



*Citation for published version:*

Parsons, TL & Rogers, T 2017, 'Dimension reduction for stochastic dynamical systems forced onto a manifold by large drift: A constructive approach with examples from theoretical biology', *Journal of Physics A: Mathematical and Theoretical*, vol. 50, no. 41, 415601. <https://doi.org/10.1088/1751-8121/aa86c7>

*DOI:*

[10.1088/1751-8121/aa86c7](https://doi.org/10.1088/1751-8121/aa86c7)

*Publication date:*

2017

*Document Version*

Peer reviewed version

[Link to publication](https://doi.org/10.1088/1751-8121/aa86c7)

This is an author-created, un-copyedited version of an article published in *Journal of Physics A: Mathematical and Theoretical*. IOP Publishing Ltd is not responsible for any errors or omissions in this version of the manuscript or any version derived from it. The Version of Record is available online at <https://doi.org/10.1088/1751-8121/aa86c7>.

## University of Bath

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Dimension reduction for stochastic dynamical systems forced onto a manifold by large drift: a constructive approach with examples from theoretical biology

**Todd L. Parsons**

Laboratoire de Probabilités et Modèles Aléatoires, CNRS UMR 7599, Université Pierre et Marie Curie, Paris, 75005, France.

E-mail: [todd.parsons@upmc.fr](mailto:todd.parsons@upmc.fr)

**Tim Rogers**

Centre for Networks and Collective Behaviour, Department of Mathematical Sciences, University of Bath, Bath, BA2 7AY, UK.

E-mail: [t.c.rogers@bath.ac.uk](mailto:t.c.rogers@bath.ac.uk)

**Abstract.** Systems composed of large numbers of interacting agents often admit an effective coarse-grained description in terms of a multidimensional stochastic dynamical system, driven by small-amplitude intrinsic noise. In applications to biological, ecological, chemical and social dynamics it is common for these models to possess quantities that are approximately conserved on short timescales, in which case system trajectories are observed to remain close to some lower-dimensional subspace. Here, we derive explicit and general formulae for a reduced-dimension description of such processes that is exact in the limit of small noise and well-separated slow and fast dynamics. The Michaelis-Menten law of enzyme-catalysed reactions, and the link between the Lotka-Volterra and Wright-Fisher processes are explored as simple worked examples. Extensions of the method are presented for infinite dimensional systems and processes coupled to non-Gaussian noise sources.

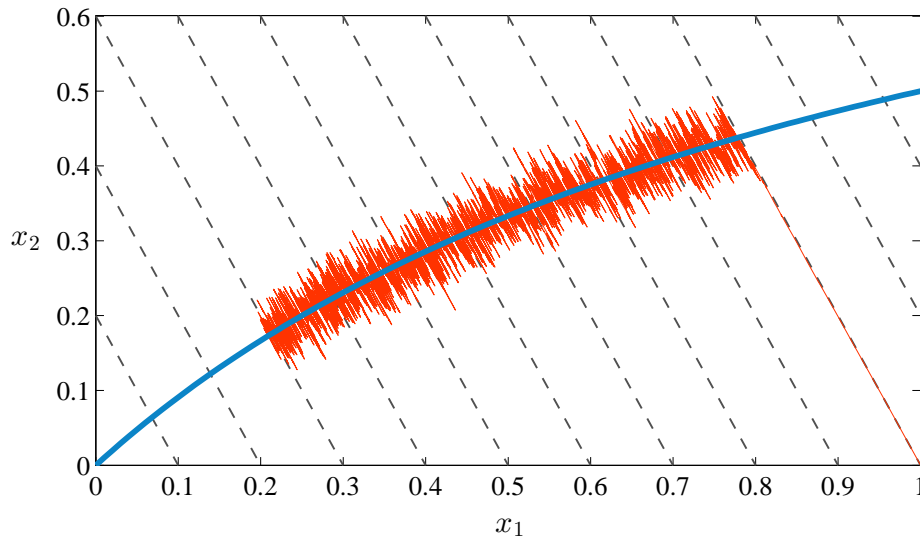
*Keywords:* Stochastic processes; Dynamical systems; Dimension reduction; Timescale separation

## 1. Introduction

To bridge the gap between observing the interactions of individuals and predicting the dynamics of whole populations is one of the core challenges of theoretical biology. Until recently, the established norm has been to take as the starting point of the analysis a continuum description for the dynamics of very large populations, usually written as a set of ordinary differential equations for population density. These equations would be motivated by assumptions such as the laws of mass-action (telling us how to derive reaction rates), and large numbers (suggesting that random fluctuations are negligible for large populations). Real populations, however, are of finite size and composed of discrete individuals whose interactions are not wholly predictable. Following an asymptotic expansion in system size, it is possible to rigorously map [1] from a microscopic description of interacting individuals to a system of stochastic differential equations that incorporate intrinsic noise arising from the discrete nature of the population and the random timing of events. It has been repeatedly shown that this noise can significantly alter the dynamics of a system at the population scale; important examples include the modelling of ecology [2, 3], epidemic spread [4, 5], and pattern formation [6, 7, 8].

Whilst the addition, or more precisely *recognition*, of stochasticity in population models can lead to richer dynamics and more relevant biological predictions, the theoretical analysis becomes far more complicated. The early development of this field focused heavily on stochastic effects present near isolated stable fixed points, where solvable linear descriptions hold true. More recently, a new direction of research into non-linear noise effects has opened up, exploiting approximation techniques based on dimension reduction. Several groups have independently found and exploited a natural separation of scales emerging in certain models of interacting populations [9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21]. Loosely speaking, it is often the case in biological models that the total size of a population adapts much more rapidly than its composition, as is evident in the disparate timescales of ecology and evolution. In dynamical systems terminology, it is observed that trajectories remain in the neighbourhood of a lower-dimensional manifold; a subspace of the system state space in which the total size of the population is a function of its composition. Intrinsic noise drives small perturbations from this manifold, which are quickly suppressed by a large deterministic drift back (see, for example, the trajectories of Michaelis-Menten dynamics in Fig. 1). The works cited above pursue various related approximation strategies, allowing for a simplified, often solvable, effective model to be derived that describes motion along the lower-dimensional manifold.

This is a kind of timescale separation that cannot be put by hand into a model as there is a complex feedback between the fast and slow degrees of freedom, which must be carefully computed. The result of system size expansion applied to an interacting population model with a separation of timescales will be an SDE for the system state  $\mathbf{x}$  that typically has the form  $\dot{\mathbf{x}} = \mathbf{f} + \varepsilon \mathbf{h} + \sqrt{\mu} \mathbf{G} \boldsymbol{\eta}(t)$ . Here  $\mathbf{f}$  describes the fast (outer)



**Figure 1. Thin Red:** Simulation of a single stochastic trajectory of an SDE of the type (1), with  $\mathbf{f}$ ,  $\mathbf{h}$  and  $\mathbf{G}$  corresponding to the Michaelis-Menten model (33, 34). **Thick Blue:** The slow manifold for this system, which the stochastic trajectory stays close to after the fast initial transient carrying it away from the initial condition (1, 0). **Dashed Black:** The flow field of the outer drift term  $\mathbf{f}$ , to which the fast motion is approximately parallel.

dynamics,  $\mathbf{h}$  the slow (inner) dynamics, and  $\boldsymbol{\eta}(t)$  the noise, with coupling matrix  $\mathbf{G}$ . The small parameters  $\varepsilon$  and  $\mu$  control the separation of timescales and strength of the noise. In this article we will consider the situation that the outer system  $\dot{\mathbf{x}} = \mathbf{f}$  has a manifold of fixed points to which the full stochastic system is attracted, and along which the slow ( $\varepsilon$ ) and noisy ( $\mu$ ) elements compete to drive the dynamics. As shown in [9]–[21], a host of surprising effects can arise from the interplay between the noise and the fast/slow dynamics.

Various theories of time scale separation in stochastic systems have been developed over decades of research in the mathematics and theoretical physics literature. In physics it has been common to work with Fokker-Planck equation, which gives a formulation of a stochastic processes in terms of the PDE for the evolution of probability density. Timescale separation in this setting amounts to integrating out one or more degrees of freedom from the PDE to reduce its dimension, see [22] for an introduction. Physicists might understand this process through its natural analogue in quantum mechanics, the Born-Oppenheimer approximation [23]. In most applications the Fokker-Planck equation will not be exactly separable, necessitating the application by hand of a carefully chosen projection operator, an approach going back to the work of Zwanzig [24]. Alternatively, working directly with the SDE description, stochastic versions of centre manifold theory [25] and local normal forms [26, 27] have been developed under the assumption of ergodicity. The most relevant theory for our setting — SDEs derived from system size expansion — is contained in the rigorous treatments of Katzenberger and Funaki [28, 29], proving the convergence of the homogenised slow/fast stochastic

Case	Procedure
Outer system ( $\varepsilon = \mu = 0$ ) is solvable	Use equations (3, 4) and (8, 9)
Manifold is one-dimensional	Use equations (9) and (15, 16)
Manifold has co-dimensional one	Use equations (23) and (24)
Manifold is $m$ -dimensional	Use equations (9) and (17 – 21)

**Table 1.** Quick reference table of equations applying to different cases of slow-manifold reduction.

system to one of lower dimension that is restricted to the slow manifold, which is valid in the long-term, in a slow timescale. We give a more detailed statement of Katzenberger’s theorem in [Appendix A](#), however, the results of these works are somewhat difficult to apply in practice as they are formulated in terms of a quantity (the flow map of the fast outer system), which is general has no closed analytical solution.

In this article we present for the first time a computationally explicit formulation of the rigorous theory of Katzenberger, in terms of quantities that can be directly computed. The end product is a single robust, systematic and provably correct procedure for timescale separation in stochastic dynamical systems with intrinsic noise, which we believe will be of considerable general use. Our main results are contained [Section 2](#), where we describe a map from a high-dimensional system of equations (1) to a lower-dimensional one (2), via explicit formulae that are summarised in [Table 1](#). The subsections contain (i) a new explicit derivation of the theory for one-dimensional manifolds using a perturbation expansion, which we hope should be useful for readers wishing to gain intuition about the method, (ii) the general procedure for arbitrary dimension, and (iii) explicit closed-form expressions for the case of manifolds of co-dimension one. In [Section 3](#) we present three exploratory examples: (i) we demonstrate the basic theory for the well understood example of Michaelis-Menten kinetics for enzyme catalysed reactions, (ii) we use our expressions for co-dimension one manifolds to give a new derivation of the relationship between Wright-Fisher diffusion and near neutral Lotka-Volterra dynamics, and (iii) we present a new extension of the method to infinite dimensional processes. Appendices contain technical details.

## 2. Reduced model description

We consider Langevin stochastic differential equations of the general type

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}) + \varepsilon\mathbf{h}(\mathbf{x}) + \sqrt{\mu}\mathbf{G}(\mathbf{x})\boldsymbol{\eta}(t), \quad (1)$$

where the state variable  $\mathbf{x} = (x_1, \dots, x_d)^T$  is an  $d$ -dimensional vector, and there are  $s$  independent Itô white noise sources  $\boldsymbol{\eta}(t) = (\eta_1(t), \dots, \eta_s(t))^T$ . The vector-valued functions  $\mathbf{f} : \mathbb{R}^d \rightarrow \mathbb{R}^d$  and  $\mathbf{h} : \mathbb{R}^d \rightarrow \mathbb{R}^d$  are the ‘outer’ and ‘inner’ parts of the drift respectively, and the matrix valued function  $\mathbf{G} : \mathbb{R}^d \rightarrow \mathbb{R}^{d,s}$  specifies the coupling of

state variables to noise sources. We assume throughout that  $\mathbf{f}$  is twice differentiable, but place no constraints on the other functions. The parameters  $\varepsilon$  and  $\mu$  determine the separation of timescales and the strength of the noise, respectively.

We do not assume an *a priori* separation into slow and fast variables, as is common in the literature, as in the applications that motivate us, an appropriate change of variables is frequently neither evident nor analytically tractable (although see [9] for an example where the coordinates can be globally constructed, and [30, 31, 32] for some recent advances in computational methods to identify the change of variable), and our method does not require that they be known.

We are interested in the case when  $\varepsilon$  and  $\mu$  are small and  $\mathbf{f}$  possesses an attracting  $m$ -dimensional submanifold of equilibria  $\Gamma \subset \mathbb{R}^d$  (*i.e.*  $\mathbf{f}(\tilde{\mathbf{x}}) = \mathbf{0}$  for all  $\tilde{\mathbf{x}} \in \Gamma$ ). For simplicity, we assume that this manifold is unique, connected, and globally attracting (*i.e.* it is a normally hyperbolic slow manifold, see *e.g.* [33]); then we expect solutions of (1) to rapidly approach and remain very close to  $\Gamma$ . In fact, it has been rigorously proved by Katzenberger [28] that the trajectories of  $\mathbf{x} \in \mathbb{R}^d$  converge those of a stochastic variable  $\tilde{\mathbf{x}} \in \Gamma$  with dynamics

$$\frac{d\tilde{\mathbf{x}}}{dt} = \varepsilon \mathbf{P}(\tilde{\mathbf{x}}) \mathbf{h}(\tilde{\mathbf{x}}) + \mu \mathbf{g}(\tilde{\mathbf{x}}) + \sqrt{\mu} \mathbf{P}(\tilde{\mathbf{x}}) \mathbf{G}(\tilde{\mathbf{x}}) \boldsymbol{\eta}(t), \quad (2)$$

where  $\mathbf{P}$  is a certain projection matrix derived from  $\mathbf{f}$ , and  $\mathbf{g}$  is a new contribution to the drift arising from the way in which fluctuations away from the manifold are suppressed; as our examples illustrate, unlike the deterministic situation, it is not sufficient to simply restrict (1) to  $\Gamma$  to obtain the slow dynamics. Our purpose here is to derive explicit expressions for  $\mathbf{P}$  and  $\mathbf{g}$ . Readers with a specific problem in mind may wish to jump straight to the appropriate result, which can be found by referring to Table 1.

Before we proceed with our main task, we give a brief sketch of the derivation of (2). Examining (1) when  $\varepsilon$  and  $\mu$  are small, one might imagine a picture in which the state of the system is quickly carried onto the manifold by the fast outer drift term  $\mathbf{f}$ . Following this fast initial transient, it may then receive multiple stochastic ‘kicks’ carrying it away from the manifold, each time only to return again via the paths described by  $\mathbf{f}$ . See Figure 1 for an illustrative example. This intuition can be made concrete by considering the flow map of the outer system. Let  $\mathbf{x}$  be a point in the state space and consider the deterministic initial value problem

$$\begin{cases} \frac{d\boldsymbol{\xi}_x}{dt} = \mathbf{f}(\boldsymbol{\xi}_x) \\ \boldsymbol{\xi}_x(0) = \mathbf{x}. \end{cases} \quad (3)$$

Since the centre manifold is globally attractive, all trajectories lead eventually to  $\Gamma$  and we may thus define the asymptotic phase [33],  $\boldsymbol{\pi} : \mathbb{R}^d \rightarrow \Gamma$ , giving the endpoint of the deterministic trajectories

$$\boldsymbol{\pi}(\mathbf{x}) = \lim_{t \rightarrow \infty} \boldsymbol{\xi}_x(t), \quad (4)$$

where  $\boldsymbol{\xi}_x$  is the solution of (3).

If we take the point  $\mathbf{x}$  to be the current location of the random variable governed by equation (1), then  $\boldsymbol{\pi}(\mathbf{x})$  defines another random variable that tracks the motion of  $\mathbf{x}$  but is constrained to the manifold. Application of Itô's formula [34] gives the Langevin equations for each spatial coordinate:

$$\begin{aligned} \frac{d}{dt}\pi_i(\mathbf{x}) &= \sum_j \frac{\partial \pi_i}{\partial x_j} \frac{dx_j}{dt} + \frac{\mu}{2} \sum_{s,j,k} G_{js}(\mathbf{x}) G_{ks}(\mathbf{x}) \frac{\partial^2 \pi_i}{\partial x_j \partial x_k} \\ &= \varepsilon \sum_j \frac{\partial \pi_i}{\partial x_j} h_j(\mathbf{x}) + \frac{\mu}{2} \sum_{s,j,k} G_{js}(\mathbf{x}) G_{ks}(\mathbf{x}) \frac{\partial^2 \pi_i}{\partial x_j \partial x_k} + \sqrt{\mu} \sum_{s,j} G_{js}(\mathbf{x}) \frac{\partial \pi_i}{\partial x_j} \eta_s(t). \end{aligned} \quad (5)$$

where the contribution from  $\mathbf{f}$  has vanished because  $\boldsymbol{\pi}(\mathbf{x})$  is contained in the slow manifold where  $\mathbf{f} = \mathbf{0}\ddagger$ . Unfortunately equation (5) is not closed since it relies on full knowledge of the random variable  $\mathbf{x}$ . However, if we believe that  $\mathbf{x}$  remains very close to  $\Gamma$  (as is the case when  $\varepsilon$  and  $\mu$  are small) then we might be motivated to consider a new random variable  $\tilde{\mathbf{x}} \in \Gamma$  which we assume is a close approximation to both  $\mathbf{x}$  and  $\boldsymbol{\pi}(\mathbf{x})$ . Substituting  $\tilde{\mathbf{x}}$  for both these quantities in (5), we obtain the closed expression

$$\frac{d\tilde{x}_i}{dt} = \varepsilon \sum_j P_{ij}(\tilde{\mathbf{x}}) h_j(\tilde{\mathbf{x}}) + \frac{\mu}{2} \sum_{s,j,k} G_{js}(\tilde{\mathbf{x}}) G_{ks}(\tilde{\mathbf{x}}) Q_{ijk}(\tilde{\mathbf{x}}) + \sqrt{\mu} \sum_{s,j} P_{ij}(\tilde{\mathbf{x}}) G_{js}(\tilde{\mathbf{x}}) \eta_s(t), \quad (7)$$

where  $\mathbf{P}$  is a matrix and  $\mathbf{Q}$  an array defined by

$$P_{ij}(\tilde{\mathbf{x}}) = \left. \frac{\partial}{\partial x_j} \pi_i(\mathbf{x}) \right|_{\mathbf{x}=\tilde{\mathbf{x}}}, \quad Q_{ijk}(\tilde{\mathbf{x}}) = \left. \frac{\partial^2}{\partial x_j \partial x_k} \pi_i(\mathbf{x}) \right|_{\mathbf{x}=\tilde{\mathbf{x}}}. \quad (8)$$

Equivalently we may rewrite (7) as equation (2), where the additional drift term is

$$\mathbf{g}(\tilde{\mathbf{x}}) = \frac{1}{2} \sum_{s,j,k} G_{js}(\tilde{\mathbf{x}}) G_{ks}(\tilde{\mathbf{x}}) Q_{ijk}(\tilde{\mathbf{x}}). \quad (9)$$

The projection  $\mathbf{P}(\tilde{\mathbf{x}})\S$  is entirely determined by the first order terms of (1), and typically it can be straightforwardly reconstructed from knowledge of the eigenvectors

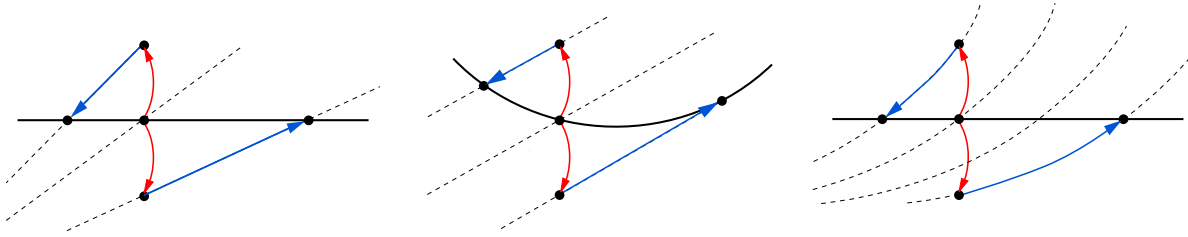
$\ddagger$  The component  $\mathbf{f}$  vanishes as  $\boldsymbol{\pi}(\boldsymbol{\xi}_{\mathbf{x}}(t)) = \boldsymbol{\pi}(\mathbf{x})$  for all  $t$ , and thus

$$0 = \left. \frac{d}{dt} \right|_{t=0} \pi_i(\boldsymbol{\xi}_{\mathbf{x}}(t)) = \sum_j \left. \frac{\partial \pi_i}{\partial x_j} \frac{d\xi_j}{dt} \right|_{t=0} = \sum_j \frac{\partial \pi_i}{\partial x_j} f_j(\mathbf{x}). \quad (6)$$

$\S$  For  $\tilde{\mathbf{x}} \in \Gamma$ , the matrix  $\mathbf{P}(\tilde{\mathbf{x}})$  is a projection: since  $\Gamma$  is composed of fixed points,  $\boldsymbol{\pi}(\boldsymbol{\pi}(\mathbf{x})) = \boldsymbol{\pi}(\mathbf{x})$  and  $\boldsymbol{\pi}(\tilde{\mathbf{x}}) = \tilde{\mathbf{x}}$ ; thus, using the chain rule,

$$\begin{aligned} P_{ij}(\tilde{\mathbf{x}}) &= \left. \frac{\partial}{\partial x_j} \pi_i(\mathbf{x}) \right|_{\mathbf{x}=\tilde{\mathbf{x}}} = \left. \frac{\partial}{\partial x_j} \pi_i(\boldsymbol{\pi}(\mathbf{x})) \right|_{\mathbf{x}=\tilde{\mathbf{x}}} \\ &= \sum_k \left. \frac{\partial}{\partial x_k} \pi_i(\boldsymbol{\pi}(\mathbf{x})) \frac{\partial}{\partial x_j} \pi_k(\boldsymbol{\pi}(\mathbf{x})) \right|_{\mathbf{x}=\tilde{\mathbf{x}}} = \sum_k P_{ik}(\tilde{\mathbf{x}}) P_{kj}(\tilde{\mathbf{x}}). \end{aligned}$$

Moreover, since for  $\mathbf{X} \in \mathbb{R}^d$ ,  $\Gamma(t) = \boldsymbol{\pi}(\tilde{\mathbf{x}} + t\mathbf{X})$  is a one-parameter curve in  $\Gamma$  with  $\dot{\gamma}(0) = \mathbf{P}(\tilde{\mathbf{x}})\mathbf{X}$ , we see that the image of  $\mathbf{P}(\tilde{\mathbf{x}})$  is the tangent plane to  $\Gamma$  at  $\tilde{\mathbf{x}}$ . One may similarly show that the kernel of the projection is the image of  $\mathbf{J}(\tilde{\mathbf{x}})$ .



**Figure 2. Left:** Here, the variation in the angle between the fast (dashed) and slow (solid) subspaces creates a bias in the location of the return to the manifold of a perturbation away from it; an upward perturbation returns quite close on the left of the origin, but an equally likely downward perturbation is carried far to the right. **Centre:** The same effect can occur as a result of curvature of the manifold. In this figure the flow fields are parallel, but the manifold curves, resulting in the same rightwards bias in the projected system. **Right:** Curvature of the flow field may also induce bias, even when the angle of intersection is constant.

of the Jacobian matrix of  $\mathbf{f}$ . The calculation of the noise-induced drift term  $\mathbf{g}$  is more complicated, having contributions from three possible sources: variation of the alignment of the flow field, curvature of the manifold, and curvature of the flow field. Each of these mechanisms can induce a bias in the direction of flow of the reduced dimension system, as illustrated in Figure 2. In the following subsections we will present explicit procedures for computing  $\mathbf{P}$  and  $\mathbf{Q}$ .

### 2.1. One-dimensional manifolds

The simplest case to treat is that of a one-dimensional manifold, as the second-order perturbation expansion is explicitly solvable. Suppose that the slow manifold  $\Gamma$  is a curve parameterised by the first spatial co-ordinate of the system<sup>||</sup>. That is, there exists function  $\gamma$  such that

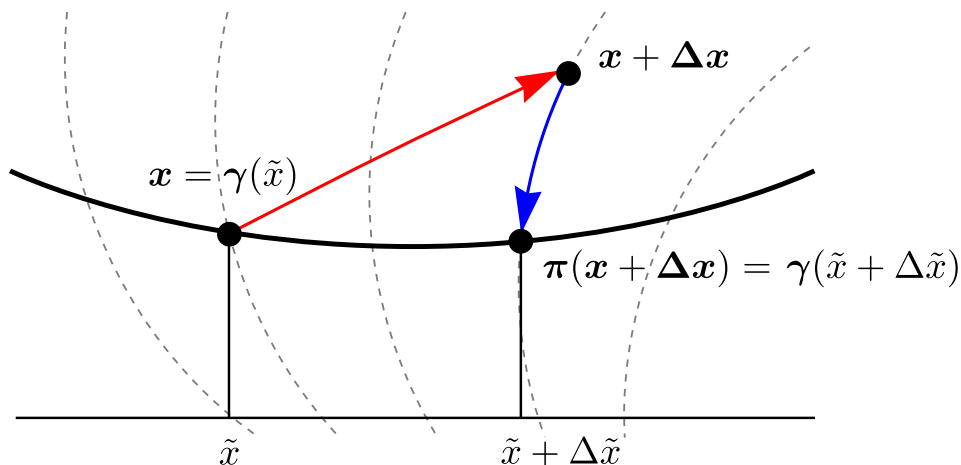
$$\mathbf{x} \in \Gamma \quad \Leftrightarrow \quad \mathbf{x} = \gamma(x_1). \quad (10)$$

In this case the dynamics of the reduced system  $\tilde{\mathbf{x}}$  defined in (2) are determined entirely by the first component, so we need only to compute the partial derivatives of  $\pi_1$ . For ease of notation we will drop the subscript 1 from now on, writing  $\tilde{x} := \tilde{x}_1$  as well as  $P_j := P_{1j}(\gamma(\tilde{x}_1))$  and  $Q_{jk} := Q_{1jk}(\gamma(\tilde{x}_1))$ .

We undertake a second-order perturbation theory, informed by the intuition that a small perturbation may move the system away from the manifold, before it returns via the outer flow field. Consider a point  $\mathbf{x} = \gamma(\tilde{x})$  on the manifold; by definition it is unmoved by the action of the outer flow field, so  $\pi(\mathbf{x}) = \gamma(\tilde{x})$ . To access expressions for  $P_j$  and  $Q_{jk}$ , we will consider the relationship between the small perturbations  $\Delta\mathbf{x}$

<sup>||</sup> Note that we have chosen this case for simplicity of presentation, and not all 1D manifolds can in fact be tackled in this way (e.g. a circular manifold would fail here). The more general case of a manifold described by an arbitrary parameterised curve is not substantially different, however, as we only ever require the local properties of the projection  $\pi$ , and for smooth manifolds there is always a local coordinate system in which the problem can be set up in the required format.





**Figure 3.** Illustration of the perturbation calculation for a 1D manifold  $\gamma$  parameterised by a coordinate  $\tilde{x}$ . We imagine the system with initial state  $\mathbf{x} = \gamma(\tilde{x})$  receives a small perturbation  $\Delta\mathbf{x}$  (red arrow), and then relaxes back to the manifold via the flow line of the fast system (blue arrow). This process is equivalent to making a corresponding perturbation  $\Delta\tilde{x}$  to the slow manifold coordinate.

and  $\Delta\tilde{x}$  such that  $\pi(\mathbf{x} + \Delta\mathbf{x}) = \gamma(\tilde{x} + \Delta\tilde{x})$ .

On the one hand, because we set the problem up so that  $\pi_1(\mathbf{x}) = \tilde{x}$ , Taylor expansion gives

$$\Delta\tilde{x} = \sum_j P_j \Delta x_j + \frac{1}{2} \sum_{j,k} Q_{jk} \Delta x_j \Delta x_k + \dots \quad (11)$$

Alternatively, we can consider the two-step process of making a perturbation to the system state and then following the outer flow field back to the manifold – see Figure 3 for an illustration. If we make second order approximations to the flow field and the manifold, then the  $\Delta\tilde{x}$  computed by this method must match that of (11). This correspondence will allow us to solve for  $P_j$  and  $Q_{jk}$ .

Near the point  $\mathbf{x} \in \Gamma$  we can approximate the action of  $\pi$  by constructing the quadratic expansion of the preimage. Specifically, it can be shown that in the neighbourhood of  $\mathbf{x}$ , the collection of nearby points that would be mapped to  $\mathbf{x}$  by  $\pi$  (i.e. the invariant foliation [33],  $\pi^{-1}(\mathbf{x}) = \{\mathbf{y} : \pi(\mathbf{y}) = \mathbf{x}\}$ ) is approximated to second order by the set of points  $\mathbf{y}$  such that

$$\mathbf{v}(\tilde{x})^T (\mathbf{y} - \gamma(\tilde{x})) + (\mathbf{y} - \gamma(\tilde{x}))^T \Theta(\tilde{x}) (\mathbf{y} - \gamma(\tilde{x})) = 0, \quad (12)$$

where  $\mathbf{v}(\tilde{x})$  is a perpendicular vector to the flow field near  $\mathbf{x} = \gamma(\tilde{x})$  and  $\Theta(\tilde{x})$  a matrix describing the curvature of the flow field near the same point. In Appendix B we give an explicit derivation of these quantities from  $\mathbf{f}$ ; for now we assume they are known. Recall that we are seeking the perturbation  $\Delta\tilde{x}$  such that  $\pi(\mathbf{x} + \Delta\mathbf{x}) = \gamma(\tilde{x} + \Delta\tilde{x})$ , to

second order. We make the following Taylor expansions of various orders:

$$\begin{aligned}
\left[ \mathbf{x} + \Delta \mathbf{x} - \boldsymbol{\gamma}(\tilde{x} + \Delta \tilde{x}) \right]_{\ell} &= \Delta x_{\ell} - \gamma'_{\ell} \Delta \tilde{x} - \frac{1}{2} \gamma''_{\ell} (\Delta \tilde{x})^2 + \dots \\
&= \sum_k (\delta_{k,\ell} - \gamma'_{\ell} P_k) \Delta x_k - \frac{1}{2} \sum_{j,k} (\gamma'_{\ell} Q_{jk} + \gamma''_{\ell} P_j P_k) \Delta x_j \Delta x_k + \mathcal{O}(\Delta \mathbf{x}^3) \\
\left[ \mathbf{v}(\tilde{x} + \Delta \tilde{x}) \right]_{\ell} &= v_{\ell} + v'_{\ell} \Delta \tilde{x} + \dots = v_{\ell} + v'_{\ell} \sum_j P_j \Delta x_j + \mathcal{O}(\Delta \mathbf{x}^2) \\
\left[ \Theta(\tilde{x} + \Delta \tilde{x}) \right]_{jk} &= \Theta_{jk} + \mathcal{O}(\Delta \mathbf{x}).
\end{aligned} \tag{13}$$

Here we use  $[\dots]_l$  and  $[\dots]_{jk}$  to indicate the  $l^{\text{th}}$  (resp.  $j, k^{\text{th}}$ ) entry of the vector or matrix in brackets,  $\delta_{k,l}$  to indicate the Kronecker delta function, and drop the argument  $\tilde{x}$  from  $\gamma'_{\ell}$ ,  $\gamma''_{\ell}$ ,  $v_{\ell}$ ,  $v'_{\ell}$  and  $\Theta_{jk}$  to avoid clutter. Following (12), the requirement that  $\boldsymbol{\pi}(\mathbf{x} + \Delta \mathbf{x}) = \boldsymbol{\gamma}(\tilde{x} + \Delta \tilde{x})$  to second order becomes

$$\begin{aligned}
0 &= \mathbf{v}(\tilde{x} + \Delta \tilde{x})^T \left( \mathbf{x} + \Delta \mathbf{x} - \boldsymbol{\gamma}(\tilde{x} + \Delta \tilde{x}) \right) \\
&\quad + \left( \mathbf{x} + \Delta \mathbf{x} - \boldsymbol{\gamma}(\tilde{x} + \Delta \tilde{x}) \right)^T \Theta(\tilde{x} + \Delta \tilde{x}) \left( \mathbf{x} + \Delta \mathbf{x} - \boldsymbol{\gamma}(\tilde{x} + \Delta \tilde{x}) \right) + \mathcal{O}(\Delta \mathbf{x}^3) \\
&= \sum_{\ell} \left[ v_{\ell} + v'_{\ell} \sum_j P_j \Delta x_j \right] \left[ \sum_k (\delta_{k,\ell} - \gamma'_{\ell} P_k) \Delta x_k - \frac{1}{2} \sum_{j,k} (\gamma'_{\ell} Q_{jk} + \gamma''_{\ell} P_j P_k) \Delta x_j \Delta x_k \right] \\
&\quad + \sum_{j,k} \Theta_{jk} \Delta x_j \Delta x_k + \mathcal{O}(\Delta \mathbf{x}^3) \\
&= \sum_k \left\{ \sum_{\ell} v_{\ell} (\delta_{k,\ell} - \gamma'_{\ell} P_k) \right\} \Delta x_k \\
&\quad + \frac{1}{2} \sum_{j,k} \left\{ \sum_{\ell} v'_{\ell} (\delta_{k,\ell} + \delta_{j,\ell} - 2\gamma'_{\ell} P_k) P_j - \sum_{\ell} v_{\ell} (\gamma'_{\ell} Q_{jk} + \gamma''_{\ell} P_j P_k) + 2\Theta_{jk} \right\} \Delta x_j \Delta x_k \\
&\quad + \mathcal{O}(\Delta \mathbf{x}^3).
\end{aligned} \tag{14}$$

Since the perturbation  $\Delta \mathbf{x}$  was arbitrary, we require each term in curly brackets above to be equal to zero. From the first order terms we conclude that

$$P_k = \frac{v_k}{\sum_{\ell} v_{\ell} \gamma'_{\ell}}, \tag{15}$$

and from the second order that

$$Q_{jk} = \frac{1}{\sum_{\ell} v_{\ell} \gamma'_{\ell}} \left( v'_k P_j + v'_j P_k + 2\Theta_{jk} - \sum_{\ell} (2v'_{\ell} \gamma'_{\ell} + v_{\ell} \gamma''_{\ell}) P_j P_k \right). \tag{16}$$

Written this way, the separate contributions from variation of the flow field (terms involving  $\mathbf{v}'$ ), curvature of flow field ( $\Theta$ ), and curvature of the manifold (the  $\boldsymbol{\gamma}''$  term) are clearly visible.

In higher dimensions, the above perturbation expansion is less useful, as it produces a larger system of equations which lacks an explicit solution. A different line of attack is necessary.

## 2.2. General case

If the linearisation of the flow field  $\phi_t$  is known in the neighbourhood of the manifold then  $\mathbf{P}$  can be reconstructed easily. Specifically, around a point  $\mathbf{z} \in \Gamma$  the state space  $\mathbb{R}^d$  can be decomposed into a product of ‘slow’ and ‘fast’ subspaces of dimension  $m$  and  $d - m$ , respectively. The slow subspace is the tangent plane to the manifold at the given point; a perturbation in one of these directions is unaffected by the action of  $\mathbf{f}$ . Conversely, the fast subspace comprises perturbation directions that collapse quickly back to the manifold. The projection matrix  $\mathbf{P}(\mathbf{z})$  acts as the identity on the slow subspace and as zero on the fast subspace.

Unfortunately, no such simple formulation is available for  $\mathbf{Q}(\mathbf{z})$  in general. This problem was explored in [13], where the following method was developed. This result is explained fully in [Appendix C](#), for now we simply present the computational steps.

### Procedure for calculating $\mathbf{P}$ and $\mathbf{Q}$ at a point $\mathbf{z} \in \Gamma$

- (i) Compute the Jacobian matrix  $\mathbf{J}$  of  $\mathbf{f}$ ,

$$J_{ij} = \left. \frac{\partial f_i}{\partial x_j} \right|_{\mathbf{x}=\mathbf{z}},$$

and diagonalise it, writing

$$\mathbf{J} = \mathbf{W} \mathbf{\Lambda} \mathbf{W}^{-1}. \quad (17)$$

where  $\mathbf{W} = (\mathbf{w}_1 \cdots \mathbf{w}_n)$  is a matrix of eigenvectors forming a basis of  $\mathbb{R}^d$ , with the  $m$  slow directions written first.  $\mathbf{\Lambda}$  is a diagonal matrix of eigenvalues with  $\lambda_1 = \cdots = \lambda_m = 0$  and  $\text{Re}(\lambda_{m+1}), \dots, \text{Re}(\lambda_n) < 0$ . Also compute the pseudo-inverse

$$\mathbf{J}^+ = \mathbf{W} \mathbf{\Lambda}^+ \mathbf{W}^{-1}, \quad (18)$$

where  $\mathbf{\Lambda}^+$  is the diagonal matrix with eigenvalues  $\lambda_1^+ = \cdots = \lambda_n^+$ , where

$$\lambda^+ = \begin{cases} 0 & \text{if } \lambda = 0 \\ 1/\lambda & \text{if } \lambda \neq 0. \end{cases}$$

- (ii) For each  $i$ , compute the Hessian of  $f_i$ ,  $\mathbf{H}_i$ , defined by

$$H_{ijk} = \left. \frac{\partial^2 f_i(\mathbf{x})}{\partial x_j \partial x_k} \right|_{\mathbf{x}=\mathbf{z}}.$$

Then find the (matrix-valued) solution  $\mathbf{X}_i$  of the Lyapunov equation

$$\mathbf{J}^T \mathbf{X}_i + \mathbf{X}_i \mathbf{J} = -\mathbf{H}_i. \quad (19)$$

NB: this is a linear problem that is straightforwardly solved [35].

(iii) Finally, the projection matrix is given by

$$\mathbf{P} = \mathbf{I} - \mathbf{J}^+ \mathbf{J}. \quad (20)$$

and for  $\mathbf{Q}$  we have

$$Q_{ijk} = \sum_l -J_{il}^+ [\mathbf{P}^T \mathbf{H}_l \mathbf{P}]_{jlk} + P_{il} [\mathbf{X}_l - \mathbf{J}^{+T} \mathbf{H}_l \mathbf{P} - \mathbf{P}^T \mathbf{H}_l \mathbf{J}^+]_{jlk}. \quad (21)$$

In the case when  $\mathbf{f}$  is of gradient form *i.e.*,  $\mathbf{f} = \nabla U$  for a scalar potential  $U$ , then a similar procedure gives an elegant closed form for the induced drift [29, 36].

### 2.3. Co-dimension one manifolds

We now use the results of the previous section to obtain explicit expressions for the derivatives in the case when  $\Gamma$  is a  $(d-1)$ -dimensional manifold. In this case, in a small neighbourhood around any point  $\mathbf{z} \in \Gamma$ , the flow field can be decomposed as  $\mathbf{f} = \phi \mathbf{r}$ , into a scalar part  $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$  that vanishes on  $\Gamma$ , and a non-vanishing vector part  $\mathbf{r} : \mathbb{R}^d \rightarrow \mathbb{R}^d$ . Using this decomposition we compute an expression for the Jacobian around a point  $\mathbf{x}$ :

$$\mathbf{J}(\mathbf{x}) = \phi(\mathbf{x}) \frac{\partial \mathbf{r}}{\partial \mathbf{x}} + \mathbf{r}(\mathbf{x}) \nabla \phi(\mathbf{x})^T,$$

In particular, evaluated at the point  $\mathbf{z}$  on the manifold we have  $\mathbf{J} = \mathbf{r} \nabla \phi^T$ . Meaning that  $\mathbf{r}$  is, up to scalar multiple, the unique eigenvector corresponding to

$$\lambda = \nabla \phi^T \mathbf{r},$$

which is the sole non-zero eigenvalue of  $\mathbf{J}$ . Note that here and hereafter we drop the argument  $\mathbf{z}$  to avoid notational clutter. As  $\mathbf{r}$  spans the non-zero eigenspaces, it is straightforward to check that the pseudo-inverse may be written as

$$\mathbf{J}^+ = \frac{1}{\lambda} \mathbf{J}. \quad (22)$$

We conclude from (20) that

$$\mathbf{P} = \mathbf{I} - \mathbf{J}^+ \mathbf{J} = \mathbf{I} - \frac{1}{\lambda} \mathbf{J}. \quad (23)$$

¶ Given a scalar function  $\psi$  of  $\mathbf{x} = (x_1, \dots, x_d)$ , we write

$$\nabla \psi = \left( \frac{\partial \psi}{\partial x_1}, \dots, \frac{\partial \psi}{\partial x_d} \right),$$

and use  $\frac{\partial \psi}{\partial \mathbf{x}}$  for the matrix with  $i, j^{\text{th}}$  entry

$$\frac{\partial^2 \psi}{\partial x_i \partial x_j}.$$

When  $\Psi$  is a vector-valued function, we write  $\frac{\partial \Psi}{\partial \mathbf{x}}$  for its Jacobian matrix, reserving  $\mathbf{J}$  for the Jacobian matrix of  $\mathbf{f}$ .

To determine  $\mathbf{Q}$ , it thus remains to solve (19),

$$\mathbf{J}^T \mathbf{X}_i + \mathbf{X}_i \mathbf{J} = -\mathbf{H}_i,$$

for  $\mathbf{H}$  and insert into (21). As we show in Appendix C, in this case, equation (C.9) can be explicitly solved in closed form to give

$$X_{ijk} = -\frac{\frac{\partial \phi}{\partial x_j} \frac{\partial \phi}{\partial x_k}}{2\lambda^3} \mathbf{r}^T \frac{\partial^2 f_i}{\partial \mathbf{x}^2} \mathbf{r}.$$

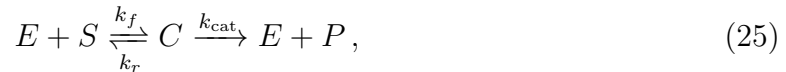
Finally, observing that  $\frac{\partial^2 f_i}{\partial x_j \partial x_k} = \frac{\partial r_i}{\partial x_j} \frac{\partial \phi}{\partial x_k} + r_i \frac{\partial^2 \phi}{\partial x_j \partial x_k}$ , substituting the above into (21) and considerable algebraic simplification yields

$$Q_{ijk} = -\frac{1}{\lambda} \left( [\mathbf{P}^T \frac{\partial^2 \phi}{\partial \mathbf{x}^2} \mathbf{P}]_{jk} r_i + \frac{\partial \phi}{\partial x_j} [\mathbf{P} \frac{\partial \mathbf{r}}{\partial \mathbf{x}}]_{ik} + \frac{\partial \phi}{\partial x_k} [\mathbf{P} \frac{\partial \mathbf{r}}{\partial \mathbf{x}}]_{ij} \right) + \frac{1}{\lambda^2} \frac{\partial \phi}{\partial x_j} \frac{\partial \phi}{\partial x_k} [\mathbf{P} \frac{\partial \mathbf{r}}{\partial \mathbf{x}}]_i. \quad (24)$$

### 3. Worked examples

#### 3.1. Simple example: Michaelis-Menten kinetics

The Michaelis-Menten law is perhaps one of the most widely-applied examples of timescale separation. It is a model for the net rate of production in a chemical reaction that is catalysed by an enzyme, in which it is assumed that the process of enzyme binding and unbinding occurs very much faster than the catalytic reaction of interest. Using the notation of chemical reactions, one may write



where  $E$  symbolises the enzyme,  $S$  the substrate,  $C$  the enzyme-substrate complex, and  $P$  the product. The parameters  $k_f$  and  $k_r$  give the rate of binding (forward) and unbinding (reverse) of the enzyme to the substrate, while  $k_{\text{cat}}$  specifies the rate of catalysis.

Assuming the reaction takes place in a domain of infinite volume, one may write rate the deterministic equations

$$\begin{aligned} \frac{dS}{dt} &= -k_f ES + k_r C, \\ \frac{dE}{dt} &= -k_f ES + (k_r + k_{\text{cat}}) C, \\ \frac{dC}{dt} &= k_f ES - (k_r + k_{\text{cat}}) C, \\ \frac{dP}{dt} &= k_{\text{cat}} C, \end{aligned} \quad (26)$$

where  $S$ ,  $C$ ,  $P$  and  $E$  now represent the *concentrations* of the various reactants. Note that this system has only two degrees of freedom due to conservation relations

$E + C = E_0$  and  $S + C + P = S_0$ , where  $E_0$  and  $S_0$  are the initial concentrations of the enzyme and substrate, respectively. If  $k_f, k_r \gg k_{\text{cat}}$  we might approximate the concentration of the complex  $C$  by the equilibrium value it would have if  $k_{\text{cat}}$  were actually zero:

$$k_f ES - k_r C \approx 0 \quad \Rightarrow \quad C \approx E_0 \frac{S}{k + S}, \quad (27)$$

where  $k = k_r/k_f$ . Introducing  $v^* = k_{\text{cat}}E_0$ , on the slower timescale the net production rate is then found to be

$$\frac{dP}{dt} = \frac{v^* S}{k + S}. \quad (28)$$

This is the Michaelis-Menten law.

In finite volume domains chemical reactions are subject to random fluctuations arising from the discrete nature of the molecules involved. A more appropriate description in these circumstances is a stochastic differential equation, with noise terms that are derived from the instantaneous reaction rates (each possible reaction introduces its own source of noise). For the reaction described above in (25) occurring in a domain of volume  $V$ , equations are derived following Kurtz [1] :

$$\begin{aligned} \frac{dS}{dt} &= -k_f(E_0 - C)S + k_r C - \sqrt{\frac{k_f(E_0 - C)S}{V}} \eta_f(t) + \sqrt{\frac{k_r C}{V}} \eta_r(t), \\ \frac{dC}{dt} &= k_f(E_0 - C)S - (k_r + k_{\text{cat}})C + \sqrt{\frac{k_f(E_0 - C)S}{V}} \eta_f(t) \\ &\quad - \sqrt{\frac{k_r C}{V}} \eta_r(t) - \sqrt{\frac{k_{\text{cat}} C}{V}} \eta_{\text{cat}}(t). \end{aligned} \quad (29)$$

(In fact this step is not strictly necessary; we could choose to work directly with the process of particle numbers, as described in [Appendix A](#)). Following similar lines to [37] a dimensionless form may be found by rescaling time  $t \mapsto k_f E_0 t$  and introducing variables

$$\mathbf{x} = \begin{pmatrix} S/S_0 \\ C/E_0 \end{pmatrix}, \quad (30)$$

and parameters

$$\varepsilon = \frac{k_{\text{cat}}}{k_f E_0} > 0, \quad \mu = \frac{1}{S_0 V}, \quad \alpha = \frac{k_r}{k_f S_0} > 0, \quad \beta = \frac{S_0}{E_0} > 0. \quad (31)$$

The result is a system of exactly the form of equation (1):

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}) + \varepsilon \mathbf{h}(\mathbf{x}) + \sqrt{\mu} \mathbf{G}(\mathbf{x}) \boldsymbol{\eta}(t), \quad (32)$$

where

$$\mathbf{f}(\mathbf{x}) = \begin{pmatrix} -x_1 + (x_1 + \alpha)x_2 \\ \beta(x_1 - (x_1 + \alpha)x_2) \end{pmatrix}, \quad \mathbf{h}(\mathbf{x}) = \begin{pmatrix} 0 \\ -x_2 \end{pmatrix}, \quad (33)$$

and

$$\mathbf{G}(\mathbf{x}) = \begin{pmatrix} -\sqrt{(1-x_2)x_1} & \sqrt{\alpha x_2} & 0 \\ \beta\sqrt{(1-x_2)x_1} & -\beta\sqrt{\alpha x_2} & -\sqrt{\varepsilon\beta x_2} \end{pmatrix}, \quad \boldsymbol{\eta}(t) = \begin{pmatrix} \eta_f(t) \\ \eta_r(t) \\ \eta_{\text{cat}}(t) \end{pmatrix}. \quad (34)$$

The slow manifold in this case is the curve  $x_1 - x_2(x_1 + \alpha) = 0$ , along which  $\mathbf{f}(\mathbf{x}) = \mathbf{0}$ . See Figure 1 for an illustration.

Let us take  $\tilde{x} = x_1$  as the slow variable and proceed to calculate a reduced system in terms of  $\tilde{x}$  only. As the manifold is one-dimensional, we are able to simply follow the procedure laid out above. We begin by writing down the formula for the slow manifold and its  $\tilde{x}$  derivatives:

$$\boldsymbol{\gamma}(\tilde{x}) = \begin{pmatrix} \tilde{x} \\ \frac{\tilde{x}}{\tilde{x}+\alpha} \end{pmatrix} \quad \boldsymbol{\gamma}'(\tilde{x}) = \begin{pmatrix} 1 \\ \frac{\alpha}{(\tilde{x}+\alpha)^2} \end{pmatrix} \quad \boldsymbol{\gamma}''(\tilde{x}) = \begin{pmatrix} 0 \\ \frac{-2\alpha}{(\tilde{x}+\alpha)^3} \end{pmatrix}. \quad (35)$$

Next, we find the Jacobian matrix on the manifold

$$\mathbf{J}(\mathbf{x}) = \begin{pmatrix} x_2 - 1 & x_1 + \alpha \\ \beta(1 - x_2) & -\beta(x_1 + \alpha) \end{pmatrix} \Rightarrow \mathbf{J}(\tilde{x}) = \begin{pmatrix} \frac{\tilde{x}}{\tilde{x}+\alpha} - 1 & \tilde{x} + \alpha \\ \beta(1 - \frac{\tilde{x}}{\tilde{x}+\alpha}) & -\beta(\tilde{x} + \alpha) \end{pmatrix}. \quad (36)$$

Diagonalising  $\mathbf{J}(\tilde{x})$  we find the left eigenvector  $\mathbf{v}(\tilde{x})$  corresponding to the eigenvalue zero, and its  $\tilde{x}$  derivative:

$$\mathbf{v}(\tilde{x}) = \begin{pmatrix} \beta \\ 1 \end{pmatrix} \quad \mathbf{v}'(\tilde{x}) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \quad (37)$$

Following equation (15) we obtain

$$\mathbf{P}_1(\tilde{x}) = \frac{(\tilde{x} + \alpha)^2}{\alpha + \beta(\tilde{x} + \alpha)^2} \begin{pmatrix} \beta & 1 \end{pmatrix}, \quad (38)$$

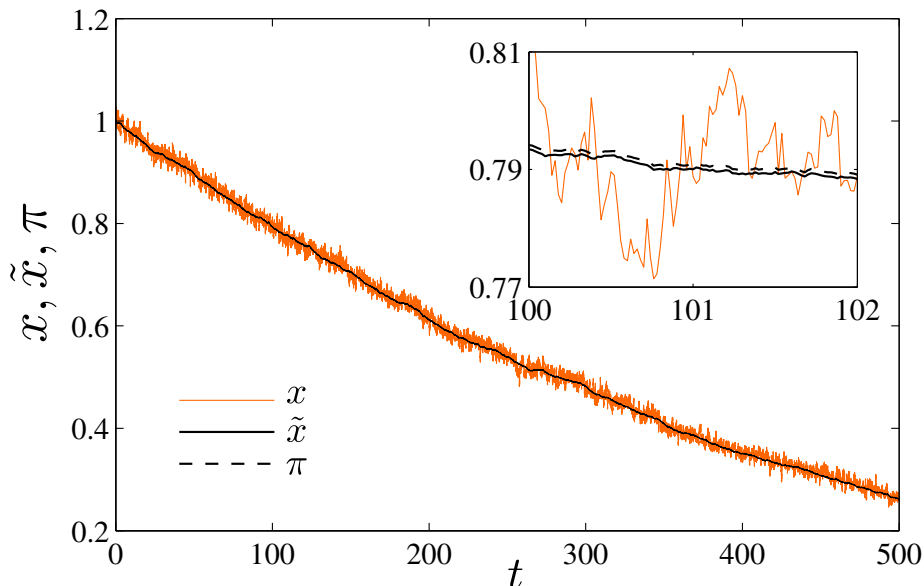
and from equation (16)

$$\mathbf{Q}_1(\tilde{x}) = 2\alpha \left( \frac{(\tilde{x} + \alpha)}{\alpha + \beta(\tilde{x} + \alpha)^2} \right)^3 \begin{pmatrix} \beta^2 & \beta \\ \beta & 1 \end{pmatrix}. \quad (39)$$

where  $\mathbf{P}_1$  and  $\mathbf{Q}_1$  indicate the vector (resp. matrix) obtained by fixing the first coordinate at the value 1. Plugging these results into the general formula (7) gives the reduced model

$$\frac{d\tilde{x}}{dt} = -\varepsilon \frac{\tilde{x}(\tilde{x} + \alpha)}{\alpha + \beta(\tilde{x} + \alpha)^2} + \varepsilon\mu \frac{\alpha\beta\tilde{x}(\tilde{x} + \alpha)^2}{(\alpha + \beta(\tilde{x} + \alpha)^2)^3} - \frac{(\tilde{x} + \alpha)^2}{\alpha + \beta(\tilde{x} + \alpha)^2} \sqrt{\varepsilon\mu \frac{\beta\tilde{x}}{\tilde{x} + \alpha}} \eta_{\text{cat}}(t). \quad (40)$$

Figure 4 shows the dynamics of  $\tilde{x}$  compared with those of  $x_1$  in the full system for a single realisation of the noise. At first sight equation (40) is considerably more complex than the traditional Michaelis-Menten law, however, carefully transforming back to the



**Figure 4.** Trajectories of  $x_1$ ,  $\tilde{x}$  and  $\pi_1(x)$  from a single stochastic simulation of the Michaelis-Menten model (29), simulated using the Euler-Maryuama method (see e.g. [38]). The inset shows details of the fluctuations near the point (100, 0.8). Note that the reduced dimension model for  $\tilde{x}$  given by equation (40) captures the dynamics of the full system under the projection  $\pi$  (hence the extremely close agreement between the solid and dashed black lines above). The original coordinate  $x_1$  is subject to additional noise in the kernel of the projection. Parameters are  $k_f = 20$ ,  $k_r = 20$ ,  $k_{\text{cat}} = 0.05$ ,  $V = 1000$ ,  $S_0 = 1$ ,  $E_0 = 0.5$ .

original coordinates, we will find a simple result. First for the dynamics of  $S = S_0 \tilde{x}$ , recalling the time change  $t \mapsto t/k_f E_0$  and returning to the original parameter names, we obtain from (40) the SDE

$$\frac{dS}{dt} = -\frac{v^* S(k+S)}{k_e + (k+S)^2} + \frac{v^* k_e S(k+S)^2}{(k_e + (k+S)^2)^3 V} - \frac{(k+S)^2}{k_e + (k+S)^2} \sqrt{\frac{v^* S}{V(k+S)}} \eta_{\text{cat}}(t) \quad (41)$$

where  $v^* = k_{\text{cat}} E_0$ ,  $k = k_r/k_f$  and  $k_e = E_0 k$ . Next for  $C = E_0 \tilde{x}/(\tilde{x} + \alpha)$ , applying Itô's lemma to (40) gives

$$\frac{dC}{dt} = -\frac{v^* k_e S}{(k+S)(k_e + (k+S)^2)} - \frac{v^* k_e S(k+S)^2}{(k_e + (k+S)^2)^3 V} - \frac{k_e}{k_e + (k+S)^2} \sqrt{\frac{v^* S}{V(k+S)}} \eta_{\text{cat}}(t). \quad (42)$$

Finally, from the conservation rule  $S + C + P = S_0$  we can combine (41) and (42) to obtain the simple form

$$\frac{dP}{dt} = \frac{v^* S}{k+S} + \sqrt{\frac{v^* S}{V(k+S)}} \eta_{\text{cat}}(t). \quad (43)$$

Applying our constructive approach to Katzenberger's results, we have thus obtained a detailed description of the system dynamics, which takes into account the non-



perpendicular nature of the projection onto the slow manifold. This is in contrast to a naïve quasi-steady state approximation, where one would simply substitute  $C = S/(k + S)$  into (29). The difference between these approaches may explain some of the issues with the quasi-steady state approximation raised in [39, 40].

3.2. Co-dimension one: the Wright-Fisher diffusion as a limit of a near-neutral stochastic Lotka-Volterra process

Consider a well mixed-population of  $d$  interacting species in an environment of carrying capacity  $K$ : there are  $K$  “slots” in the environment that at most one individual may occupy. Let  $X_i$  denote the number of individuals of species  $i$ , and suppose that each individual of species  $i$  gives birth at rate  $b_i$  and dies at rate  $d_i$ . Further, suppose that the offspring is only viable if it lands in an empty patch, or if it lands in an occupied patch and out-competes the resident; say that an individual of type  $i$  successfully displaces a resident of type  $j$  with probability  $c_{ij}$ . Then, there are three types of events:

- (i)  $X_i$  increases by 1 at rate  $b_i X_i \left(1 - \frac{\sum_j X_j}{K}\right)$ ,
- (ii)  $X_i$  decreases by 1 at rate  $d_i X_i$ , or,
- (iii)  $X_i$  increases by 1 and  $X_j$  decreases by 1 at rate  $b_i X_i \left(\frac{c_{ij} X_j}{K}\right)$ .

This gives a stochastic model of a population with density-dependent competition; *n.b.* the total population size is *not* fixed at  $K$ , but is rather allowed to fluctuate stochastically with an upper bound of  $K$ , as we allow the possibility of empty slots in the environment.

Let  $x_i(t)$  denote the density of species  $i$  (*i.e.*  $\frac{X_i(t)}{K}$ ). As in the previous section, this system may be approximated by a system of stochastic differential equations,

$$\begin{aligned} \frac{dx_i}{dt} = & \left( (b_i - d_i) - \sum_j (b_i - b_i c_{ij} + b_j c_{ji}) x_j \right) x_i \\ & + \sqrt{\frac{b_i x_i (1 - \sum_j x_j)}{K}} \eta_{b,i}(t) - \sqrt{\frac{d_i x_i}{K}} \eta_{d,i}(t) \\ & + \sum_j \sqrt{\frac{b_i c_{ij} x_i x_j}{K}} \eta_{i,j}(t) - \sum_j \sqrt{\frac{b_j c_{ji} x_i x_j}{K}} \eta_{j,i}(t). \end{aligned}$$

We will be interested in finding a non-trivial diffusion when  $K$  is very large, and thus choose our small parameter is  $\mu = \frac{1}{K}$ . To explore the link between population genetics and population dynamics, we will further postulate that there exist values  $\epsilon_1, \dots, \epsilon_d$  so that

$$b_i = b \left(1 + \frac{\epsilon_i}{K}\right), \quad d_i = d + \frac{\nu_i}{K}, \quad \text{and} \quad c_{ij} = c + \frac{a_{ij}}{K},$$

for all  $i, j$ ; this corresponds to the weak selection hypothesis of classical population genetics [41], intended to capture the often small effect of point mutations: all species

differ in their demographic rates and their ability to compete for sites by terms of  $O\left(\frac{1}{K}\right)$ . Then,  $\varepsilon = \frac{1}{K}$ ,

$$f_i(\mathbf{x}) = x_i \left( (b-d) - b \sum_{j=1}^K x_j \right),$$

and

$$h_i(\mathbf{x}) = x_i \left( (b\varepsilon_i - d\nu_i) - b \sum_j ((1-c)\varepsilon_i - c\varepsilon_j - a_{ij} + a_{ji}) x_j \right)$$

Under these assumptions,

$$\Gamma = \left\{ \mathbf{x} \in \mathbb{R}^d : \sum_{j=1}^K x_j = 1 - \frac{d}{b} \right\}$$

and for  $\mathbf{x} \in \Gamma$ , the derivatives (23) and (24) simplify to

$$P_{ij}(\mathbf{x}) = \delta_{ij} - \frac{x_i}{1 - \frac{d}{b}} \quad \text{and} \quad Q_{ijk}(\mathbf{x}) = -\frac{1}{1 - \frac{d}{b}} \left( \delta_{ij} + \delta_{ik} - \frac{2x_i}{1 - \frac{d}{b}} \right),$$

whereas for  $\mathbf{x} \in \Gamma$ ,

$$h_i(\mathbf{x}) = x_i \left( d(\varepsilon_i - \nu_i) + c \sum_j (\varepsilon_i - \varepsilon_j) x_j + b \sum_j (a_{ij} - a_{ji}) x_j \right).$$

A straightforward if lengthy calculation shows that  $\mathbf{g}(\mathbf{x}) = O\left(\frac{1}{K^2}\right)$ .

Substituting into our general formula (7) then gives

$$\begin{aligned} \frac{d\tilde{x}_i}{dt} &= \frac{1}{K} \left( h_i(\tilde{\mathbf{x}}) - \frac{\tilde{x}_i}{1 - \frac{d}{b}} \sum_j h_j(\tilde{\mathbf{x}}) \right) \\ &+ \sum_j \left( \delta_{ij} - \frac{\tilde{x}_i}{1 - \frac{d}{b}} \right) \left( \sqrt{\frac{d\tilde{x}_j}{K}} (\eta_{b,j}(t) - \eta_{d,j}(t)) + \sum_k \sqrt{\frac{bc\tilde{x}_j\tilde{x}_k}{K}} (\eta_{j,k}(t) - \eta_{k,j}(t)) \right), \end{aligned}$$

or, changing variables to  $p_i = \frac{\tilde{x}_i}{1 - \frac{d}{b}}$ , so  $p_i$  is the proportion of species  $i$ ,

$$\begin{aligned} \frac{dp_i}{dt} &= \frac{1}{K} p_i \left( s_i(\mathbf{p}) - \sum_j s_j(\mathbf{p}) p_j \right) \\ &+ \sum_j (\delta_{ij} - p_i) \left( \frac{1}{\sqrt{1 - \frac{d}{b}}} \sqrt{\frac{dp_j}{K}} (\eta_{b,j}(t) - \eta_{d,j}(t)) + \sum_k \sqrt{\frac{bc p_j p_k}{K}} (\eta_{j,k}(t) - \eta_{k,j}(t)) \right), \end{aligned}$$

where

$$s_i(\mathbf{p}) = d(\varepsilon_i - \nu_i) + c \left( 1 - \frac{d}{b} \right) \sum_j (\varepsilon_i - \varepsilon_j) p_j + (b-d) \sum_j (a_{ij} - a_{ji}) p_j.$$

The corresponding Fokker-Planck equation for the density  $f(\mathbf{p}, t)$  is then

$$\frac{\partial f}{\partial t} = -\frac{1}{K} \frac{\partial}{\partial p_i} \left[ p_i \left( s_i(\mathbf{p}) - \sum_j s_j(\mathbf{p}) p_j \right) f \right] + \frac{1}{2} \frac{2 \left( bc + \frac{d}{1-\frac{d}{b}} \right)}{K} \frac{\partial^2}{\partial p_i \partial p_j} [p_i (\delta_{ij} - p_j) f]$$

which we recognise as the equation for the Wright-Fisher diffusion, where the (frequency dependent) selection coefficient is  $\frac{s_i(\mathbf{p})}{K}$  and the effective population size is  $N_e = \frac{(1-\frac{d}{b})K}{2(c(b-d)+d)}$ ;  $(1 - \frac{d}{b}) K$  is the population size at the deterministic equilibrium, whereas the other terms reflect variance in the total population size. This gives an alternate derivation of the results presented in [9, 10, 12, 20].

### 3.3. Continuous degrees of freedom: example of competition-limited diffusion

The methods of Section 2 can readily be extended to infinite dimensional settings. Two recent examples come from work exploring the role of stochasticity in spatial ecological models [14, 19]. Here we work through a simple illustrative example of diffusing particles coupled by a competitive birth-death interaction; we will show that this competition acts to limit the speed of diffusion of the population. Interested readers are referred to [42], where the continuum limit of this example has been studied in considerable depth.

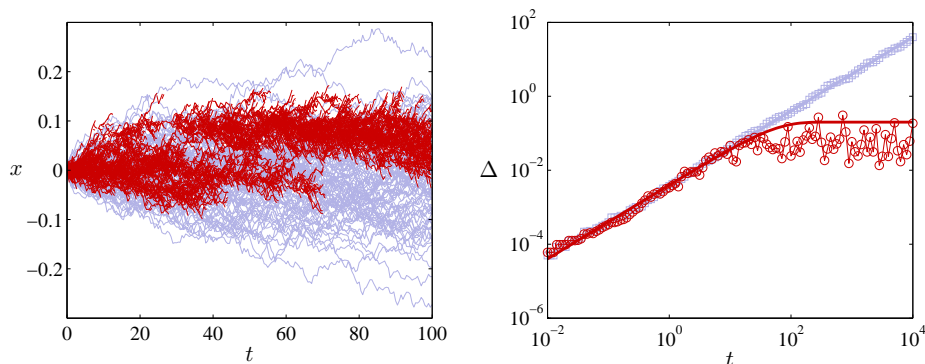
Consider the following stochastic process. At time  $t$  there are  $N(t)$  individual particles wandering in a one-dimensional space, each following their own Brownian motion with diffusion constant  $D = \sqrt{2\varepsilon}$ . With rate one, each particle may independently ‘reproduce’, creating a daughter particle that initially shares the location of the parent, but thereafter moves independently. Particles ‘die’ with rate proportional to their total number; specifically, the death rate for each particle is  $\mu(N(t) - 1)$ . We assume the constants  $\mu$  and  $\varepsilon$  are small, but of the same order.

Since the location of the particles does not influence the birth or death rates, it is easy to see that the total number of particles follows a logistic growth law, quickly reaching an equilibrium  $N(t) \approx \mu^{-1}$ . The total population size remains at this level while the spatial distribution of particles evolves slowly over a much longer timescale. We are interested in the long-term behaviour of the distribution of particle locations. Introduce the population density

$$u(x) = \mu \sum_{n=1}^{N(t)} \delta(x - X_n(t)) , \tag{44}$$

where  $X_n$  is the location of particle  $n$  at time  $t$ ,  $\delta$  is the Dirac delta function, and we suppress the dependence of  $u$  on  $t$  to reduce clutter. Simulations suggest that the competitive interaction of the particles limits the extent to which they are able to diffuse away from each other (Figure 5, left panel). This observation can be made quantitative by computing the mean square distance between pairs of particles,

$$\Delta[u] := \mu^2 \sum_{n,m} (X_n(t) - X_m(t))^2 = \iint (x - y)^2 u(x) u(y) dx dy . \tag{45}$$



**Figure 5.** Simulation of competition-limited diffusion (dark red), contrasted with a collection of  $N$  independent Brownian particles (light purple). The left panel shows the particle trajectories, on the right is shown the mean square distance between pairs of particles, together with the analytical approximation to  $\mathbb{E}\Delta[v]$ , (62). Parameters are  $\varepsilon = 0.05$ ,  $\mu = 0.01$ .

The right panel of Figure 5 shows the time evolution of  $\Delta$  for the population, compared to the growth  $\Delta \sim t$  observed for independent diffusing particles. The solid lines show our theoretical prediction for this phenomenon, which we will now derive using timescale separation.

Following a system-size expansion [8], we find that the time-evolution of  $u(x)$  is described to close approximation by the stochastic partial differential equation (SPDE)

$$\frac{\partial}{\partial t} u(x) = \varepsilon \frac{\partial^2}{\partial x^2} u(x) + u(x) \left( 1 - \int u(y) dy \right) + \sqrt{\mu u(x) \left( 1 + \int u(y) dy \right)} \eta(x, t), \quad (46)$$

where  $\eta(x, t)$  is spatio-temporal white noise and the integrals run over the real line.

Equation (46) has the same essential structure as our basic object of interest (1). If we identify

$$\begin{aligned} f[u](x) &= u(x) \left( 1 - \int u(y) dy \right) \\ h[u](x) &= \frac{\partial^2}{\partial x^2} u(x) \\ G[u](x, s) &= \delta(x - s) \sqrt{u(x) \left( 1 + \int u(y) dy \right)}, \end{aligned} \quad (47)$$

then (48) becomes

$$\frac{\partial}{\partial t} u(x) = f[u](x) + \varepsilon h[u](x) + \sqrt{\mu} \int G[u](x, s) \eta(s, t) ds. \quad (48)$$

The integral here is the analogue of the matrix-vector multiplication  $\mathbf{G}(\mathbf{x})\boldsymbol{\eta}(t)$  appearing in (1). The delta function appearing in  $G[u]$  means that the noise in our example is spatially uncorrelated; this may not hold for other models.

In this section we will show how the timescale separation techniques discussed above may also be applied to equations of the form (48), applying the techniques without rigorous justification.

First we examine the outer part  $\partial u/\partial t = f[u]$ . In our example, the PDE

$$\frac{\partial}{\partial t}u(x) = u(x) \left( 1 - \int u(y) dy \right), \quad (49)$$

is straightforward to solve:

$$u(x, t) = \frac{u(x, 0) e^t}{1 + (e^t - 1) \int u(y, 0) dy}, \quad (50)$$

which describes the fast relaxation of  $u$  to a state in which it has total mass one. In this infinite-dimensional setting, the map that describes the long-time limit of the outer solution (previously defined in (4)) is an operator  $\pi$ , whose action is specified by

$$\pi[u](x) = \frac{u(x)}{\int u(y) dy}. \quad (51)$$

We suppose that there exists a suitable space of functions  $\mathcal{U}$  describing possible solutions of (48). Exactly what kind of space is a deep question beyond our present focus. The analogue of the slow manifold is the subspace  $\mathcal{V} \subset \mathcal{U}$  containing functions  $v$  satisfying  $f[v] = 0$ , or equivalently for our example,  $\int v(y) dy = 1$ . We aim to derive an equation describing slow stochastic evolution in  $\mathcal{V}$  that well-approximates the behaviour of solutions to the full system (48).

Where previous calculations involved partial differentiation, we now apply a functional derivative. In analogue to the definitions in (8) we introduce

$$P[v](x, y) = \left. \frac{\delta}{\delta u(y)} \pi[u](x) \right|_{u=v}, \quad Q[v](x, y, z) = \left. \frac{\delta^2}{\delta u(y) \delta u(z)} \pi[u](x) \right|_{u=v}. \quad (52)$$

The reduced system may then be written down:

$$\begin{aligned} \frac{\partial}{\partial t}v(x) = & \int P[v](x, y) \left[ \varepsilon h(y) dy + \sqrt{\mu} \int G[v](y, s) \eta(s, t) ds \right] \\ & + \frac{\mu}{2} \iiint G[v](y, s) G[v](z, s) Q[v](x, y, z) dy dz ds. \end{aligned} \quad (53)$$

For the example at hand we compute

$$\begin{aligned} \frac{\delta}{\delta u(y)} \pi[u](x) &= \frac{\delta(x-y)}{\int u(z) dz} - \frac{u(x)}{(\int u(z) dz)^2}, \\ \frac{\delta^2}{\delta u(y)^2} \pi[u](x) &= \frac{2u(x)}{(\int u(z) dz)^3} - \frac{2\delta(x-y)}{(\int u(z) dz)^2}, \end{aligned} \quad (54)$$

and thus

$$P[v](x, y) = \delta(x-y) - v(x), \quad Q[v](x, y, y) = 2v(x) - 2\delta(x-y). \quad (55)$$

Note that we only need the  $z = y$  parts of  $Q[v](x, y, z)$  because of the delta function in  $G$ . Plugging (47) and (55) into (53), we obtain the reduced model

$$\frac{\partial}{\partial t}v(x) = \varepsilon \frac{\partial^2}{\partial x^2}v(x) + \sqrt{2\mu} \int [\delta(x - y) - v(x)] \sqrt{v(y)} \eta(y, t) dy. \quad (56)$$

Comparing (56) to the original equation (46) we see two main differences: the non-linearity in the drift has vanished, but the noise is now spatially coupled.

To compute a prediction for the mean squared distance between particles, it is simpler to work in Fourier space. Introducing  $\tilde{v}(k) = \int e^{-2\pi i k x} v(x) dx$ , we note first that

$$\mathbb{E}\Delta[v] = \iint z^2 e^{2\pi i k z} \mathbb{E}|\tilde{v}(k)|^2 dk dz = -\frac{1}{4\pi^2} \frac{\partial^2}{\partial k^2} \mathbb{E}|\tilde{v}(k)|^2 \Big|_{k=0} \quad (57)$$

Translating (56) to Fourier space we find

$$\frac{\partial}{\partial t} \tilde{v}(k) = -4\varepsilon\pi^2 k^2 \tilde{v}(k) + \sqrt{2\mu} \int \tilde{G}[\tilde{v}](k, x) \eta(x, t) dx, \quad (58)$$

where

$$\tilde{G}[\tilde{v}](k, x) = (e^{-2\pi i k x} - \tilde{v}(k)) \sqrt{\int e^{2\pi i \ell x} \tilde{v}(\ell) d\ell}. \quad (59)$$

In mean, this process behaves exactly as a straightforward diffusion:

$$\frac{d}{dt} \mathbb{E}[\tilde{v}(k)] = -4\varepsilon\pi^2 k^2 \mathbb{E}[\tilde{v}(k)]. \quad (60)$$

However, the noise introduces a correction to the variance following Itô's formula. Specifically,

$$\begin{aligned} \frac{d}{dt} \mathbb{E}|\tilde{v}(k)|^2 &= -8\varepsilon\pi^2 k^2 \mathbb{E}|\tilde{v}(k)|^2 + \frac{1}{2} \iiint \tilde{G}[\tilde{v}](\ell, x) \tilde{G}[\tilde{v}](m, x) \frac{\delta^2 |\tilde{v}(k)|^2}{\delta \tilde{v}(\ell) \delta \tilde{v}(m)} dx d\ell dm \\ &= -8\varepsilon\pi^2 k^2 \mathbb{E}|\tilde{v}(k)|^2 + 2\mu(1 - \mathbb{E}|\tilde{v}(k)|^2). \end{aligned} \quad (61)$$

Solving (61) and plugging into (57) gives the prediction

$$\mathbb{E}\Delta[v] = \frac{2\varepsilon}{\mu} (1 - e^{-2\mu t}). \quad (62)$$

This result is shown as the dark red curve in Figure 5. In particular, notice that whilst the mean square distance between diffusing particles grows indefinitely, in the competition coupled process it attains a finite limit  $2\varepsilon/\mu$ .

#### 4. Discussion

The purpose of this article has been to show the derivation and application of a systematic computational framework for dimension reduction in stochastic dynamical

systems that exhibit a separation of timescales via a globally stable normal hyperbolic slow manifold *i.e.* in the limit of small noise the limiting deterministic dynamical system defined by  $\mathbf{f}$  possesses a single, connected and globally attractive manifold of fixed points. The method is exact in the limit of small noise and well-separated slow and fast dynamics, and experimentally found to be valid as an approximation scheme over a sensible parameter range. We have also presented extensions of the method for infinite dimensional systems and processes coupled to general noise sources.

In some applications more general scenarios may occur, we now briefly discuss two of interest. Some models may exhibit more than one connected manifold of equilibria or dynamic bifurcations, *i.e.*, points where the critical manifold ceases to be normally hyperbolic [43]; in this case the theory developed here will apply locally to trajectories in the basin of attraction of each manifold individually, but further analysis will be necessary to describe the statistics of noise-driven transitions between manifolds. A possibly more exciting direction for further research is the analysis of noisy behaviour around more general attractors such as limit cycles, limit tori and strange attractors. In the case of limit cycles some work exists on stochastic extensions to Floquet theory [44], however, this is a linear description that cannot capture any bias analogous to the noise-induced drift in the slow manifold setting.

Finally, it is worth returning to discuss the motivation for this work. As mentioned earlier, variations of the work of Katzenberger have been independently rediscovered by several groups in recent years, almost all of whom have been interested in questions about the role of noise in ecology and evolution. Historically, many theoretical results in this field have been derived from models that assume for convenience a fixed population size. In the deterministic limit this assumption is not important, but we are now beginning to realise that the inclusion of noise can induce radically different and sometimes unexpected behaviour. Mathematically, this is a consequence of the noise-induced drift term  $\mathbf{g}$  that appears in our equation (2), and more generally of the seemingly endless capacity of Itô's lemma to cause surprise. There have been some tentative explorations of the possible evolutionary and ecological consequences of these effects [12, 45, 46], but much more is yet to be discovered.

## Acknowledgements

We thank George Constable, Alan McKane and Christopher Quince for useful discussions and Tom Kurtz for bringing [28] to our attention. Work on this paper began when both TLP and TR were visitors at the Isaac Newton Institute for Mathematical Sciences programme *Understanding Microbial Communities; Function, Structure and Dynamics*. TLP is supported by the CNRS, and TR by the Royal Society.

### Appendix A. Katzenberger’s Theorem

Above we developed our results in the context of Itô SDEs, however, [28] proved a more general result that allows us to consider a much broader class of noise processes: semimartingales. Semimartingales are the most general class of stochastic processes for which one may define a stochastic integral and stochastic differential equations (Brownian motion is included as a special case). Suitably adapted, most of the familiar results for SDEs and white-noise integrals, including Itô’s formula, remain true in the more general setting [47].

To define a semimartingale, we must first make a few auxiliary definitions. A Markov process  $M(t)$  is a *martingale* if

$$\mathbb{E} [M(t)|M(s)] = M(s).$$

A random variable  $\tau$  taking values in  $[0, \infty)$  is a *stopping time* if one can determine if  $\tau < t$  without knowledge of the future beyond  $t$ ; an example of a stopping time is the first time a diffusion started from 0 exits an interval  $[-a, a]$ .  $M(t)$  is a *local martingale* if there is a sequence of stopping times  $\tau_n \rightarrow \infty$  such that  $M(\min\{t, \tau_n\})$  is a martingale for each  $n$ .

A function is *càdlàg* if it is continuous from the right and has left-hand limits at every point.

The *total variation* of a function  $f$  on an interval  $[a, b]$  is

$$V_b^a(f) = \min_{\{t_i\}} \sum_i |f(t_{i+1}) - f(t_i)|,$$

where the minimum is over all partitions  $a = t_0 < t_1 < \dots < t_n = b$  of  $[a, b]$ . A stochastic process  $A(t)$  is of *finite variation* if it is càdlàg and has finite total variation on all intervals  $[a, b]$  (note that  $A(t)$  is allowed to have jump discontinuities).

Finally,  $Z(t)$  is a semimartingale if it may be written as the sum of a local martingale and a finite variation process,

$$Z(t) = M(t) + A(t).$$

Diffusion processes are the prototypical example of semimartingales, but the class is much broader, and includes processes with jumps, such as Lévy processes; *e.g.* if  $N(t)$  is a Poisson process, then  $M(t) = N(t) - t$  is a local martingale and  $A(t) = t$  is of finite variation, so  $N(t)$  is a semimartingale. Integration with respect to a semimartingale is defined analogously to the Stieltjes integral, except that we require the approximating sum to converge in probability, and, as with the Itô integral, the integrand is always evaluated at the left endpoint of each interval in the partition.

More generally, we can define vector and matrix valued martingales, local martingales, finite variation processes and semimartingales,  $\mathbf{M}(t)$ ,  $\mathbf{A}(t)$ , and  $\mathbf{Z}(t)$ , by requiring the components,  $M_i(t)$  *etc.*, have the corresponding property.

We can now formulate Katzenberger’s result. Let



- (i)  $\mathbf{Z}_n(t)$  be a convergent sequence of vector valued semimartingales such that the jumps  $\Delta \mathbf{Z}_n(t) \rightarrow 0$  as  $n \rightarrow \infty$ ,
- (ii)  $A_n(t)$  be a sequence of non-decreasing finite variation processes such that  $\Delta A_n(t) \rightarrow 0$ , and

$$\int_a^b dA_n(s) = A_n(b) - A_n(a) \rightarrow \infty$$

as  $n \rightarrow \infty$ ; Katzenberger notes that most frequently in applications,  $A_n(t) = \alpha_n t$  for some sequence  $\alpha_n \rightarrow \infty$  (*n.b.*, in this formulation, this explosion in  $A_n(s)$  corresponds to the drift becoming infinitely strong, rather than the noise infinitely weak, as in (1). The two are equivalent, if one changes the timescale accordingly; recall we had

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}) + \varepsilon \mathbf{h}(\mathbf{x}) + \sqrt{\mu} \mathbf{G}(\mathbf{x}) \boldsymbol{\eta}(t).$$

If instead, we consider the homogenised process  $\tilde{\mathbf{x}}(t) = \mathbf{x}(\mu^{-1}t)$ , we get

$$\frac{d\tilde{\mathbf{x}}}{dt} = \frac{1}{\mu} \mathbf{f}(\tilde{\mathbf{x}}) + \frac{\varepsilon}{\mu} \mathbf{h}(\tilde{\mathbf{x}}) + \mathbf{G}(\tilde{\mathbf{x}}) \boldsymbol{\eta}(t),$$

with a drift that blows up as  $\mu \rightarrow 0$ ).

- (iii)  $\mathbf{f}$  and  $\Gamma$  be as before,
- (iv)  $\mathbf{G}_n(\mathbf{x})$  be a sequence of matrix-valued functions converging to a limit  $\mathbf{G}(\mathbf{x})$ , and
- (v)  $\mathbf{x}_n(t)$  be a sequence of stochastic processes satisfying the (semimartingale) SDE

$$d\mathbf{x}_n = \mathbf{f}(\mathbf{x}_n) dA_n + \mathbf{G}_n(\mathbf{x}_n) d\mathbf{Z}_n. \tag{A.1}$$

Then, as before, subject to a few technical considerations, as  $n \rightarrow \infty$ ,  $\mathbf{x}_n$  converges to a diffusion process on  $\Gamma$  satisfying

$$\frac{dz}{dt} = \mathbf{g}(z) + \mathbf{P}(z) \mathbf{G}(z) \boldsymbol{\eta}(t), \tag{A.2}$$

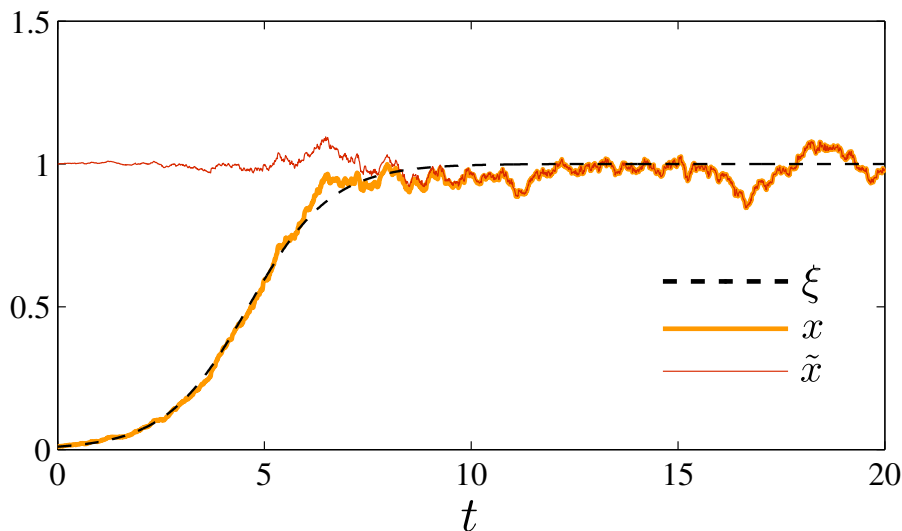
where  $\mathbf{g}$  is as in equation (9) and  $\boldsymbol{\eta}$  is white noise.

Some care is required in understanding the sense of convergence in [28]; if  $\mathbf{x}_n(0)$  converges weakly to  $\mathbf{z} \in \Gamma$  in  $\mathbb{R}^d$  (*i.e.*, for all continuous functions  $F : \mathbb{R}^d \rightarrow \mathbb{R}$ ,  $\mathbb{E}[F(\mathbf{x}_n(0))] \rightarrow \mathbb{E}[F(\mathbf{z})]$ ) then  $\mathbf{x}_n(t)$  converges weakly to  $\mathbf{z}(t)$  in the space of càdlàg functions:

$$\mathbb{E}[F(\mathbf{x}_n(t))] \rightarrow \mathbb{E}[F(\mathbf{z}(t))]$$

for all continuous functions  $F$  from the space of càdlàg functions on  $[0, \infty)$  to  $\mathbb{R}$  (see [48, 49] for a definition of the topology on càdlàg functions and results on weak convergence). When  $\mathbf{x}_n(0)$  converges to a limit  $\mathbf{x}$  that is in the basin of attraction of  $\Gamma$ , but not in  $\Gamma$ , additional care is required: in this case, the process will jump instantaneously from  $\mathbf{x}$  to  $\boldsymbol{\pi}(\mathbf{x}) \in \Gamma$ , which is not compatible with convergence in the weak topology on càdlàg functions. However, if one considers

$$\hat{\mathbf{x}}_n(t) = \mathbf{x}_n(t) - \boldsymbol{\xi}(A_n(t)) + \boldsymbol{\pi}(\mathbf{x}),$$



**Figure A1.** Illustration of  $\tilde{x}_1$  on the fast timescale for a prototypical stochastic dynamical system with a slow manifold  $\Gamma = \{\mathbf{x} : x_1 = 1\}$ . In the slow timescale the initial transit to the manifold is compressed into an instantaneous jump at  $t = 0$ . For reference, the system used is  $\dot{x}_1 = x_1(1 - x_1) + \sqrt{\mu x_1(1 + x_1)}\eta(t)$ ,  $\dot{x}_2 = x_2(1 - x_2)$  with  $\mu = 0.002$ ,  $x_{1,2}(0) = 0.01$ .

(recall,  $\boldsymbol{\xi}(t)$  is the solution to the outer system, (3)) then  $\hat{\mathbf{x}}_n(0) \rightarrow \boldsymbol{\pi}(\mathbf{x}) \in \Gamma$  and  $\hat{\mathbf{x}}_n(t)$  converges weakly to the diffusion  $\mathbf{z}(t)$  on  $\Gamma$  as before; intuitively  $\hat{\mathbf{x}}_n(t)$  is obtained by removing the initial transient phase when  $\mathbf{x}_n(t)$  follows the trajectories of the outer system, and starting the process instead from the endpoint of that trajectory,  $\boldsymbol{\pi}(\mathbf{x})$  (see Figure A1).

The term  $\boldsymbol{\xi}(A_n(t))$  makes explicit the time scale change that is only implicit in (A.1): for simplicity, consider briefly the case when  $A_n(t)$  is differentiable and  $\mathbf{Z}_n(t)$  is identically zero. Then, if  $\boldsymbol{\xi}(t)$  solves  $d\boldsymbol{\xi} = \mathbf{f}(\boldsymbol{\xi}) dt$ ,

$$\boldsymbol{\xi}(A_n(t)) = \int_0^{A_n(t)} \mathbf{f}(\boldsymbol{\xi}(u)) du = \int_0^t \mathbf{f}(\boldsymbol{\xi}(A_n(u))) A_n'(u) du = \int_0^t \mathbf{f}(\boldsymbol{\xi}(A_n(u))) dA_n(u),$$

so, defining  $\mathbf{x}_n(t) = \boldsymbol{\xi}(A_n(t))$ , we have  $d\mathbf{x}_n = \mathbf{f}(\mathbf{x}_n) dA_n$  i.e., (A.1) is describing the evolution of a homogenized process in the slow time scale  $A_n(t)$ . In particular, when  $A_n(t) = \alpha_n t$ , the middle integral above gives us

$$\mathbf{x}_n(t) = \alpha_n \int_0^t \mathbf{f}(\mathbf{x}_n(u)) du,$$

by which we see explicitly how a large drift is absorbed into a rescaled time.

#### Appendix A.1. Density dependent population processes

While Katzenberger's result might seem unnecessarily abstract, it allows one to apply the same slow-manifold reduction to a number of individual-based, discrete

stochastic processes that include a number of well-known examples from applications. In [50, 51, 1, 52], Kurtz introduced and studied what he called density dependent population processes. While his original motivation was chemical reaction networks, the class also includes many examples of interest in biology and epidemiology.

A sequence of Markov processes  $\mathbf{x}_n(t)$  is a *density dependent population process* if  $\mathbf{x}_n$  takes values in  $\frac{1}{n}\mathbb{Z}^d$ , and, if  $q_{\mathbf{x},\mathbf{y}}^{(n)}$  is the jump rate between  $\mathbf{x}, \mathbf{y} \in \frac{1}{n}\mathbb{Z}^d$ , then

$$q_{\mathbf{x},\mathbf{y}}^{(n)} = n\lambda_{n(\mathbf{y}-\mathbf{x})}(\mathbf{x})$$

for some non-negative function  $\lambda_{\mathbf{l}}(\mathbf{x})$  on  $\mathbb{R}^d$ , where  $\mathbf{l} = n(\mathbf{y} - \mathbf{x}) \in \mathbb{Z}^d$ . More generally, one can consider the case of functions  $\lambda_{\mathbf{l}}^{(n)}(\mathbf{x})$  that depend on  $n$ , provided  $\lambda_{\mathbf{l}}^{(n)}(\mathbf{x})$  converges to a limit  $\lambda_{\mathbf{l}}(\mathbf{x})$  sufficiently quickly as  $n \rightarrow \infty$ ; see [53].

The parameter  $n$  corresponds to the “system size” in [54], and can be interpreted differently according to the context, as *e.g.* total population size, area, or volume. For example, consider the stochastic logistic process  $X_n(t)$  with birth and death rates

$$Q_{X,X+1}^{(n)} = \beta X \left(1 - \frac{X}{n}\right) \quad Q_{X,X-1}^{(n)} = \delta X.$$

Here,  $n$  plays the role of the carrying capacity in the deterministic logistic equation, *i.e.* the number of individuals the environment can support: individuals have an intrinsic per-capita birth rate  $\beta$ , but the offspring will only survive if it arrives in an unoccupied spot in the habitat. Nondimensionalising, we might consider instead the process  $x_n(t) = \frac{1}{n}X_n(t)$ , with rates

$$q_{x,x+\frac{1}{n}}^{(n)} = n\beta x(1-x) \quad q_{x,x-\frac{1}{n}}^{(n)} = n\delta x.$$

The latter is an example of a density-dependent population process, with

$$\lambda_1(x) = \beta x(1-x) \quad \lambda_{-1}(x) = \delta x.$$

In [50], Kurtz shows that provided

$$\sum_{\mathbf{l} \in \mathbb{Z}^d} \|\mathbf{l}\| \sup_{\mathbf{x} \in \mathcal{K}} \lambda_{\mathbf{l}}(\mathbf{x}) < \infty$$

for all closed and bounded sets  $\mathcal{K}$ , then if

$$\mathbf{f}(\mathbf{x}) = \sum_{\mathbf{l} \in \mathbb{Z}^d} \mathbf{l} \lambda_{\mathbf{l}}(\mathbf{x})$$

is differentiable and  $\mathbf{x}_n(t) \rightarrow \mathbf{x}_0$ , then for any fixed  $T > 0$ ,

$$\lim_{n \rightarrow \infty} \sup_{t \leq T} |\mathbf{x}_n(t) - \mathbf{x}(t)| = 0,$$

where  $\mathbf{x}(t)$  is the solution of  $\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x})$  with  $\mathbf{x}(0) = \mathbf{x}_0$ .

If one assumes that  $\lambda_i(\mathbf{x})$  is non-zero for only finitely many transitions, say  $\mathbf{l}_1, \dots, \mathbf{l}_s$ , then, letting  $\mathbf{G}(\mathbf{x})$  be the matrix with  $i^{\text{th}}$  column  $\mathbf{l}_i \sqrt{\lambda_i(\mathbf{x})}$ ,  $\boldsymbol{\eta}(t)$  be an  $s$ -dimensional Itô white noise, and  $\mathbf{z}_n(t)$  be the solution of

$$\frac{d\mathbf{z}_n}{dt} = \mathbf{f}(\mathbf{z}_n) + \frac{1}{\sqrt{n}} \mathbf{G}(\mathbf{z}_n) \boldsymbol{\eta}(t),$$

then for any fixed  $T > 0$ , there exists a constant  $C_T$  such that

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \sup_{t \leq T} |\mathbf{x}_n(t) - \mathbf{z}_n(t)| > \frac{C_T \log n}{n} \right) = 0.$$

In our current setting, if  $\mathbf{f}(\mathbf{x})$  is twice continuously differentiable and once again has a globally attractive  $m$ -dimensional manifold of equilibria  $\Gamma$ , then the process  $\mathbf{z}_n(t) = \mathbf{x}_n(nt)$  satisfies the conditions of [28], so that as  $n \rightarrow \infty$ ,  $\mathbf{z}_n(t)$  converges to a diffusion  $\mathbf{z}(t)$  satisfying equation (A.2) for  $\mathbf{f}$  and  $\mathbf{G}$  defined as above. This result was applied to the study population genetic and epidemiological models in [12, 13].

## Appendix B. Local representations of one-dimensional manifolds

In this section, we will discuss how one may obtain a parameterisation  $\boldsymbol{\gamma}$  of a one-dimensional slow manifold  $\Gamma$  and compute the quadratic expansion of the flow field (i.e. the quantities  $\mathbf{v}$  and  $\boldsymbol{\Theta}$ ) in the neighbourhood of a point  $\mathbf{x} \in \Gamma$ .

We start by fixing a basis of generalised eigenvectors of the Jacobian at  $\mathbf{x}_0$ ,  $\mathbf{J}(\mathbf{x}_0)$ , say  $\mathbf{w}_1, \dots, \mathbf{w}_n$ , and letting  $\mathbf{W}$  be the corresponding change of basis matrix with the  $\mathbf{w}_i$  as columns. Let  $\mathbf{w}_1$  to be the eigenvector corresponding to the eigenvalue 0 (which we take to be unique up to scalar multiplication). Then,

$$\mathbf{W}^{-1} \mathbf{J}(\mathbf{x}_0) \mathbf{W} = \begin{bmatrix} 0 & \\ & \mathbf{J}_2 \end{bmatrix}, \quad (\text{B.1})$$

where  $\mathbf{J}_2$  is a block-diagonal matrix, with each block acting invariantly on one of the eigenspaces corresponding to the non-zero eigenvalues.

We introduce a new coordinate system

$$\mathbf{z} = \mathbf{W}^{-1} (\mathbf{x} - \mathbf{x}_0).$$

In this coordinate system, we will construct a parameterisation  $\boldsymbol{\gamma}(z_1)$  of  $\Gamma$  such that that  $\mathbf{x}_0 = \boldsymbol{\gamma}(0)$ .

In the new coordinate system, the dynamics are then given by  $\frac{dz}{dt} = \hat{\mathbf{f}}(\mathbf{z})$ , where

$$\hat{\mathbf{f}}(\mathbf{z}) = \mathbf{W}^{-1} \mathbf{f}(\mathbf{x}_0 + \mathbf{W}\mathbf{z})$$

(thus, the Jacobian of  $\hat{\mathbf{f}}$  at  $\mathbf{0}$ , say  $\hat{\mathbf{J}}$ , is  $\mathbf{W}^{-1} \mathbf{J}(\mathbf{x}_0) \mathbf{W}$ ). Setting  $\mathbf{z}_2 = (z_2, \dots, z_d)$ , we may write this as

$$\begin{aligned} \frac{dz_1}{dt} &= \varphi_1(z_1, \mathbf{z}_2) \\ \frac{dz_2}{dt} &= \mathbf{J}_2 \mathbf{z}_2 + \boldsymbol{\varphi}_2(z_1, \mathbf{z}_2). \end{aligned} \quad (\text{B.2})$$

where  $\boldsymbol{\varphi}_2(z_1, \mathbf{z}_2) = (\varphi_2, \dots, \varphi_d)$  is quadratic. We may thus Taylor expand  $\varphi_i(\mathbf{z})$  about  $\mathbf{0}$  as

$$\varphi_i(\mathbf{z}) = \sum_{j,k=1}^d c_{ijk} z_j z_k + O(|\mathbf{z}|^3).$$

Computing  $\boldsymbol{\gamma}$  or  $\Theta$ , is essentially the task of characterising the centre and stable manifolds at  $\mathbf{x}_0$  respectively. The centre manifold theorem (we follow the treatment in [55]) tells us that at  $\mathbf{x}_0$  the centre manifold is tangent to  $\mathbf{w}_1$ , whereas the stable manifold is tangent to the space spanned by  $\mathbf{w}_2, \dots, \mathbf{w}_d$ . Moreover, we may locally represent each manifold as the graph of a function over the tangent space. In particular, in the new coordinate system, there exists a function

$$\boldsymbol{\gamma}_2(z_1) = (\gamma_2(z_1), \dots, \gamma_d(z_1))$$

such that  $\boldsymbol{\gamma}(z) = (z, \boldsymbol{\gamma}_2(z))^T$  is a point on  $\Gamma$  for all  $z_1$  sufficiently close to 0, and a function  $\vartheta(\mathbf{z}_2)$  such that  $(\mathbf{z}_2, \vartheta(\mathbf{z}_2))$  is a point in the stable manifold near  $\mathbf{x}_0$  for  $\mathbf{z}_2$  sufficiently close to  $\mathbf{0}$ . We will demonstrate the calculation of  $\boldsymbol{\gamma}_2(z_1)$  below; the calculation of  $\vartheta(\mathbf{z}_2)$  is similar, so we will simply give the result. Finally, we will show how one obtains  $\Theta$  from  $\vartheta(\mathbf{z}_2)$ .

To begin, we observe that in our new coordinate system  $\mathbf{x}_0$  is the origin and  $\Gamma$  is tangent to the  $z_1$  axis (*i.e.* the span of  $\mathbf{w}_1$ ), so we must have  $\boldsymbol{\gamma}'_2(0) = \frac{d\boldsymbol{\gamma}_2}{dz_1} = \mathbf{0}$ . We thus look for  $\boldsymbol{\gamma}_2(z_1)$  of the form

$$\gamma_i(z_1) = a_i z_1^2 + O(z_1^3).$$

(as we shall only be interested in the first and second order derivatives of  $\boldsymbol{\gamma}$  at  $\mathbf{x}_0$  – *i.e.* at  $z_1 = 0$  – this is adequate for our purposes).

Substituting into (B.2), for points on  $\Gamma$  we have

$$\begin{aligned} \frac{dz_1}{dt} &= \varphi_1(z_1, \boldsymbol{\gamma}_2(z_1)) \\ \frac{d}{dt} \boldsymbol{\gamma}_2(z_1) &= \mathbf{J}_2 \boldsymbol{\gamma}_2(z_1) + \boldsymbol{\varphi}_2(z_1, \boldsymbol{\gamma}_2(z_1)), \end{aligned} \tag{B.3}$$

or, expanding the latter using the chain rule,

$$\varphi_1(z_1, \boldsymbol{\gamma}_2(z_1)) \frac{d\boldsymbol{\gamma}_2}{dz_1} = \mathbf{J}_2 \boldsymbol{\gamma}_2(z_1) + \boldsymbol{\varphi}_2(z_1, \boldsymbol{\gamma}_2(z_1)).$$

Substituting our series expressions for the  $\varphi_i$  and  $h_i$ , to lowest order this gives us

$$2c_{i11} a_i z_1^3 + O(z_1^4) = \left( \sum_{j=2}^d \hat{J}_{ij} a_j + c_{i11} \right) z_1^2 + O(z_1^3)$$

*i.e.* we may obtain the quantities  $a_i$ ,  $i = 2, \dots, d$  by solving the system of equations

$$\sum_{j=2}^d \hat{J}_{ij} a_j = -c_{i11}, \quad i = 2, \dots, d.$$

Noting that  $(a_2, \dots, a_d)^T = \frac{1}{2} \frac{d^2 \gamma_2}{dz_1^2}(0)$  whereas  $(c_{111}, \dots, c_{d11})^T = \frac{1}{2} \frac{\partial^2 \varphi_2}{\partial z_1^2}(\mathbf{0})$ , we can solve the previous equation as

$$\frac{d^2 \gamma_2}{dz_1^2}(0) = -\hat{\mathbf{J}}^{-1} \frac{\partial^2 \varphi_2}{\partial z_1^2}(\mathbf{0}).$$

To return to our original functions as expressed in the original coordinate system, we first observe that for  $i = 2, \dots, d$ ,

$$\frac{d^2 \varphi_i}{dz_1^2}(\mathbf{0}) = \frac{\partial^2 \hat{f}_i}{\partial z_1^2}(\mathbf{0}),$$

whereas

$$\frac{\partial^2 \hat{\mathbf{f}}}{\partial z_1^2}(\mathbf{0}) = \mathbf{W}^{-1} \sum_{j,k=1}^d \frac{\partial^2 \mathbf{f}}{\partial x_j \partial x_k}(\mathbf{x}_0) W_{j1} W_{k1}.$$

In particular, recalling (B.1), we see that for  $i = 2, \dots, d$ ,  $\frac{d^2 \gamma_i}{dz_1^2}(0)$  agrees with the  $i^{\text{th}}$  entry of

$$-\mathbf{W}^{-1} \mathbf{J}^+ \sum_{j,k=1}^d \frac{\partial^2 \mathbf{f}}{\partial x_j \partial x_k}(\mathbf{x}_0) W_{j1} W_{k1},$$

*i.e.*

$$-\left[ \mathbf{W}^T \frac{\partial^2 [\mathbf{W}^{-1} \mathbf{J}^+ \mathbf{f}]_i}{\partial \mathbf{x}^2}(\mathbf{x}_0) \mathbf{W} \right]_{11},$$

where, as before,  $\mathbf{J}^+$  is the pseudo-inverse of  $\mathbf{J}$ , which is defined by  $\mathbf{J}^{-1}$  on the image of  $\mathbf{J}$  and is  $\mathbf{0}$  on the kernel of  $\mathbf{J}$ .

Thus,

$$\begin{aligned} \gamma(0) &= \mathbf{x}_0 \\ \gamma'(0) &= \mathbf{w}_1 \\ \gamma''(0) &= \sum_{i=2}^d \frac{d^2 \gamma_i}{dz_1^2}(0) \mathbf{w}_i = - \sum_{i=2}^d \left[ \mathbf{W}^T \frac{\partial^2 [\mathbf{W}^{-1} \mathbf{J}^+ \mathbf{f}]_i}{\partial \mathbf{x}^2}(\mathbf{x}_0) \mathbf{W} \right]_{11} \mathbf{w}_i \end{aligned}$$

and

$$\gamma(z_1) = \mathbf{x}_0 + z_1 \mathbf{w}_1 - \frac{1}{2} z_1^2 \sum_{i=2}^d \left[ \mathbf{W}^T \frac{\partial^2 [\mathbf{W}^{-1} \mathbf{J}^+ \mathbf{f}]_i}{\partial \mathbf{x}^2}(\mathbf{x}_0) \mathbf{W} \right]_{11} \mathbf{w}_i + O(z_1^3)$$

is the desired parametrisation of  $\Gamma$  in the  $\mathbf{z}$  coordinates.

Proceeding similarly, we find that

$$\vartheta(\mathbf{z}_2) = \mathbf{z}_2^T (\hat{\mathbf{J}}^T)^{-1} \frac{\partial^2 \hat{f}_1}{\partial \mathbf{z}_2^2}(\mathbf{0}) \mathbf{z}_2$$

and  $(\hat{\mathbf{J}}^T)^{-1} \frac{\partial^2 \hat{f}_1}{\partial \mathbf{z}_2^2}(\mathbf{0})$  and

$$\mathbf{W}^T (\mathbf{J}^T)^+ \frac{\partial^2 [\mathbf{W}^{-1} \mathbf{f}]_1}{\partial \mathbf{x}^2}(\mathbf{x}_0) \mathbf{W}$$

have equal  $j, k^{\text{th}}$  entry for all  $j, k = 2, \dots, d$  (the first row of the latter is zero, but the first column need not be). Thus, if we set

$$\Theta(\mathbf{x}_0) = \mathbf{P}(\mathbf{x}_0)\mathbf{W}^T (\mathbf{J}^T)^+ \frac{\partial^2 [\mathbf{W}^{-1}\mathbf{f}]_1}{\partial \mathbf{x}^2}(\mathbf{x}_0)\mathbf{W},$$

then the stable manifold at  $\mathbf{x}_0$  is thus the set of all points  $\mathbf{z}$  such that

$$z_1 = \mathbf{z}_2^T \Theta(\mathbf{x}_0) \mathbf{z}_2.$$

Now, if we choose  $\mathbf{v}(\mathbf{x}_0)$  so that

$$\mathbf{v}(\mathbf{x}_0)^T \mathbf{w}_i = \begin{cases} 1 & \text{if } i = 1, \text{ and} \\ 0 & \text{otherwise,} \end{cases}$$

then for a point  $\mathbf{x} = \mathbf{x}_0 + \Delta \mathbf{x}$ ,  $z_1 = \mathbf{v}(\mathbf{x}_0)^T \Delta \mathbf{x}$ , whereas

$$\mathbf{z}_2 = \Delta \mathbf{x} - (\mathbf{v}(\mathbf{x}_0)^T \Delta \mathbf{x}) \mathbf{w}_1,$$

so that  $\mathbf{x}$  is in the stable manifold at  $\mathbf{x}_0$  (to lowest order in  $\Delta \mathbf{x}$ ) provided

$$\mathbf{v}(\mathbf{x}_0)^T \Delta \mathbf{x} - (\Delta \mathbf{x} - (\mathbf{v}(\mathbf{x}_0)^T \Delta \mathbf{x}) \mathbf{w}_1)^T \Theta(\mathbf{x}_0) (\Delta \mathbf{x} - (\mathbf{v}(\mathbf{x}_0)^T \Delta \mathbf{x}) \mathbf{w}_1) = 0,$$

or, rearranging,

$$\mathbf{v}(\mathbf{x}_0)^T \Delta \mathbf{x} - \Delta \mathbf{x}^T (\mathbf{I} - \mathbf{w}_1 \mathbf{v}(\mathbf{x}_0)^T)^T \Theta(\mathbf{x}_0) (\mathbf{I} - \mathbf{w}_1 \mathbf{v}(\mathbf{x}_0)^T) \Delta \mathbf{x} = 0.$$

### Appendix C. Derivation of general case

First we examine the projection matrix  $\mathbf{P}$ . Consider the outer system

$$\frac{d\xi}{dt} = \mathbf{f}(\xi), \quad \xi(0) = \mathbf{x}, \tag{C.1}$$

where  $\mathbf{x}$  lies close to a point  $\mathbf{z}$  on the manifold. Varying the initial conditions yields

$$\frac{d}{dt} \frac{\partial \xi_i}{\partial x_j} = \frac{\partial}{\partial x_j} f_i(\xi) = \sum_k \frac{\partial \xi_k}{\partial x_j} \frac{\partial}{\partial \xi_k} f_i(\xi) = \sum_k J_{ik}(\xi) \frac{\partial \xi_k}{\partial x_j} \tag{C.2}$$

*i.e.*

$$\frac{d}{dt} \frac{\partial \xi}{\partial \mathbf{x}} = \mathbf{J}(\xi(t)) \frac{\partial \xi}{\partial \mathbf{x}}$$

where  $\mathbf{J}$  is the Jacobian matrix of  $\mathbf{f}$ . Now, since  $\xi(0) = \mathbf{x}$ ,

$$\frac{\partial \xi}{\partial \mathbf{x}}(0, \mathbf{x}) = \mathbf{I}$$

and thus this variational equation has solution

$$\frac{\partial \xi_i}{\partial x_j} = \mathbf{\Pi}(0, t)$$

where  $\mathbf{\Pi}(s, t)$  is the fundamental matrix solving

$$\frac{d}{dt}\mathbf{\Pi}(s, t) = \mathbf{J}(\boldsymbol{\xi}(\mathbf{x}, t))\mathbf{\Pi}(s, t), \quad \mathbf{\Pi}(s, s) = I.$$

When  $\mathbf{x}$  is taken to be  $\mathbf{z} \in \Gamma$ , since  $\boldsymbol{\xi}(\mathbf{z}, t) = \mathbf{z}$  for all  $\mathbf{z} \in \Gamma$ , we have

$$\mathbf{\Pi}(s, t) = e^{(t-s)\mathbf{J}(\mathbf{z})}.$$

so, in this case,  $\frac{\partial \boldsymbol{\pi}}{\partial \mathbf{x}}(\mathbf{z}, t) = e^{t\mathbf{J}(\mathbf{z})}$  (*i.e.* informally,  $\frac{d}{dt} \frac{\partial \boldsymbol{\xi}}{\partial \mathbf{x}} \approx \mathbf{J}(\mathbf{z}) \frac{\partial \boldsymbol{\xi}}{\partial \mathbf{x}}$ ). Under this approximation the equation is linear and admits the solution  $\frac{\partial \boldsymbol{\xi}}{\partial \mathbf{x}} = e^{t\mathbf{J}(\mathbf{z})}$ .

From the definitions (4) and (8) we recover  $\mathbf{P}$  by taking the limit of large  $t$ ,

$$\mathbf{P}(\mathbf{z}) = \lim_{t \rightarrow \infty} e^{t\mathbf{J}(\mathbf{z})}. \quad (\text{C.3})$$

To compute the limit we consider the action of  $e^{t\mathbf{J}(\mathbf{z})}$  on an eigenvector of the Jacobian<sup>+</sup>. If  $\mathbf{u}_i$  is tangent to the manifold then the corresponding eigenvalue  $\lambda_i$  is zero and so  $e^{t\lambda_i} = 1$  and  $\mathbf{P}(\mathbf{z})$  leaves  $\mathbf{u}_i$  unchanged. Alternatively, if  $\mathbf{u}_i$  corresponds to a direction of fast collapse then its eigenvalue is negative and  $e^{t\lambda_i} \rightarrow 0$ , so  $\mathbf{u}_i$  is annihilated by  $\mathbf{P}(\mathbf{z})$ .

Let  $\mathbf{U} = (\mathbf{u}_1, \dots, \mathbf{u}_m)$  be a basis of the tangent plane to the manifold at  $\mathbf{z}$  (the slow subspace) and let  $\mathbf{V} = (\mathbf{v}_1, \dots, \mathbf{v}_m)$  a basis of the orthogonal complement of the fast subspace. Then we may write

$$\mathbf{P}(\mathbf{z}) = \mathbf{U}(\mathbf{V}^T \mathbf{U})^{-1} \mathbf{V}^T. \quad (\text{C.4})$$

In the above we assumed that the tangent plane to the manifold was precisely the kernel of the Jacobian, in which case  $\mathbf{U}$  would be the first  $m$  columns of the right eigenvector matrix, and  $\mathbf{V}^T$  the bottom  $d - m$  rows of the left eigenvector matrix. This may not hold if the manifold is not hyperbolic (for example if  $\mathbf{f}$  has a component like  $-x_i^3$ , which is stable but not linearly so), however, equation (C.4) remains true for all flow fields, provided we somehow have access to bases  $\mathbf{U}$  and  $\mathbf{V}$ .

Let us move on to calculate  $\mathbf{Q}$ . We start by obtaining some simple identities: first note that by the definition of  $\boldsymbol{\pi}$ , we have  $f_i(\boldsymbol{\pi}(\mathbf{x})) = 0$  for all  $\mathbf{x}$ . Differentiating this, we obtain

$$\sum_m \frac{\partial f_i}{\partial x_m}(\boldsymbol{\pi}(\mathbf{x})) \frac{\partial \pi_m}{\partial x_j} = 0, \quad (\text{C.5})$$

or, in matrix form,  $\mathbf{J}(\boldsymbol{\pi}(\mathbf{x})) \frac{\partial \boldsymbol{\pi}}{\partial \mathbf{x}} = \mathbf{0}$ . Replacing  $\mathbf{x}$  by  $\mathbf{z} \in \Gamma$ , and recalling that  $\frac{\partial \boldsymbol{\pi}}{\partial \mathbf{x}}(\mathbf{z}) = \mathbf{P}(\mathbf{z})$ , we have

$$\mathbf{J}(\mathbf{z})\mathbf{P}(\mathbf{z}) = \mathbf{0},$$

*i.e.*  $\mathbf{J}(\mathbf{z})$  annihilates all the slow directions, as we have already observed. Differentiating (C.5), we obtain

$$\sum_{m,n} \frac{\partial^2 f_i}{\partial x_m \partial x_n}(\boldsymbol{\pi}(\mathbf{x})) \frac{\partial \pi_m}{\partial x_j} \frac{\partial \pi_n}{\partial x_k} + \sum_m \frac{\partial f_i}{\partial x_m}(\boldsymbol{\pi}(\mathbf{x})) \frac{\partial^2 \pi_m}{\partial x_j \partial x_k} = 0,$$

<sup>+</sup> To simplify the discussion we assume that  $\mathbf{J}(\mathbf{z})$  is diagonalisable and that its kernel contains only the tangent plane to the manifold. Neither assumption is necessary.



which we can write in vector form as

$$\mathcal{H}_{jk} \left( \frac{\partial \boldsymbol{\pi}}{\partial \mathbf{x}} \right) + \mathbf{J}(\boldsymbol{\pi}(\mathbf{x})) \frac{\partial^2 \boldsymbol{\pi}}{\partial x_j \partial x_k} = 0, \quad (\text{C.6})$$

where, for any  $n \times n$ -matrix  $A$ ,  $\mathcal{H}_{jk}(A)$  is the vector with  $i^{\text{th}}$  entry

$$\mathcal{H}_{ijk}(A) = \mathbf{e}_j^T A^T \frac{\partial^2 f_i}{\partial \mathbf{x}^2} A \mathbf{e}_k,$$

where  $\mathbf{e}_j$  is the  $j^{\text{th}}$  standard basis vector, and we have written  $\frac{\partial^2 f_i}{\partial \mathbf{x}^2}$  for the Hessian matrix with  $j, k^{\text{th}}$  entry  $\frac{\partial^2 f_i}{\partial x_j \partial x_k}$ . *i.e.*, since  $\frac{\partial \boldsymbol{\pi}}{\partial x_j} = \frac{\partial \boldsymbol{\pi}}{\partial \mathbf{x}} \mathbf{e}_j$ ,

$$\mathcal{H}_{ijk} \left( \frac{\partial \boldsymbol{\pi}}{\partial \mathbf{x}} \right) = \left( \frac{\partial \boldsymbol{\pi}}{\partial x_j} \right)^T \frac{\partial^2 f_i}{\partial \mathbf{x}^2} \frac{\partial \boldsymbol{\pi}}{\partial x_k} = \sum_{m,n} \frac{\partial^2 f_i}{\partial x_m \partial x_n}(\boldsymbol{\pi}(\mathbf{x})) \frac{\partial \pi_m}{\partial x_j} \frac{\partial \pi_n}{\partial x_k}.$$

Now, recalling that at  $\mathbf{z} \in \Gamma$ ,  $\boldsymbol{\pi}(\mathbf{z}) = \mathbf{z}$ ,  $\frac{\partial \boldsymbol{\pi}}{\partial \mathbf{x}} = \mathbf{P}(\mathbf{z})$ , and  $\frac{\partial^2 \pi_i}{\partial x_j \partial x_k}(\mathbf{z}) = Q_{ijk}(\mathbf{z})$ , we can write (C.6) as

$$\mathbf{J}(\mathbf{z}) \mathbf{Q}_{jk}(\mathbf{z}) = -\mathcal{H}_{jk}(\mathbf{P}(\mathbf{z})), \quad (\text{C.7})$$

where we continue with the convention that  $\mathbf{Q}_{jk}(\mathbf{z})$  is the vector with  $i^{\text{th}}$  entry  $Q_{ijk}(\mathbf{z})$ . Applying  $\mathbf{P}(\mathbf{z})$  to both sides of (C.7) gives

$$\mathbf{P}(\mathbf{z}) \mathcal{H}_{jk}(\mathbf{P}(\mathbf{z})) = \mathbf{0}$$

so we see  $\mathcal{H}_{jk}(\mathbf{P}(\mathbf{z}))$  is entirely contained in the eigenspace of fast directions. Notice that restricted to the fast subspace,  $\mathbf{J}(\mathbf{z})$  is a full-rank operator, so that, *regarded as an operator on the fast subspace*, (C.7) has a unique solution, which we will write as

$$-\mathbf{J}(\mathbf{z})^+ \mathcal{H}_{jk}(\mathbf{P}(\mathbf{z})).$$

where we recall that  $\mathbf{J}(\mathbf{z})^+$  is the pseudo-inverse of  $\mathbf{J}(\mathbf{z})$ , which acts as the inverse of  $\mathbf{J}(\mathbf{z})$  when restricted to the fast directions and which annihilates all vectors in the slow directions.

However, regarded as an equation on all of  $\mathbb{R}^d$ , the solution to (C.7) is not unique, but rather takes the form

$$\mathbf{Q}_{jk}(\mathbf{z}) = -\mathbf{J}(\mathbf{z})^+ \mathcal{H}_{jk}(\mathbf{P}(\mathbf{z})) + \mathbf{S}_{jk}(\mathbf{z})$$

for some vector  $\mathbf{S}_{jk}(\mathbf{z})$  in the slow directions.

To obtain  $\mathbf{S}_{jk}(\mathbf{z})$ , we proceed as we did to obtain  $\mathbf{P}(\mathbf{z})$ , differentiating (C.2) to obtain

$$\frac{d}{dt} \frac{\partial^2 \xi_i}{\partial x_j \partial x_k} = \sum_l \mathbf{J}_{il}(\boldsymbol{\xi}) \frac{\partial^2 \xi_l}{\partial x_j \partial x_k} + \sum_{m,n} \frac{\partial^2 f_i}{\partial x_m \partial x_n}(\boldsymbol{\xi}) \frac{\partial \xi_m}{\partial x_j} \frac{\partial \xi_n}{\partial x_k}$$

which again write in vector form as

$$\frac{d}{dt} \frac{\partial^2 \boldsymbol{\xi}}{\partial x_j \partial x_k} = \mathbf{J}(\boldsymbol{\xi}) \frac{\partial^2 \boldsymbol{\xi}}{\partial x_j \partial x_k} + \mathcal{H}_{jk} \left( \frac{\partial \boldsymbol{\xi}}{\partial \mathbf{x}} \right). \quad (\text{C.8})$$

This may be formally solved by Duhamel's principle to give

$$\frac{\partial^2 \boldsymbol{\xi}}{\partial x_j \partial x_k} = \int_0^t \boldsymbol{\Pi}(s, t) \boldsymbol{\mathcal{H}}_{jk} \left( \frac{\partial \boldsymbol{\xi}}{\partial \mathbf{x}}(x, s) \right) ds$$

where  $\boldsymbol{\Pi}(s, t)$  is the fundamental matrix from above.

As before, when  $\mathbf{x}$  is taken to be a point  $\mathbf{z} \in \Gamma$ , since  $\boldsymbol{\xi}(\mathbf{z}, t) = \mathbf{z}$  for all  $\mathbf{z} \in \Gamma$ , we have  $\boldsymbol{\Pi}(s, t) = e^{(t-s)\mathbf{J}(\mathbf{z})}$  and  $\frac{\partial \boldsymbol{\pi}}{\partial \mathbf{x}}(\mathbf{z}, t) = e^{t\mathbf{J}(\mathbf{z})}$ , and the solution to (C.8) simplifies to

$$\int_0^t e^{(t-s)\mathbf{J}(\mathbf{z})} \boldsymbol{\mathcal{H}}_{jk} \left( e^{s\mathbf{J}(\mathbf{z})} \right) ds.$$

Thus,

$$\mathbf{Q}_{jk}(\mathbf{z}) = \lim_{t \rightarrow \infty} \frac{\partial^2 \boldsymbol{\xi}}{\partial x_j \partial x_k} = \lim_{t \rightarrow \infty} \int_0^t e^{(t-s)\mathbf{J}(\mathbf{z})} \boldsymbol{\mathcal{H}}_{jk} \left( e^{s\mathbf{J}(\mathbf{z})} \right) ds.$$

Now,  $e^{t\mathbf{J}(\mathbf{z})} \rightarrow \mathbf{P}(\mathbf{z})$  as  $t \rightarrow \infty$ , and

$$\lim_{t \rightarrow \infty} \boldsymbol{\mathcal{H}}_{jk} \left( e^{t\mathbf{J}(\mathbf{z})} \right) = \boldsymbol{\mathcal{H}}_{jk}(\mathbf{P}(\mathbf{z})),$$

both of which are non-zero, so it is not immediately obvious that the integral above converges. However, the information obtained above allows us to resolve these issues. We start by observing that

$$\mathbf{S}_{jk}(\mathbf{z}) = \mathbf{P}(\mathbf{z}) \mathbf{Q}_{jk}(\mathbf{z}) = \lim_{t \rightarrow \infty} \int_0^t \mathbf{P}(\mathbf{z}) e^{(t-s)\mathbf{J}(\mathbf{z})} \boldsymbol{\mathcal{H}}_{jk} \left( e^{s\mathbf{J}(\mathbf{z})} \right) ds,$$

and, since  $e^{(t-s)\mathbf{J}(\mathbf{z})}$  acts like the identity matrix on the slow directions,  $\mathbf{P}(\mathbf{z}) e^{(t-s)\mathbf{J}(\mathbf{z})} = \mathbf{P}(\mathbf{z})$ , so that

$$\mathbf{S}_{jk}(\mathbf{z}) = \lim_{t \rightarrow \infty} \int_0^t \mathbf{P}(\mathbf{z}) \boldsymbol{\mathcal{H}}_{jk} \left( e^{s\mathbf{J}(\mathbf{z})} \right) ds = \mathbf{P}(\mathbf{z}) \int_0^\infty \boldsymbol{\mathcal{H}}_{jk} \left( e^{s\mathbf{J}(\mathbf{z})} \right) ds.$$

Moreover, we've already observed that  $\mathbf{P}(\mathbf{z}) \boldsymbol{\mathcal{H}}_{jk}(\mathbf{P}(\mathbf{z})) = \mathbf{0}$ , so

$$\mathbf{S}_{jk}(\mathbf{z}) = \mathbf{P}(\mathbf{z}) \int_0^\infty \boldsymbol{\mathcal{H}}_{jk} \left( e^{s\mathbf{J}(\mathbf{z})} \right) - \boldsymbol{\mathcal{H}}_{jk}(\mathbf{P}(\mathbf{z})) ds,$$

and we are left with evaluating the integral

$$\begin{aligned} & \int_0^\infty \boldsymbol{\mathcal{H}}_{ijk} \left( e^{s\mathbf{J}(\mathbf{z})} \right) - \boldsymbol{\mathcal{H}}_{ijk}(\mathbf{P}(\mathbf{z})) ds \\ &= \int_0^\infty \mathbf{e}_j^T e^{s\mathbf{J}(\mathbf{z})^T} \frac{\partial^2 f_i}{\partial \mathbf{x}^2}(\mathbf{z}) e^{s\mathbf{J}(\mathbf{z})} \mathbf{e}_k - \mathbf{e}_j^T \mathbf{P}(\mathbf{z})^T \frac{\partial^2 f_i}{\partial \mathbf{x}^2}(\mathbf{z}) \mathbf{P}(\mathbf{z}) \mathbf{e}_k \\ &= \mathbf{e}_j^T \left( \int_0^\infty e^{s\mathbf{J}(\mathbf{z})^T} \frac{\partial^2 f_i}{\partial \mathbf{x}^2}(\mathbf{z}) e^{s\mathbf{J}(\mathbf{z})} - \mathbf{P}(\mathbf{z})^T \frac{\partial^2 f_i}{\partial \mathbf{x}^2}(\mathbf{z}) \mathbf{P}(\mathbf{z}) ds \right) \mathbf{e}_k. \end{aligned}$$

Now,

$$\begin{aligned} & \int_0^\infty e^{s\mathbf{J}(z)^T} \frac{\partial^2 f_i}{\partial \mathbf{x}^2}(z) e^{s\mathbf{J}(z)} - \mathbf{P}(z)^T \frac{\partial^2 f_i}{\partial \mathbf{x}^2}(z) \mathbf{P}(z) ds \\ &= \int_0^\infty (e^{s\mathbf{J}(z)} - \mathbf{P}(z))^T \frac{\partial^2 f_i}{\partial \mathbf{x}^2}(z) (e^{s\mathbf{J}(z)} - \mathbf{P}(z)) ds \\ &+ \int_0^\infty (e^{s\mathbf{J}(z)} - \mathbf{P}(z))^T \frac{\partial^2 f_i}{\partial \mathbf{x}^2}(z) \mathbf{P}(z) ds + \int_0^\infty \mathbf{P}(z)^T \frac{\partial^2 f_i}{\partial \mathbf{x}^2}(z) (e^{s\mathbf{J}(z)} - \mathbf{P}(z)) ds, \end{aligned}$$

and, since  $e^{t\mathbf{J}(z)} - \mathbf{P}(z)$  vanishes on the slow directions, and acts as  $e^{t\mathbf{J}(z)}$  restricted to the fast directions,

$$\int_0^\infty e^{s\mathbf{J}(z)} - \mathbf{P}(z) ds = -\mathbf{J}(z)^+$$

whereas

$$\mathbf{X}_i(z) = \int_0^\infty (e^{s\mathbf{J}(z)} - \mathbf{P}(z))^T \frac{\partial^2 f_i}{\partial \mathbf{x}^2}(z) (e^{s\mathbf{J}(z)} - \mathbf{P}(z)) ds \quad (\text{C.9})$$

is the unique solution to the Lyapunov equation

$$\mathbf{J}(z)^T \mathbf{X}_i(z) + \mathbf{X}_i(z) \mathbf{J}(z) = -\frac{\partial^2 f_i}{\partial \mathbf{x}^2}(z)$$

in the fast subspace [56]. Thus,

$$\mathbf{S}_{jk}(z) = \mathbf{P}(z) \tilde{\mathbf{S}}_{jk}(z),$$

where

$$\tilde{\mathbf{S}}_{ijk}(z) = \mathbf{e}_j^T \left( \mathbf{X}_i(z) - (\mathbf{J}(z)^+)^T \frac{\partial^2 f_i}{\partial \mathbf{x}^2} \mathbf{P}(z) - \mathbf{P}(z)^T \frac{\partial^2 f_i}{\partial \mathbf{x}^2} \mathbf{J}(z)^+ \right) \mathbf{e}_k$$

and, finally,

$$\mathbf{Q}_{jk}(z) = -\mathbf{J}(z)^+ \mathcal{H}_{jk}(\mathbf{P}(z)) + \mathbf{P}(z) \tilde{\mathbf{S}}_{jk}(z).$$

Finally, we remark that in the co-dimension one case, (C.9) can be evaluated directly. Adopting the notation of 2.3, we have  $\mathbf{f} = \phi \mathbf{r}$ ,  $\mathbf{J} = \mathbf{r} \nabla \phi^T$ ,  $\lambda = \nabla \phi^T \mathbf{r}$ , and

$$\mathbf{P} = \mathbf{I} - \frac{1}{\lambda} \mathbf{J}.$$

Then, for an arbitrary vector  $\mathbf{Y}$ ,  $\mathbf{J}\mathbf{Y} = \mathbf{r} \nabla \phi^T \mathbf{Y}$ , so that

$$\mathbf{J}^2 \mathbf{Y} = \mathbf{r} \nabla \phi^T \mathbf{r} \nabla \phi^T \mathbf{Y} = \lambda (\nabla \phi^T \mathbf{Y}) \mathbf{r},$$

$\mathbf{J}^n \mathbf{Y} = \lambda^{n-1} (\nabla \phi^T \mathbf{Y}) \mathbf{r}$ , and

$$\begin{aligned} e^{s\mathbf{J}} \mathbf{Y} &= \sum_{n=0}^{\infty} \frac{s^n}{n!} \mathbf{J}^n \mathbf{Y} \\ &= \mathbf{Y} + (\nabla \phi^T \mathbf{Y}) \mathbf{r} \lambda^{n-1} \sum_{n=1}^{\infty} \frac{s^n}{n!} \\ &= \mathbf{Y} + \frac{(\nabla \phi^T \mathbf{Y})}{\lambda} (e^{\lambda t} - 1) \mathbf{r}. \end{aligned}$$

Thus,  $e^{s\mathbf{J}} - \mathbf{P}\mathbf{Y} = \frac{(\nabla\phi^T\mathbf{Y})}{\lambda}e^{\lambda t}\mathbf{r}$  and, recalling that  $\lambda < 0$ , we have that

$$\begin{aligned} X_{ijk} &= \mathbf{e}_j^T \left( \int_0^\infty \mathbf{e}_i^T (e^{s\mathbf{J}} - \mathbf{P})^T \frac{\partial^2 f_i}{\partial \mathbf{x}^2} (e^{s\mathbf{J}} - \mathbf{P}) ds \right) \mathbf{e}_k \\ &= \frac{(\nabla\phi^T \mathbf{e}_j)(\nabla\phi^T \mathbf{e}_k)}{\lambda^2} \mathbf{r}^T \frac{\partial^2 f_i}{\partial \mathbf{x}^2} \mathbf{r} \int_0^\infty e^{2\lambda t} dt \\ &= -\frac{\frac{\partial\phi}{\partial x_j} \frac{\partial\phi}{\partial x_k}}{2\lambda^3} \mathbf{r}^T \frac{\partial^2 f_i}{\partial \mathbf{x}^2} \mathbf{r}. \end{aligned}$$

- [1] T. G. Kurtz. Strong approximation theorems for density dependent Markov chains. *Stochastic Processes Appl.*, 6(3):223 – 240, 1978.
- [2] Alan J McKane and Timothy J Newman. Predator-prey cycles from resonant amplification of demographic stochasticity. *Physical review letters*, 94(21):218102, 2005.
- [3] Andrew J Black and Alan J McKane. Stochastic formulation of ecological models and their applications. *Trends in ecology & evolution*, 27(6):337–345, 2012.
- [4] David Alonso, Alan J McKane, and Mercedes Pascual. Stochastic amplification in epidemics. *Journal of the Royal Society Interface*, 4(14):575–582, 2007.
- [5] Rachel Kuske, Luis F Gordillo, and Priscilla Greenwood. Sustained oscillations via coherence resonance in sir. *Journal of theoretical biology*, 245(3):459–469, 2007.
- [6] Tommaso Biancalani, Duccio Fanelli, and Francesca Di Patti. Stochastic turing patterns in the brusselator model. *Physical Review E*, 81(4):046215, 2010.
- [7] Thomas Butler and Nigel Goldenfeld. Fluctuation-driven turing patterns. *Physical Review E*, 84(1):011112, 2011.
- [8] Alan J McKane, Tommaso Biancalani, and Tim Rogers. Stochastic pattern formation and spontaneous polarisation: the linear noise approximation and beyond. *Bulletin of mathematical biology*, 76(4):895–921, 2014.
- [9] T. L. Parsons and C. Quince. Fixation in haploid populations exhibiting density dependence II: The quasi-neutral case. *Theor. Popul. Biol.*, 72(4):468–479, 2007.
- [10] T. L. Parsons, C. Quince, and J. B. Plotkin. Expected times to absorption and fixation for quasi-neutral and neutral haploid populations exhibiting density dependence. *Theor. Popul. Biol.*, 74(4):302–310, 2008.
- [11] R. Durrett and L. Popovic. Degenerate diffusions arising from gene duplication models. *Ann. Appl. Prob.*, 19(1):15–48, 2009.
- [12] T. L. Parsons, C. Quince, and J. B. Plotkin. Some consequences of demographic stochasticity in population genetics. *Genetics*, 185:1345–1354, 2010.
- [13] T. L. Parsons. *Asymptotic Analysis of Some Stochastic Models from Population Dynamics and Population Genetics*. PhD thesis, University of Toronto, 2012.
- [14] T. Rogers, A. J. McKane, and A. G. Rossberg. Demographic noise can lead to the spontaneous formation of species. *EPL (Europhysics Letters)*, 97(4):40008, 2012.
- [15] Tim Rogers, Alan J McKane, and Axel G Rossberg. Spontaneous genetic clustering in populations of competing organisms. *Physical Biology*, 9(6):066002, 2012.
- [16] YenTing Lin, Hyejin Kim, and CharlesR. Doering. Features of fast living: On the weak selection for longevity in degenerate birth-death processes. *Journal of Statistical Physics*, 148(4):647–663, 2012.
- [17] George W A Constable, Alan J McKane, and Tim Rogers. Stochastic dynamics on slow manifolds. *Journal of Physics A: Mathematical and Theoretical*, 46(29):295002, 2013.
- [18] George W. A. Constable and Alan J. McKane. Fast-mode elimination in stochastic metapopulation models. *Phys. Rev. E*, 89:032141, Mar 2014.
- [19] Simone Pigolotti and Roberto Benzi. Selective advantage of diffusing faster. *Phys. Rev. Lett.*, 112:188102, May 2014.

- [20] George W. A. Constable and Alan J. McKane. Models of genetic drift as limiting forms of the lotka-volterra competition model. *Phys. Rev. Lett.*, 114:038101, Jan 2015.
- [21] Thiparat Chotibut and David R. Nelson. Evolutionary dynamics with fluctuating population sizes and strong mutualism. *Phys. Rev. E*, 92:022718, Aug 2015.
- [22] Crispin W Gardiner et al. *Handbook of stochastic methods*, volume 4. Springer Berlin, 1985.
- [23] M. Born and R. Oppenheimer. Zur quantentheorie der molekeln. *Annalen der Physik*, 389(20):457–484, 1927.
- [24] Robert Zwanzig. Ensemble method in the theory of irreversibility. *The Journal of Chemical Physics*, 33(5):1338–1341, 1960.
- [25] Petra Boxler. A stochastic version of center manifold theory. *Probability Theory and Related Fields*, 83(4):509–545, 1989.
- [26] Ludwig Arnold and Peter Imkeller. Normal forms for stochastic differential equations. *Probability Theory and Related Fields*, 110(4):559–588, 1998.
- [27] Anthony J Roberts. Normal form transforms separate slow and fast modes in stochastic dynamical systems. *Physica A: Statistical Mechanics and its Applications*, 387(1):12–38, 2008.
- [28] G. S. Katzenberger. Solutions of a stochastic differential equation forced onto a manifold by a large drift. *Ann. Probab.*, 19(4):1587–1628, 1991.
- [29] T. Funaki and H. Nagai. Degenerative convergence of diffusion process toward a submanifold by strong drift. *Stochastics*, 44(1-2):1–25, 1993.
- [30] Alexandra Goeke, Christian Schilli, Sebastian Walcher, and Eva Zerz. Computing quasi-steady state reductions. *Journal of Mathematical Chemistry*, 50(6):1495–1513, 2012.
- [31] Alexandra Goeke, Sebastian Walcher, and Eva Zerz. Determining “small parameters” for quasi-steady state. *Journal of Differential Equations*, 259(3):1149–1180, 2015.
- [32] Alexandra Goeke, Sebastian Walcher, and Eva Zerz. Classical quasi-steady state reduction—a mathematical characterization. *Physica D: Nonlinear Phenomena*, 345:11–26, 2017.
- [33] Neil Fenichel. Asymptotic stability with rate conditions. *Indiana U. Math. J.*, 23(12):1109–1137, 1974.
- [34] K. Itô and H.P.J. McKean. *Diffusion Processes and their Sample Paths*. Springer Verlag, 1974.
- [35] R. H. Bartels and GW Stewart. Solution of the matrix equation  $ax + xb = c$  [f4]. *Communications of the ACM*, 15(9):820–826, 1972.
- [36] Ibrahim Fatkullin, Gregor Kovacic, Eric Vanden-Eijnden, et al. Reduced dynamics of stochastically perturbed gradient flows. *Commun. Math. Sci.*, 8(2):439–461, 2010.
- [37] FG Heineken, HM Tsuchiya, and R Aris. On the mathematical status of the pseudo-steady state hypothesis of biochemical kinetics. *Mathematical Biosciences*, 1(1):95–113, 1967.
- [38] Desmond J Higham. An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM review*, 43(3):525–546, 2001.
- [39] Jae Kyoung Kim, Krešimir Josić, and Matthew R Bennett. The validity of quasi-steady-state approximations in discrete stochastic simulations. *Biophysical journal*, 107(3):783–793, 2014.
- [40] Jae Kyoung Kim, Krešimir Josić, and Matthew R Bennett. The relationship between stochastic and deterministic quasi-steady state approximations. *BMC systems biology*, 9(1):87, 2015.
- [41] W. J. Ewens. *Mathematical Population Genetics*. Springer-Verlag,, Berlin/New York, 1979.
- [42] Alison Etheridge and Peter March. A note on superprocesses. *Probability Theory and Related Fields*, 89(2):141–147, 1991.
- [43] N. Berglund and B. Gentz. *Noise-induced phenomena in slow-fast dynamical systems: a sample-paths approach*. Springer Science Business Media, 2006.
- [44] Richard P Boland, Tobias Galla, and Alan J McKane. Limit cycles, complex floquet multipliers, and intrinsic noise. *Physical Review E*, 79(5):051131, 2009.
- [45] Axel G Rossberg, Tim Rogers, and Alan J McKane. Are there species smaller than 1 mm? *Proceedings of the Royal Society of London B: Biological Sciences*, 280(1767):20131248, 2013.
- [46] George WA Constable and Alan J McKane. Population genetics on islands connected by an arbitrary network: An analytic approach. *Journal of theoretical biology*, 358:149–165, 2014.

- [47] P. E. Protter. *Stochastic Integration and Differential Equations*. Springer, Berlin, 2nd edition, 2004.
- [48] P. Billingsley. *Convergence of Probability Measures*. Wiley, New York, 1968.
- [49] S. N. Ethier and T. G. Kurtz. *Markov Processes: Characterization and Convergence*. John Wiley and Sons, New York, 1986.
- [50] T. G. Kurtz. Solutions of ordinary differential equations as limits of pure jump markov processes. *J. Appl. Prob.*, 7(1):pp. 49–58, 1970.
- [51] T. G. Kurtz. Limit theorems for sequences of jump markov processes approximating ordinary differential processes. *J. Appl. Prob.*, 8(2):344–356, 1971.
- [52] T. G. Kurtz. *Approximation of Population Processes*, volume 36. Society for Industrial and Applied Mathematics, Philadelphia, 1981.
- [53] P. K. Pollett. On a model for interference between searching insect parasites. *J. Austral. Math. Soc. Ser. B*, 32(02):133–150, 1990.
- [54] N. G. van Kampen. *Stochastic Processes in Physics and Chemistry*. Elsevier, Amsterdam, San Diego, Oxford, London, 1992.
- [55] P. Glendinning. *Stability, Instability, and Chaos: an Introduction to the Theory of Nonlinear Differential Equations*. Cambridge University Press, Cambridge, 1994.
- [56] R. E. Bellman. *Introduction to Matrix Analysis*. Society for Industrial and Applied Mathematics, Philadelphia, 1995.