(t)$, where

$$\dot{z}(t) = \sum_{j=1}^{J} \lambda_j(z(t)) e_j. \tag{3}$$

This Law-of-Large-Numbers-type result is referred to as the mean field limit.

If instead of fixing T, we allow $T \to \infty$ as $\varepsilon \to 0$ so that $\varepsilon \log T > 0$, i.e. look at (3) on exponentially long time intervals, then something different occurs: large deviations control the dynamics [8,17]. The system spends all of its time near the attractors of (3) and, on exponentially long timescales, jumps amongst them. For example, if we assume that the limiting equation (3) has P stable fixed points x_p, $p = 1, \dots, P$, which attract any initial condition of (3), then in the long-time limit the system effectively becomes a P-state Markov chain on these fixed points with rates typically given by

$$k^\varepsilon_{p,p'} = \nu^\varepsilon_{p,p'} \exp(-I_{p,p'}/\varepsilon) \tag{4}$$

where $I_{p,p'}$ is the rate function of large deviation for the hopping event $x_p \to x_{p'}$ and $\nu^\varepsilon_{p,p'}$ is a prefactor which is independent of ε to leading order in ε.

We describe yet another case in this paper. Specifically, we consider a scenario where the chain involves two groups of variables, one fast and one slow. Adding this additional timescale can have significant effects: we show that the interplay between the new timescale and the one associated with ϵ in (2) can (in a suitable limit) lead to dynamics which is deterministic, yet radically different from (3). It should also be stressed that this choice of scaling for a Markov chain is in no way contrived: the authors analyze a specific example of this sort of system in the context of chemical kinetics in [2], and in section 3 below we present an example of this type of scaling in a model of a molecular motor. The model we investigate is related to those developed in [5–7,12,15]; in particular it has been shown numerically in [16] that stochastic models of this type display regular behavior in certain parameter regimes (this may also explain some experimental data: see [18]).

The analysis of this paper is reminiscent of [9,10,13,1]. In those works, the authors showed that in the case of an ordinary differential equation perturbed by white noise, the small noise perturbation can lead to nontrivial yet coherent dynamics induced by the noise, a phenomena termed *self-induced* *stochastic resonance*. The present paper can be thought of as an analogue to [10] for the case of Markov chains.

The plan of the paper is as follows: in section 2 we will state Theorem 1 concerning the existence of a new scaling limit in the simplest setting which leads to nontrivial deterministic dynamics in the solution. While simple, this setting arises naturally in the context of molecular motors, as illustrated in section 3. Several generalizations are then suggested in section 4. Finally, the appendix contains the proof of Theorem 1.

2 A new scaling limit

Instead of (2), consider a Markov process on $(x,y)\in\varepsilon\mathbb{Z}\times\varepsilon\mathbb{Z}$ whose generator is

$$\begin{aligned}&(L_{\alpha,\varepsilon}f)(x,y)\\&=(\varepsilon\alpha)^{-1}\big(\lambda_+(x,y)(f(x+\varepsilon,y)-f(x,y))+\lambda_-(x,y)(f(x-\varepsilon,y)-f(x,y))\big)\\&+\varepsilon^{-1}\big(\mu_+(x,y)(f(x,y+\varepsilon)-f(x,y))+\mu_-(x,y)(f(x,y-\varepsilon)-f(x,y))\big),\end{aligned}\tag{5}$$

where $\alpha>0$ and $\varepsilon>0$ are parameters and we make the following assumption:

Assumption 1 *$\lambda_+(x,y)$ and $\lambda_-(x,y)$ are positive, smooth, and bounded functions of $(x,y)\in\mathbb{R}\times\mathbb{R}$.*

In this new chain, the rates in the x-direction can be made much faster (by choosing α small) than those in the y-direction. We investigate what happens when $\varepsilon\to 0$, $\alpha\to 0$ on specific sequences.

Letting $\varepsilon\to 0$ at fixed α, we arrive at the mean field limit (i.e. the equivalent of (3)):

$$\begin{cases}\dot{x}=\alpha^{-1}f(x,y)\\ \dot{y}=g(x,y).\end{cases}\tag{6}$$

where we have defined

$$f(x,y)=\lambda_+(x,y)-\lambda_-(x,y),\qquad g(x,y)=\mu_+(x,y)-\mu_-(x,y).\tag{7}$$

If we now take $\alpha\to 0$, so that the timescale separation becomes large as well, (6) becomes a singularly-perturbed ordinary differential equation, and we further impose the following assumptions:

Assumption 2 *For each $y\in\mathbb{R}$, there are exactly three points $x_-(y)<x_s(y)<x_+(y)$ with*

$$f(x_-(y),y)=f(x_s(y),y)=f(x_+(y),y)=0.$$

Furthermore, the curves $S_-=\{(x,y):x=x_-(y)\}$, $S_s=\{(x,y):x=x_s(y)\}$ and $S_+=\{(x,y):x=x_+(y)\}$ are smooth.

Assumption 3 *The points $x_\pm(y)$ are locally attracting in the x-direction, and $x_s(y)$ is locally repelling in the y direction, namely:*

$$\frac{\partial f}{\partial x}(x_\pm(y),y)<0,\quad \frac{\partial f}{\partial x}(x_s(y),y)>0.$$

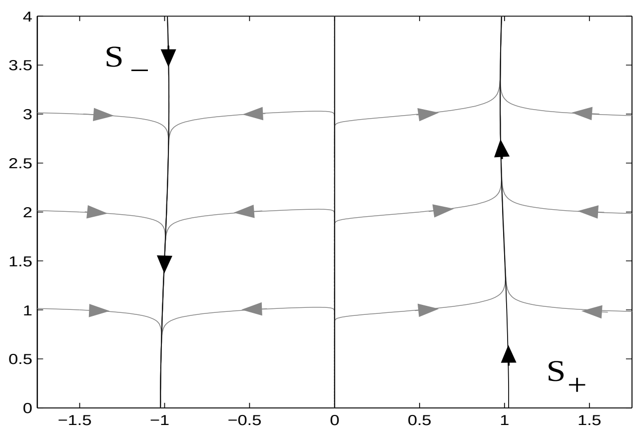

Fig. 1 Schematic representation of the flow associated with a system (6) which satisfies Assumptions 2–4. The two attracting slow manifolds, $S_\pm$ are also shown.

Assumption 4 $G_-(y) := g(x_-(y), y) < 0, \quad G_+(y) := g(x_+(y), y) > 0.$

Assumptions 2–4 mean that if $\alpha \ll 1$ in (6), there are three slow manifolds for (6); by assumption 3 the outer two are attracting and the inner one is repelling (see figure 1). Moreover, by assumption 4 the solutions of (6) move up on the right slow manifold, and down on the left one. If we consider any initial condition for (6), the trajectory is then as follows when $\alpha \ll 1$ (see Figure 1): the solution moves rapidly on a $O(\alpha)$-timescale to one or the other of the slow manifolds without changing much in y (which manifold is chosen by determining which side of $x_s(y)$ the initial condition lies on), then stays in a small neighborhood of the slow manifold and moves along it following the limiting equation [4]:

$$\dot{y} = G_+(y), \quad x = x_+(y) \quad \text{or} \quad \dot{y} = G_-(y), \quad x = x_-(y). \tag{8}$$

The presence of the noise in the Markov chain (5) may alter the picture above. Specifically, if both α and ε are sent to zero on a sequence on which α decays more quickly, the resulting speed-up in the x-direction may make the noise-induced hopping events from S_+ to S_- or S_- to S_+ as fast as the deterministic motion along the slow manifold $S_\pm$. At fixed y, the hopping events from S_- to S_+ and S_+ to S_- are large deviation events arising on the timescales $\alpha \exp(I_-(y)/\varepsilon)$ and $\alpha \exp(I_+(y)/\varepsilon)$, respectively, where $I_\pm(y)$ are the rate functions

$$\begin{aligned} I_-(y) &= \int_{x_-(y)}^{x_s(y)} \log\left(\frac{\lambda_-(x,y)}{\lambda_+(x,y)}\right) dx \qquad (S_- \to S_+), \\ I_+(y) &= \int_{x_+(y)}^{x_s(y)} \log\left(\frac{\lambda_-(x,y)}{\lambda_+(x,y)}\right) dx \qquad (S_+ \to S_-). \end{aligned} \tag{9}$$

This suggests that an interesting interplay between the motion along $S_\pm$ and the hoppings between these manifolds will arise in the distinguished limit when

$$\varepsilon \to 0, \quad \alpha \to 0, \quad \varepsilon \log \alpha^{-1} \to \beta, \tag{10}$$

where β is a parameter to be specified below. If we choose a point y_- with $I_-(y_-) = \beta$, then $\alpha \exp(I_-(y_-)/\varepsilon) \to 1$ in the distinguished limit (10); similarly, if y_+ is such that $I_+(y_+) = \beta$, then $\alpha \exp(I_+(y_+)/\varepsilon) \to 1$ in the limit (10). Points along S_- (resp. S_+) where $I_-(y) > \beta$ (resp. $I_+(y) > \beta$) are points where the rate of hopping toward S_+ (resp. S_-) is much slower than the motion along the manifold (and hence jumps should not occur in the limit (10)). Conversely, points along S_- (resp. S_+) where $I_-(y) < \beta$ (resp. $I_+(y) < \beta$) are points where the rate of hopping toward S_+ (resp. S_-) is much faster than the motion along the manifold (and hence jumps should occur in the limit (10)). This may lead to a limit cycle under the following additional assumption:

Assumption 5 *$I_-(y)$ is monotone increasing in y and $I_+(y)$ is monotone decreasing in y; both functions are bounded from above and below; and their graphs intersect, i.e.*

$$0 < m_- = \min_y I_-(y) < M_+ = \max_y I_+(y),$$
$$0 < m_+ = \min_y I_+(y) < M_- = \max_y I_-(y).$$

We will denote by $y_\star$ the (unique) point at which $I_-(y)$ and $I_+(y)$ intersect, and their value at this point by β_{max}, i.e.

$$I_+(y_\star) = I_-(y_\star) = \beta_{\mathsf{max}}.$$

We further define $\beta_{\mathsf{min}} = \max(m_-, m_+)$ and note that $\beta_{\mathsf{min}} < \beta_{\mathsf{max}} < \min(M_-, M_+)$ by construction.

Assumption 5 means that as the system moves along either of the slow manifolds, the activation energy it needs to jump across to the other decreases. Additionally, the activation energy to go from S_- to S_+ is lower than the activation energy to go from S_+ to S_- when $y < y_\star$, and conversely when $y > y_\star$. This means that, if a jump occurs from S_- to S_+ (resp. S_+ to S_-) at a point $y < y_\star$ (resp. $y > y_\star$), the probability to jump back immediately from S_+ to S_- (resp. S_- to S_+) will be very small.

We are now ready to state our main theorem. Assuming that $\beta_{\mathsf{min}} < \beta < \beta_{\mathsf{max}}$, let y_+ be the unique point such that $I_+(y_+) = \beta$, y_- the point such that $I_-(y_-) = \beta$, and define $t_\pm$ by

$$t_- = \int_{y_+}^{y_-} \frac{d\eta}{G_-(\eta)}, \quad t_+ = \int_{y_-}^{y_+} \frac{d\eta}{G_+(\eta)}.$$

Define also $(\xi(t), \eta(t))$ for $t > 0$ as follows. For $t \in [0, t_-)$, $\eta(t)$ is the solution to the second equation in (8) with initial condition $\eta(0) = y_+$ and $\xi(t) = x_-(\eta(t))$; for $t \in [t_-, t_+ + t_-)$, $\eta(t)$ is the solution to the first equation in (8) with initial condition $\eta(t_-) = y_-$ and $\xi(t) = x_+(\eta(t))$; then extend $(\xi(t), \eta(t))$ by periodicity with period $t_- + t_+$. Thus, $(\xi(t), \eta(t))$ is the periodic trajectory we obtain by following S_- from y_+ to y_-, jumping horizontally to S_+, following S_+ from y_- to y_+, jumping back to S_-, and repeating periodically.

We then have:

Theorem 1 *Let $(X^{\alpha,\varepsilon}(t), Y^{\alpha,\varepsilon}(t))$ be any realization of (5) with initial condition $(X^{\alpha,\varepsilon}(0), Y^{\alpha,\varepsilon}(0))$. Pick a $\beta \in (\beta_{\mathsf{min}}, \beta_{\mathsf{max}})$ and define $\alpha(\varepsilon)$ so that $\varepsilon \log(\alpha(\varepsilon))^{-1} = \beta$. Then, under Assumptions 1–5, there exists a phase-shift $t_\star \in [0, t_- + t_+)$ such that for any $h > 0$ and any $T > 0$,*

$$\begin{aligned} &\lim_{\varepsilon\to 0} \mathbb{P}_{x,y}\Big\{ \sup_{0\le t\le T} \Big| Y^{\alpha(\varepsilon),\varepsilon}(t) - \eta(t) \Big| > h \Big\} = 0, \\ &\lim_{\varepsilon\to 0} \mathbb{P}_{x,y}\Big\{ \int_0^T \Big| X^{\alpha(\varepsilon),\varepsilon}(t) - \xi(t) \Big| \, dt > h \Big\} = 0, \end{aligned} \tag{11}$$

where the convergence is uniform with respect to initial conditions in any compact subset of $\mathbb{R}^2$.

Remark 1 The type of convergence is different for each component of the stochastic process. As will be clear from the mechanics of the proof, unlike $Y^{\alpha(\varepsilon),\varepsilon}(t)$, which stays uniformly close to the limiting trajectory $(\xi(t), \eta(t))$ on $[0, T]$, $X^{\alpha(\varepsilon),\varepsilon}(t)$ makes excursions $O(1)$ away from the limiting trajectory many times with probability one. In fact, because of these excursions, as $\varepsilon \to 0$ the graph of the process $(X^{\alpha(\varepsilon),\varepsilon}(t), Y^{\alpha(\varepsilon),\varepsilon}(t))$ becomes dense in a finite region containing the slow manifolds whose size depends on β but not on ε. Yet, the integral estimate for $X^{\alpha(\varepsilon),\varepsilon}(t)$ in (11) holds because these excursions away from the limiting trajectory take no time in the limit.

Sketch of proof. The proof of Theorem 1 is somewhat long and technical, so we defer it to the Appendix and simply give here the main ideas of the argument.

Since we are choosing a scaling limit where $\alpha = \alpha(\varepsilon) = \exp(-\beta/\varepsilon) \ll \varepsilon$, we expect to see many steps in the x-direction before we see any steps in the y-direction. In fact, since we expect to wait $O(\varepsilon)$ time between jumps in y, we may ask: given that $Y^{\alpha(\varepsilon),\varepsilon}(t) = y$, which sites will $X^{\alpha(\varepsilon),\varepsilon}(t)$ visit in the interval $[t, t + O(\varepsilon)]$ (before the next jump in y occurs)? To answer this question, consider the "y-fixed" process generated by

$$\begin{aligned} \tilde{L}^\varepsilon_y f(x) = (\alpha(\varepsilon)\varepsilon)^{-1}\big(&\lambda_+(x,y)(f(x+\varepsilon) - f(x)) \\ &+ \lambda_-(x,y)(f(x-\varepsilon) - f(x))\big) \end{aligned} \tag{12}$$

and let us focus on what happens to the left of S_s (the picture to the right is similar). For any realization $X^\varepsilon_y(t)$ of (12), let

$$\tau^\varepsilon_{x,y}(x') = \inf\{t : X^\varepsilon_y(t) = x' \le x_s(y), X^\varepsilon_y(0) = x < x'\} \tag{13}$$

be the first exit time out of the interval $(-\infty, x']$. It is a well-known result from large deviation theory [17] that, as $\varepsilon \to 0$, $\tau^\varepsilon_{x,y}(x')$ converges towards a Poisson process with intensity[1]

$$k^\varepsilon_y(x') \asymp \exp(-\varepsilon^{-1}(I_-(x', y) - \beta)), \tag{14}$$

[1] $f(\varepsilon) \asymp g(\varepsilon)$ if $\log f(\varepsilon)/\log g(\varepsilon) \to 1$ as $\varepsilon \to 0$.

where

$$I_-(x,y) = \int_{x_-(y)}^{x} \log\left(\frac{\lambda_-(x',y)}{\lambda_+(x',y)}\right) dx'. \tag{15}$$

(The extra factor β in (14) arises because of the presence of $\alpha(\varepsilon) = e^{-\beta/\varepsilon}$ in (12).) If $y > y_-$, then by definition $I_-(x_s(y), y) > \beta$, so that as $\varepsilon \to 0$, the probability that $X_y^\varepsilon(t)$ crosses $x_s(y)$ on any time interval $[t, t+O(\varepsilon)]$ tends to zero exponentially fast in the limit as $\varepsilon \to 0$ (however notice that, with probability 1 in this limit, $X_y^\varepsilon(t)$ reaches every point to the left of $x_s(y)$ where $I_-(x_s(y), y) \le \beta$, thus preventing strong convergence of the $X^{\alpha(\varepsilon),\varepsilon}(t)$ process). In contrast, if we choose $y < y_-$, then $I_-(x_s(y), y) < \beta$ and $X_y^\varepsilon(t)$ crosses $x_s(y)$ with probability 1 in much less than $O(\varepsilon)$ time. Going back to the original process $(X^{\alpha(\varepsilon),\varepsilon}(t), Y^{\alpha(\varepsilon),\varepsilon}(t))$, this means that $X^{\alpha(\varepsilon),\varepsilon}(t)$ will cross $x_s(y)$ with probability 1 in the limit as $\varepsilon \to 0$ if $Y^{\alpha(\varepsilon),\varepsilon}(t) < y_-$. In contrast, if $Y^{\alpha(\varepsilon),\varepsilon}(t) > y_-$, $X^{\alpha(\varepsilon),\varepsilon}(t)$ will not cross $x_s(y)$ before the next jump in y occurs.

Now, to understand the motion when $Y^{\alpha(\varepsilon),\varepsilon}(t) > y_-$, we can use another estimate from large deviation theory about the y-fixed process. Consistent with the property that a jump from S_- to S_+ across S_s is exponentially unlikely if $Y^{\alpha(\varepsilon),\varepsilon}(t) > y_-$, consider the equilibrium distribution $\nu_y^\varepsilon(x)$ of the process $\tilde{L}_y^\varepsilon$ for $x < x_s(y)$ with a reflecting boundary condition at $x = x_s(y)$. It has the property that

$$\nu_y^\varepsilon(x) \asymp e^{-\varepsilon^{-1} I_-(x,y)} \quad \text{as } \varepsilon \to 0 \tag{16}$$

on any compact set which does not contain $x_s(y)$. Since $\alpha^{-1}(\varepsilon) = e^{\beta/\varepsilon} \gg 1$, the process $X^{\alpha(\varepsilon),\varepsilon}(t)$ will stay in this y-slice long enough that the occupation number of each point will be close to that of $\nu_y^\varepsilon(x)$, which, by (16), implies that $X^{\alpha,\varepsilon}(t)$ will spend most of its time near $x_-(y)$ since $I_-(x_-(y), y) = 0$ but $I_-(x,y) > 0$ if $x \le x_s(y)$ and $x \ne x_-(y)$. This in turn allows us to approximate $Y^{\alpha(\varepsilon),\varepsilon}(t)$ by the restricted process $Y_{\mathsf{res}}^\varepsilon(t)$ generated by

$$(L_{\mathsf{res}}^\varepsilon f)(y) = \varepsilon^{-1}\big(\tilde{\mu}_+(y)(f(y+\varepsilon) - f(y)) + \tilde{\mu}_-(y)(f(y-\varepsilon) - f(y))\big), \tag{17}$$

where

$$\tilde{\mu}_\pm(y) = \mu_\pm(x_-(y), y).$$

It is also clear that $Y_{\mathsf{res}}^\varepsilon(t)$ converges strongly to the solution of

$$\dot{y} = G_-(y) \tag{18}$$

by Kurtz' theorem (see Lemma 1 below). Thus, the solution of (18) allows one to approximate that of $Y^{\alpha(\varepsilon),\varepsilon}(t)$ before a jump from S_- to S_+ occurs when $Y^{\alpha(\varepsilon),\varepsilon}(t) = y_-$.

Since $\beta < \beta_{\mathsf{max}}$ by assumption, the process, once it crosses $x_s(y)$ from the left, is very unlikely to cross back at the same value of y, so a similar scenario starts over on the other side. Specifically, as long $Y^{\alpha(\varepsilon),\varepsilon}(t) < y_+$, $\dot{y} = G_+(y)$ will approximate $Y^{\alpha(\varepsilon),\varepsilon}(t)$, and as soon as $Y^{\alpha(\varepsilon),\varepsilon}(t) > y_+$, a jump from S_+ to S_- occurs with probability one. At this point, everything repeats, leading to the periodic motion as specified in Theorem 1.

3 A molecular motor example

The new limiting behavior investigated in section 2, or at least a slight modification thereof, arises in a simplified model of a molecular motor. This model is a spatial discretization of the type of Brownian ratchet models considered in [5, 15]; viewed another way, it is a one-state Kolomeisky-Fisher model with no backward steps [6, 7, 12].

In this model, the molecular motor moves by steps along one-dimensional filament referred to as the track. The motor drags behind it a massive cargo, which also moves by steps in the same direction as the motor. Here we show how the coupling to a massive cargo can give rise to regular stepping of the motor in certain parameter regimes, thereby explaining a phenomenon which has been observed in numerical experiments [16].

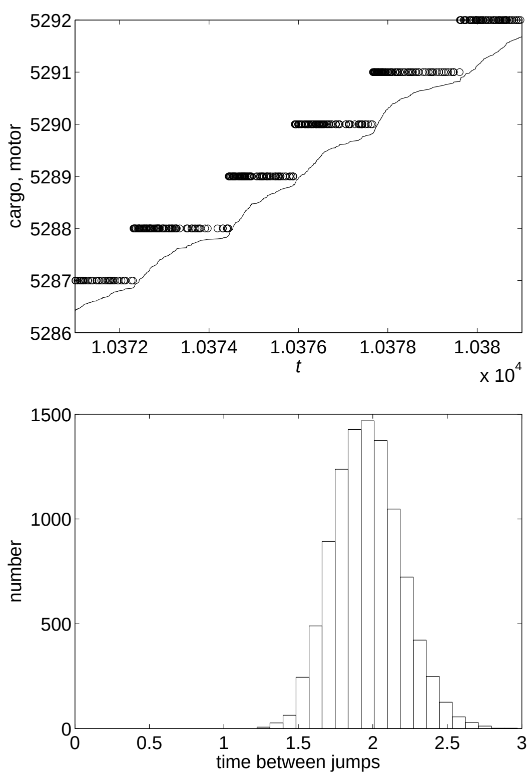

Fig. 2 One realization of (20) when $\varepsilon = 10^{-2}$, $\alpha = 10^{-6}$, $\kappa = 1.0$, $\gamma = 0.2$. The upper panel contains a trace of the trajectories; the curve represents the position of the cargo, and the dots represent the position of the motor. The lower panel is a histogram of the time between successive jumps of the motor.

Let x_M denote the position of a motor and x_C the position of a cargo. We assume that the motor is only allowed to take large steps along the one-dimensional track and we nondimensionalize these steps to be of size unity, so that $x_M \in \mathbb{Z}$. For the cargo, we take $x_C \in \varepsilon\mathbb{Z}$ and thus assume that the motion of the cargo is nearly continuous. Finally, we assume that the rates of transition depend on x_C and x_M, but only through the distance $d = x_M - x_C$. Thus, we model the motor by the Markov chain whose generator is given by

$$\begin{aligned}(L_{\alpha,\varepsilon} f)(x_M, x_C) = (\varepsilon\alpha)^{-1} \big(\lambda_+^\varepsilon(d)(f(x_M+1, x_C) - f(x_M, x_C))\big) \\ + \varepsilon^{-1}\mu_+(d)(f(x_M, x_C+\varepsilon) - f(x_M, x_C)) \\ + \varepsilon^{-1}\mu_-(d)(f(x_M, x_C-\varepsilon) - f(x_M, x_C)),\end{aligned} \tag{19}$$

where we define

$$\begin{aligned}\lambda_+^\varepsilon(d) &= e^{-d/\varepsilon}, \\ \mu_+(d) &= (\kappa+\gamma)d, \\ \mu_-(d) &= \gamma d.\end{aligned} \tag{20}$$

This model is somewhat different than the one studied in section 2, in that the size of the steps that the motor takes is independent of ε, whereas the rate $\lambda_+^\varepsilon(d)$ at which it takes these steps depends on ε (which is consistent with Kolomeisky-Fisher's model [6,7,12]). We will now show that, in the limit (10) when $\varepsilon \log \alpha^{-1} \to \beta > 0$, a regular limiting behavior emerges by a mechanism similar to the one investigated in section 2.

Letting $\alpha \to 0$ corresponds physically to making the cargo slower than the motor, which it typically is because it is much heavier than the motor. In this limit, the motion of the cargo becomes continuous and governed by

$$\dot{x}_C = -\kappa(x_C - x_M). \tag{21}$$

In this equation, x_M would be fixed if we would neglect the effect of the noise in (19) which makes the motor jump along the track. But in the limit (10), x_M is not fixed: there is a timescale matching similar to that investigated in section 2 in which the rare hopping events of the motor along the track become as fast as the motion of the cargo. For any fixed separation d, the timescale of the jumps of the motor is $\alpha e^{d/\varepsilon}$. On the other hand, the cargo moves on the $O(1)$-timescale via (21), so we expect that as long as $d > \beta$, the motor will almost certainly not jump, and as soon as $d = \beta$, it will jump.

This prediction is confirmed by the numerical results presented in Figure 2. They show that the jumping times of the motor are fairly regular as expected. Furthermore, the theory predicts that in the limit $\varepsilon \log \alpha^{-1} \to \beta$, they should arise when $d = \beta$ and the histogram Figure 2 should become a delta function. The mean and standard deviation of the jump times are

$$\mu = 1.959, \quad \sigma = 0.228,$$

giving a coefficient of variation of 11.7%. Note, moreover, that for the parameter values used, we have

$$\beta = \varepsilon \log \alpha^{-1} = 0.1382,$$

and thus the theory predicts that the cargo relaxes until the separation is $d = 0.1382$, at which time the motor jumps forward one unit making the separation $d = 1.1382$. From (21), the time of relaxation t_R it takes for x_C to relax from $x_C = x_M + 1.1382$ to $x_C = x_M + 0.1382$ is $t_R = 2.1085$, compared to $\mu = 1.959$ observed in the numerical experiments. Thus the numerically computed mean has a relative error of about 7% from the theoretically calculated value, even when the timescale separation is only two orders of magnitude.

4 Concluding remarks

To conclude, let us note that there are various ways in which the theory above can be extended. One can consider the notion of bifurcation in this context. Specifically, we can investigate what happens if, as we take the limit $\varepsilon, \alpha(\varepsilon) \to 0$, we also take $m_- \to 0$. This defines a one-parameter family of systems where the minimal activation energy is going to zero in the limit as well. One example of this was studied in [2] (see also [1] for an analogous study in the context of SDE). We consider this question in more generality in a separate paper [3].

It would also be interesting to consider what happens when the assumptions in section 2 are somewhat relaxed. For instance, interesting new phenomena may arise when there are more than two slow manifolds, or when the fast x-variable and the y-variable are multidimensional. In these situations, more complicated behaviors than the simple limit cycle found in section 2 may arise. These situations can be investigated along the same lines as what was done in the proof of Theorem 1. We shall leave these problems for future investigations.

Acknowledgements During the course of this work the authors benefited greatly from discussion with Weinan E, George Oster and Cyrill Muratov.

Appendix: Proof of Theorem 1

We define $L_y^\varepsilon = \alpha \tilde{L}_y^\varepsilon$, namely

$$L_y^\varepsilon f(x) = \varepsilon^{-1}\big(\lambda_+(x,y)(f(x+\varepsilon) - f(x)) + \lambda_-(x,y)(f(x-\varepsilon) - f(x))\big), \qquad (22)$$

where $\tilde{L}_y^\varepsilon$ is defined in (12). Recall the definition of $I_-(x', y)$ in (15). Fix y and let $X_y^\varepsilon(t)$ be any realization of (22) with $X_y^\varepsilon(0) = x$, and then define

$$\tau_{\mathsf{esc}}(x, x', y) = \inf_{t>0} \left\{t : X_y^\varepsilon(t) > x'\right\}.$$

For $\nu > 0$ define $x_\nu(y)$ to be the unique point in $(x_-(y), x_s(y))$ with

$$I_-(x_\nu(y), y) = \beta + \nu,$$

if such a point exists. (It is unique because the integrand in (15) is positive for $x \in (x_-(y), x_s(y))$, and it exists if ν is not too large.) Also define y_ν so that $x_\nu(y) = x_s(y)$. (This is unique by Assumption 5.)

Define $B_q(x) = \{z \in \mathbb{R} : |z - x| < q\}$. Choose $\delta > 0, \delta' < \delta/2$. Let $X_y^\varepsilon(t)$ be any realization of (22) with $X_y^\varepsilon(0) = x$, and define the stopping times $\tau_k(x, x', y, \delta, \delta')$ by $\tau_{-1}(x, x', y, \delta, \delta') = 0$ and

$$\tau_{2j}(x, x', y, \delta, \delta') = \inf_{t > \tau_{2j-1}} \{t : X_y^\varepsilon(t) \in \overline{B_{\delta'}(x_-(y))} \text{ or } X_y^\varepsilon(t) > x'\}, \tag{23}$$

$$\tau_{2j+1}(x, x', y, \delta, \delta') = \inf_{t > \tau_{2j}} \{t : X_y^\varepsilon(t) \notin B_\delta(x_-(y))\}. \tag{24}$$

Further define

$$a_j = \tau_{2j+1} - \tau_{2j}, \quad b_j = \tau_{2j} - \tau_{2j-1}. \tag{25}$$

We will use four lemmas which are very close to results stated in [17].

Lemma 1 (Kurtz' Theorem) *Consider the Markov process with generator*

$$L^\varepsilon f(x) = \sum_{i=1}^k \varepsilon^{-1} \lambda_i(x)(f(x + \varepsilon e_i) - f(x)),$$

where $e_i \in \mathbb{Z}^d$ and $\lambda_i \colon \Omega \subset \mathbb{R}^d \to \mathbb{R}$ are uniformly bounded and Lipschitz continuous. Let $x^\varepsilon(t)$ be any realization of this MC, and let $x^0(t)$ solve

$$\frac{d}{dt} x^0(t) = \sum_{i=1}^k \lambda_i(x^0(t)) e_i.$$

For any finite T, and for δ positive and sufficiently small, there exist $C_1, C_2 > 0$ such that for all $\varepsilon > 0$, if $x^\varepsilon(0) = x^0(0)$,

$$\mathbb{P}(\sup_{t \in [0,T]} \left| x^\varepsilon(t) - x^0(t) \right| \geq \delta) \leq C_1 e^{-C_2 \delta^2 / \varepsilon}.$$

Further, C_1, C_2 only depend on the rates λ_i as follows: if we assume that there exist $\overline{\lambda}$ and M such that

$$\sup_{i, x \in \Omega} \lambda_i(x) < \overline{\lambda}, \text{ and } |\lambda_i(x) - \lambda_i(y)| < M |x - y| \tag{26}$$

for all $x, y \in \Omega$, then C_1, C_2 need depend only on $\overline{\lambda}$ and M.

Lemma 2 *For any δ and ε positive but sufficiently small and $X_y^\varepsilon(t)$ any realization of (22) with $X_y^\varepsilon(0) \in B_\delta(x_-(y))$, we have*

$$\mathbb{P}(\varepsilon \log \tau_{\mathsf{esc}}(x, x', y) < I_-(x', y) - 3\delta) \leq 2 e^{-\delta/\varepsilon}.$$

Lemma 3 *For any δ and ε positive but sufficiently small, $\delta' < \delta/2$, there is a $T = T(\delta) < \infty$ such that for all $k \in \mathbb{Z}^+$,*

$$\mathbb{P}(a_j > kT | X_y^\varepsilon(\tau_{2j}(x, x', y, \delta, \delta')) < x') \leq (1 - \exp(-C_a \delta/\varepsilon))^k,$$
$$\mathbb{P}(b_j > kT | X_y^\varepsilon(\tau_{2j-1}(x, x', y, \delta, \delta')) < x') \leq \exp(-k C_b \delta^2/\varepsilon).$$

The constants C_a, C_b can be chosen independently of $y \in K$ (again, they depend only on uniform bounds as in (26).)

Lemma 4 *For all j odd, any positive δ, and ε positive but sufficiently small, x*

$$e^{-(I_-(x', y) + \delta)/\varepsilon} \leq \mathbb{P}(X_y^\varepsilon(\tau_{j+1}) \geq x' | X_y^\varepsilon(\tau_j) < x') \leq e^{-(I_-(x', y) - \delta)/\varepsilon}.$$

We will not prove these lemmas; in fact, they are slightly modified versions of results contained in [17]. The arguments in [17] are demonstrated without the parametric dependence on y, but one can check that all of the constants in the lemmas depend only on the bounds as in (26), and thus these lemmas apply uniformly on any compact set K. For reference, Lemmas 1, 3, 4 correspond to Theorem 5.3, Lemma 6.32, and Lemma 6.36 of [17]. Lemma 2 is not stated as such in [17] but is part of the proof of Theorem 6.17 (see p. 154).

Proof of Theorem 1. Throughout the proof, we will represent arbitrary positive constants by C_1, C_2, etc. Let T_α be any random variable with the property that

$$1 - e^{-C_1 \alpha t} \leq \mathbb{P}(T_\alpha \leq t) \leq 1 - e^{-C_2 \alpha t}. \tag{27}$$

for some $C_2 < C_1$.

Now assume that $x < x_{2\delta}(y)$. Define $X_y^0(t)$ with $X_y^0(0) = x$ and

$$\frac{d}{dt} X_y^0(t) = \lambda_+(X_y^0, y) - \lambda_-(X_y^0, y). \tag{28}$$

For all $\delta' > 0$, there exists a finite-time trajectory for (28) for which

$$X_y^0(0) = x, \quad X_y^0(T^{\mathsf{f}}) = x_-(y) + \delta'.$$

Then we have

$$\mathbb{P}(X_y^\varepsilon(t) \notin B_\delta(x_-(y)) \text{ for all } t \in [0, T^{\mathsf{f}}]) \leq C_3 \exp(-\delta^2/4\varepsilon). \tag{29}$$

To see this, note that $X_y^0(T^{\mathsf{f}}) \in B_{\delta'}(x_-(y))$, and $X_y^0(T^{\mathsf{f}}) \notin B_\delta(x_-(y))$ implies

$$\left| X_y^\varepsilon(T^{\mathsf{f}}) - X_y^0(T^{\mathsf{f}}) \right| \geq \delta/2,$$

and apply Lemma 1. Now, apply Lemma 2 with $x \in B_\delta(x_-(y))$ and $x' = x_{4\delta}(y)$. By the Markov property, if $X_y^\varepsilon(t)$ visits $B_\delta(x_-(y))$ before escaping to $x_{4\delta}(y)$, then $\tau_{\mathsf{esc}}(x, x_{4\delta}(y), y)$ is bounded as in Lemma 2. Thus, for any $x < x_{2\delta}(y)$ and $T > 0$,

$$\mathbb{P}(\tau(x, x_{4\delta}(y), y) < Te^{(\beta+\delta)/\varepsilon}) \leq 2e^{-\delta/\varepsilon} + C_3 e^{-\delta^2/4\varepsilon} \leq C_4 e^{-\delta^2/4\varepsilon}. \tag{30}$$

By the definition of T_α,

$$\mathbb{P}(T_\alpha > Te^{(\beta+\delta)/\varepsilon}) \leq C_5 e^{-C_6 \delta/\varepsilon}. \tag{31}$$

Combining (30) and (31) gives

$$\mathbb{P}(\tau(x, x_{4\delta}(y), y) < T_\alpha) \leq C_4 e^{-\delta^2/4\varepsilon} + C_5 e^{-C_6\delta/\varepsilon} < C_7 e^{-\delta^2/4\varepsilon}, \tag{32}$$

if δ is small enough. The probability of an escape to the right of $x_{4\delta}(y)$ in less than time T_α is exponentially small.

We now compute the probability of starting in any slice with initial condition less than $x_{2\delta}(y)$. Again, $x \in B_\delta(x_-(y))$ implies that

$$\mathbb{P}(\tau(x, x_{2\delta}(y), y) < Te^{(\beta+\delta)/\varepsilon}) \leq 2e^{-\delta/3\varepsilon}. \tag{33}$$

The probability of reaching $x_{2\delta}(y)$ even once before T_α is exponentially small, and therefore for any realization $X_y^\varepsilon(t)$ of (22) with $X_y^\varepsilon(0) \in B_\delta(x_-(y))$,

$$\mathbb{P}(X_y^\varepsilon(T_\alpha) \geq x_{2\delta}(y)) \leq 2e^{-\delta/3\varepsilon}. \tag{34}$$

Combining (29) with (34) gives

$$\mathbb{P}(X_y^\varepsilon(T_\alpha) \geq x_{2\delta}(y)) \leq 2e^{-\delta/3\varepsilon} + C_7 e^{-\delta^2/4\varepsilon} \tag{35}$$

for any initial condition $X_y^\varepsilon(0) < x_{2\delta}(y)$.

Summarizing what we have shown so far: consider any realization $X_y^\varepsilon(t)$ of (22) with $y > y_{4\delta}$, and let T_α satisfy (27). Then by (30) and (35), for any fixed $\delta > 0$, there exists $C_8 > 0$ with

$$\mathbb{P}(\tau_{\mathsf{esc}}(x, x_{4\delta}(y), y) < T_\alpha) + \mathbb{P}(X_y^\varepsilon(T_\alpha) > x_{2\delta}(y)) \le C_8 e^{-\delta^2/4\varepsilon}. \qquad (36)$$

Consider the process $(X^{\alpha,\varepsilon}(\alpha t), Y^{\alpha,\varepsilon}(\alpha t))$ with generator L_y^ε and define the stopping times u_i as

$$u_0 = 0, \quad u_{i+1} = \inf_{t > u_i}\{t : Y^{\alpha,\varepsilon}(\alpha t) \ne Y^{\alpha,\varepsilon}(\alpha u_i)\}.$$

Since $\mu_\pm(x,y)$ are bounded above on any compact set K, each difference $(\alpha u_{i+1} - \alpha u_i)$ is distributed as T_α in (27). Further define the random variable N by

$$N = \inf_i\{Y^{\alpha,\varepsilon}(\alpha u_i) < y_{4\delta}\}. \qquad (37)$$

Since $x_{4\delta}(y) < x_s(y)$ for all $y > y_{4\delta}$, it follows that

$$\mathbb{P}(X^{\alpha,\varepsilon}(\alpha t) > x_s(Y^{\alpha,\varepsilon}(\alpha t)) \text{ for any } t \in [0, \alpha u_N]) \le N C_8 e^{-\delta^2/4\varepsilon}. \qquad (38)$$

We will show below that $N = O(\varepsilon^{-1})$ for $\varepsilon \to 0$; the probability in (38) goes to zero as long as we can control the growth of N. To establish that $N = O(\varepsilon^{-1})$ as $\varepsilon \to 0$, we will show that the process $(X^{\alpha,\varepsilon}(\alpha t), Y^{\alpha,\varepsilon}(\alpha t))$ stays in a small neighborhood of the manifold S_-. Recalling the definitions of a_j and b_j in (25), the amount of time we spend outside of $B_\delta(x_-(y))$ is bounded above by $\sum b_j$. Lemma 3 gives

$$\mathbb{P}(b_j \ge kT) \le e^{-kC_9\delta^2/\varepsilon}.$$

Choosing T as in Lemma 3, we compute

$$\begin{aligned}
\mathbb{E}(b_j) &\le T + \sum_{k=1}^{\infty}(k+1)T\,\mathbb{P}(kT \le b_j \le (k+1)T) \\
&\le T + \sum_{k=1}^{\infty}(k+1)Te^{-kC_9\delta^2/\varepsilon} \\
&\le T(1 + 2e^{-C_9\delta^2/\varepsilon}).
\end{aligned}$$

Using the same argument as that establishing (29) gives

$$\mathbb{P}(a_j < 3T) \le e^{-C_{10}\delta^2/\varepsilon},$$

and thus

$$\mathbb{E}(a_j) \ge 3T(1 - e^{-C_{10}\delta^2/\varepsilon}),$$

by the Markov inequality. We write

$$\mathbb{E}(a_j) = A, \quad \mathbb{E}(b_j) = B,$$

and from above $A > 2B$. It is a standard result (e.g. see p. 14 of [17]) that

$$\begin{aligned}
\mathbb{P}\left(\sum_{j=1}^{J} a_j < \sum_{j=1}^{J} b_j\right) &\le \mathbb{P}\left(J^{-1}\sum_{j=1}^{J} a_j < 2A/3 \text{ and } J^{-1}\sum_{j=1}^{J} b_j > 4B/3\right) \\
&\le \exp(-C_{11}J).
\end{aligned} \qquad (39)$$

Now, define

$$J = \inf\{j : \tau_{2j} > \min(\tau_{\mathsf{esc}}(x, x_{4\delta}(y), y), T_\alpha)\}. \tag{40}$$

Then

$$\begin{aligned}\mathbb{P}(J \le \mathcal{J}) &\le \sum_{i=1}^{\mathcal{J}} \mathbb{P}(X^{\alpha,\varepsilon}(\tau_{2i}) \ge x_{4\delta}(y) \text{ and } X^{\alpha,\varepsilon}(\tau_{2i-1}) < x_{4\delta}(y)) \\ &\quad + \mathbb{P}(T_\alpha \le \tau_{2\mathcal{J}}) \\ &\le \mathcal{J} e^{-(\beta+3\delta)/\varepsilon} + e^{-C_{12}\mathcal{J}\alpha}.\end{aligned}$$

Choosing

$$\mathcal{J} = e^{(\beta+2\delta)/\varepsilon}$$

gives

$$\mathbb{P}(J \le e^{(\beta+2\delta)/\varepsilon}) \le 2e^{-\delta/\varepsilon},$$

so that we can take $J = e^{(\beta+2\delta)/\varepsilon}$ in (39) with probability exponentially close to one, and so then

$$\mathbb{P}\left(\sum_{j=1}^{J} a_j < \sum_{j=1}^{J} b_j\right) \le 3e^{-\delta/\varepsilon}.$$

Specifically, the process spends at least half of the total time inside $B_{\delta'}(x_-(y))$, and no more than half outside of $B_\delta(x_-(y))$. For any realization $X_y^\varepsilon(t)$, we denote G as

$$G = \left\{k : X_y^\varepsilon(t) \notin B_{2\delta}(x_-(y)) \text{ for some } t \in [\tau_{2k}, \tau_{2k+1}]\right\},$$

and then the proportion of time the process $X_y^\varepsilon(t)$ spends outside $B_{2\delta}(x_-(y))$ before escaping to the right at $x_{4\delta}(y)$, which we denote $\mathcal{T}_{2\delta}(x, x_{4\delta}(y), y)$, is bounded above by

$$\mathcal{T}_{2\delta}(x, x_{4\delta}(y), y) \le \sum_{k \in G} b_k.$$

Again using the arguments establishing (29),

$$p := \mathbb{P}(X_y^\varepsilon(t) \notin B_{2\delta}(x_-(y)) \text{ for some } t \in [\tau_{2k}, \tau_{2k+1}]) \le C_{13} e^{-C_{14}\delta^2/\varepsilon}, \tag{41}$$

so that

$$\mathcal{T}_{2\delta}(x, x_{4\delta}(y), y) \le \#(G)/\mathcal{J}.$$

If we take $\mathcal{J}$ Bernoulli trials, each with success p, the mean number of successes is

$$p\mathcal{J} = \mathcal{J} C_{13} e^{-C_{14}\delta^2/\varepsilon}.$$

By similar arguments to those establishing (39),

$$\mathbb{P}(\mathcal{T}_{2\delta}(x, x_{4\delta}(y), y) > 2C_{13} e^{-C_{14}\delta^2/\varepsilon}) \le e^{-C_{15}\mathcal{J}}, \tag{42}$$

and again $\mathcal{J}$ is exponentially large. In summary, the probability that the process $X_y^\varepsilon(t)$ spends more than an exponentially small amount of time outside of $B_{2\delta}(x_-(y))$ is itself exponentially small.

To complete the argument, we ignore for now every realization for which the process spends a larger ratio of time than $2C_{13}e^{-C_{14}\delta^2/\varepsilon}$ outside of $B_{2\delta}(x_-(y))$. For all other realizations,

$$\mathbb{P}(X_y^\varepsilon(u_i) \notin B_{2\delta}(x_-(y))) \le 2C_{13} e^{-C_{14}\delta^2/\varepsilon}, \text{ for each } i.$$

Recalling the definition of N as in (37),

$$\mathbb{P}(X_y^\varepsilon(u_i) \notin B_{2\delta}(x_-(y)) \text{ for any } i) \le 2NC_{13} e^{-C_{14}\delta^2/\varepsilon}.$$

As long as N is not exponentially large, then this quantity can be made as small as necessary. We again ignore all realizations for which $X_y^\varepsilon(u_i) \notin B_{2\delta}(x_-(y))$ for any i (again this removes only an exponentially-small set of realizations).

Thus in all that follows we focus only on those realizations with the property that for each i, $X_y^\varepsilon(u_i) \in B_{2\delta}(x_-(y))$.

We are now equipped to show the estimates (11). We define

$$\mu_\pm^{\mathsf{sup},\delta}(y) = \sup_{|x-x_-(y)|<2\delta} \mu_\pm(x,y), \quad \mu_\pm^{\mathsf{inf},\delta}(y) = \inf_{|x-x_-(y)|<2\delta} \mu_\pm(x,y).$$

We further define

$$(L_{\mathsf{sup}}^{\varepsilon,\delta} f)(x) = \varepsilon^{-1}\big(\mu_+^{\mathsf{sup},\delta}(y)(f(x+\varepsilon)-f(x)) + \mu_-^{\mathsf{inf},\delta}(y)(f(x-\varepsilon)-f(x))\big),$$
$$(L_{\mathsf{inf}}^{\varepsilon,\delta} f)(x) = \varepsilon^{-1}\big(\mu_+^{\mathsf{inf},\delta}(y)(f(x+\varepsilon)-f(x)) + \mu_-^{\mathsf{sup},\delta}(y)(f(x-\varepsilon)-f(x))\big),$$

and recall that $L_{\mathsf{res}}^\varepsilon$ is defined in (17). We define $Y_{\mathsf{sup}}^{\varepsilon,\delta}(t)$ (resp. $Y_{\mathsf{inf}}^{\varepsilon,\delta}(t)$, $Y_{\mathsf{res}}^\varepsilon(t)$) to be realizations of the process generated by $L_{\mathsf{sup}}^{\varepsilon,\delta}$ (resp. $L_{\mathsf{inf}}^{\varepsilon,\delta}$, $L_{\mathsf{res}}^\varepsilon$) with $Y_{\mathsf{sup}}^{\varepsilon,\delta}(0) = Y_{\mathsf{inf}}^{\varepsilon,\delta}(0) = Y_{\mathsf{res}}^\varepsilon(0) = Y^{\alpha,\varepsilon}(0)$. It is clear that for any y^*,

$$\mathbb{P}(Y^{\alpha,\varepsilon}(t) < y^*) \le \mathbb{P}(Y_{\mathsf{inf}}^{\varepsilon,\delta}(t) < y^*),$$
$$\mathbb{P}(Y^{\alpha,\varepsilon}(t) > y^*) \le \mathbb{P}(Y_{\mathsf{sup}}^{\varepsilon,\delta}(t) > y^*).$$

Define $y(t)$ to solve

$$\dot y = G_-(y), \quad y(0) = Y^{\alpha,\varepsilon}(0). \tag{43}$$

By applying Lemma 1, for any $\zeta > 0, T > 0$,

$$\lim_{\varepsilon\to 0} \mathbb{P}(\sup_{t\in[0,T]} |Y_{\mathsf{res}}^\varepsilon(t) - y(t)| > \zeta) = 0.$$

Since $\mu_\pm(x,y)$ are smooth,

$$\lim_{\delta\to 0} L_{\mathsf{sup}}^{\varepsilon,\delta} = L_{\mathsf{res}}^\varepsilon, \quad \lim_{\delta\to 0} L_{\mathsf{inf}}^{\varepsilon,\delta} = L_{\mathsf{res}}^\varepsilon$$

as operators on the space of functions defined on $\varepsilon\mathbb{Z}$. From the Trotter-Kato theorem [14], this means that for any $f\colon \varepsilon\mathbb{Z} \to \mathbb{R}$,

$$\begin{aligned}
&\lim_{\delta\to 0} \sup_{t\in[0,T]} \Big\{\mathbb{E}(f(Y_{\mathsf{inf}}^{\varepsilon,\delta}(t))|Y_{\mathsf{inf}}^{\varepsilon,\delta}(0)=y) - \mathbb{E}(f(Y_{\mathsf{res}}^\varepsilon(t))|Y_{\mathsf{res}}^\varepsilon(0)=y)\Big\} = 0,\\
&\lim_{\delta\to 0} \sup_{t\in[0,T]} \Big\{\mathbb{E}(f(Y_{\mathsf{sup}}^{\varepsilon,\delta}(t))|Y_{\mathsf{sup}}^{\varepsilon,\delta}(0)=y) - \mathbb{E}(f(Y_{\mathsf{res}}^\varepsilon(t))|Y_{\mathsf{res}}^\varepsilon(0)=y)\Big\} = 0.
\end{aligned} \tag{44}$$

Define

$$t_1 = \int_{y_-}^{Y^{\alpha,\varepsilon}(0)} \frac{d\eta}{G_-(\eta)},$$

and choose any $t^* \in [0, t_1]$. Let $y(t)$ be the solution of (43) with $y(0) = Y^{\alpha,\varepsilon}(0)$, and denote $y^* = y(t^*)$. By Lemma 1, for any $\kappa > 0, \theta > 0$ there is an $\varepsilon_0 > 0$ such that $\varepsilon < \varepsilon_0$ implies

$$\mathbb{P}(Y_{\mathsf{res}}^\varepsilon(t^*) \in [y^*-\theta, y^*+\theta]) > 1-\kappa.$$

Moreover, using (44) with $f = \mathbf{1}[y^*-\theta, y^*+\theta]$, for any $\zeta > 0$, there exists a $\delta_0 > 0$ such that $\delta < \delta_0$ implies

$$\mathbb{P}(Y_{\mathsf{sup}}^{\varepsilon,\delta}(t^*) \in [y^*-\theta, y^*+\theta]) > 1-\kappa-\zeta,$$
$$\mathbb{P}(Y_{\mathsf{inf}}^{\varepsilon,\delta}(t^*) \in [y^*-\theta, y^*+\theta]) > 1-\kappa-\zeta.$$

From the estimate on $Y_{\mathsf{sup}}^{\varepsilon,\delta}(t)$,

$$\mathbb{P}(Y^{\alpha,\varepsilon}(t^*) < y^* + \theta) > 1 - \kappa - \zeta,$$

and from the estimate on $Y_{\mathsf{inf}}^{\varepsilon,\delta}(t)$,

$$\mathbb{P}(Y^{\alpha,\varepsilon}(t^*) > y^* - \theta) > 1 - \kappa - \zeta.$$

Define $\varepsilon(\delta)$ to be the largest ε for which (30), (32), (35), (36) hold for any fixed δ. (Since these estimates hold for ε and δ sufficiently small, we can choose $\varepsilon(\delta)$ for each δ, and moreover $\varepsilon(\delta) \to 0$ as $\delta \to 0$.) Thus for any $\theta > 0$,

$$\lim_{\substack{\delta\to 0\\ \varepsilon=\varepsilon(\delta)\\ \alpha=\alpha(\varepsilon)}} \mathbb{P}\left(|Y^{\alpha,\varepsilon}(t^*) - y^*| > \theta\right) = 0. \tag{45}$$

We deduce that N as defined in (37) blows up like ε^{-1} as $\varepsilon \to 0$. Choose any initial conditions $X^{\alpha,\varepsilon}(0), Y^{\alpha,\varepsilon}(0)$ with $X^{\alpha,\varepsilon}(0) < x_s(Y^{\alpha,\varepsilon}(0))$, and again consider $y(t)$ as in (43). Combining (45) with the fact that $X_y^\varepsilon(t)$ spends an exponentially small amount of time outside of $B_{2\delta}(x_-(y))$ for any fixed δ, and that $N = O(\varepsilon^{-1})$, establishes (11) for all $t \in [0, t_1]$. The process stays near the slow manifold S_- for almost all time, and in fact $Y^{\alpha,\varepsilon}(t)$ is pathwise convergent to the underlying deterministic slow motion on the manifold S_-.

Now, assume that at some $t > 0$, $Y^{\alpha,\varepsilon}(t) = y < y_-$. Since

$$I_-(y) = I_-(x_s(y), y) < \beta,$$

then

$$\lim_{\varepsilon\to 0} \varepsilon \log \tau_{\mathsf{esc}}(x_-(y), x_s(y), y) < \beta, \tag{46}$$

and thus $\tau_{\mathsf{esc}}(x_-(y), x_s(y), y)) \ll \alpha^{-1}$ for ε small enough. From this we can conclude that in the limit, if the process $X^{\alpha,\varepsilon}(t), Y^{\alpha,\varepsilon}(t)$ ever goes below y_-, then it jumps across S_s into a neighborhood of S_+ with probability one. Arguments similar to those above apply for the process near S_+; we obtain pathwise convergence for the process near S_+. From this we get the periodic orbit described in Section 2. Where we enter this periodic orbit depends on the initial condition, and this determines the phase-shift $t_\star$. □

References

1. R. E. Lee DeVille, Cyrill B. Muratov, and Eric Vanden-Eijnden. Two distinct mechanisms of coherence in randomly perturbed dynamical systems. *Physical Review E (3)*, 72(3):031105, 10, 2005.
2. R. E. Lee DeVille, Cyrill B. Muratov, and Eric Vanden-Eijnden. Non-meanfield deterministic limits in chemical reaction kinetics far from equilibrium. *Journal of Chemical Physics*, accepted May 2005.
3. R. E. Lee Deville and Eric Vanden-Eijnden. Bifurcation theory for stochastically perturbed dynamical systems. *in preparation*, to appear 2006.
4. Wiktor Eckhaus. *Asymptotic analysis of singular perturbations*, volume 9 of *Studies in Mathematics and its Applications*. North-Holland Publishing Co., Amsterdam, 1979.
5. Timothy C. Elston and Charles S. Peskin. The role of protein flexibility in molecular motor function: coupled diffusion in a titled periodic potential. *SIAM Journal of Applied Mathematics*, 60(3):842–867, 2000.
6. Michael E. Fisher and Anatoly B. Kolomeisky. The force exerted by a molecular motor. *Proceedings of the National Academy of Sciences*, 96:6597–6601, June 1999.

7. Michael E. Fisher and Anatoly B. Kolomeisky. Simple mechanochemistry describes the dynamics of kinesin molecules. *Proceedings of the National Academy of Sciences*, 98(14):7748–7753, 2001.
8. M. I. Freidlin and A. D. Wentzell. *Random perturbations of dynamical systems.* Springer-Verlag, New York, second edition, 1998.
9. Mark Freidlin. On stochastic perturbations of dynamical systems with fast and slow components. *Stochastics and Dynamics*, 1(2):261–281, 2001.
10. Mark I. Freidlin. On stable oscillations and equilibriums induced by small noise. *Journal of Statistical Physics*, 103(1-2):283–300, 2001.
11. Iosif I. Gikhman and Anatoli V. Skorokhod. *The theory of stochastic processes. I.* Classics in Mathematics. Springer-Verlag, Berlin, 2004. Translated from the Russian by S. Kotz, Reprint of the 1974 edition.
12. Anatoly B. Kolomeisky and Michael E. Fisher. A simple kinetic model describes the processivity of myosin-V. *Biophysical Journal*, 84:1642–1650, 2003.
13. Cyrill B. Muratov, Eric Vanden-Eijnden, and Weinan E. Self-induced stochastic resonance in excitable systems. *Physica D*, 210(3-4):227–240, 2005.
14. George C. Papanicolaou. Introduction to the Asymptotic Analysis of Stochastic Equations. *Modern modeling of Continuum Phenomena,* Lectures in Applied Mathematics series volume 16, AMS, Providence, RI 1977.
15. Charles S. Peskin and George F. Oster. Coordinated hydrolysis explains the mechanical behavior of kinesin. *Biophysical Journal*, 68:202–210, 1995.
16. Maria J. Schilstra and Stephen R. Martin. Viscous load imposes a regular gait on myosin-V. *Journal of the Royal Society – Interface*, 3:153–165, 2006.
17. Adam Shwartz and Alan Weiss. *Large deviations for performance analysis.* Stochastic Modeling Series. Chapman & Hall, London, 1995. Queues, communications, and computing, With an appendix by Robert J. Vanderbei.
18. Kavitha Thirumurugan, Takeshi Sakamoto, John A. Hammer III, James R. Sellers, and Peter J. Knight. The cargo-binding domain regulates structure and activity of myosin 5. *Nature*, 442(7099):212–215, July 13 2006.