Normal form transforms separate slow and fast modes in stochastic dynamical systems

A. J. Roberts*

July 25, 2007

Abstract

Modelling stochastic systems has many important applications. Normal form coordinate transforms are a powerful way to untangle interesting long term macroscale dynamics from insignificant detailed microscale dynamics. We explore such coordinate transforms of stochastic differential systems when the dynamics has both slow modes and quickly decaying modes. The thrust is to derive normal forms useful for macroscopic modelling of complex stochastic microscopic systems. Thus we not only must reduce the dimensionality of the dynamics, but also endeavour to separate *all* slow processes from *all* fast time processes, both deterministic and stochastic. Quadratic stochastic effects in the fast modes contribute to the drift of the important slow modes. Some examples demonstrate that the coordinate transform may be only locally valid or may be globally valid depending upon the dynamical system. The results will help us accurately model, interpret and simulate multiscale stochastic systems.

Contents

*Computational Engineering and Sciences Research Centre, Department of Mathematics & Computing, University of Southern Queensland, Toowoomba, Queensland 4352, Australia. mailto:aroberts@usq.edu.au

1 Introduction

Normal form coordinate transformations provide a sound basis for simplifying multiscale nonlinear dynamics [15, 12, e.g.]. In systems with fast and slow dynamics, a coordinate transform is sought that decouples the slow from the fast. The decoupled slow modes then provide accurate predictions for the long term dynamics. Arguably, such normal form coordinate transformations provide a much more insightful view of simplifying dynamics than other, more popular, techniques. Averaging is perhaps the most popular technique for simplifying dynamics [38, Chapters 11–13, e.g.], especially for stochastic dynamics that we explore here [26, 17, e.g.]. But averaging fails in many cases. For example, consider the simple, linear, slow-fast system of stochastic differential equations (SDEs)

$$dx = \epsilon y\, dt \quad \text{and} \quad dy = -y\, dt + dW\,, \tag{1}$$

where for small parameter ϵ the variable $x(t)$ evolves slowly compared to the fast variable $y(t)$. Let us compare the predictions of averaging and a 'normal form' coordinate transform. First consider averaging: the fast variable y, being an Ornstein–Uhlenbeck process, rapidly approaches its limiting PDF that is symmetric in y. Then averaging the x equation leads to the prediction $d\bar{x} = 0\,dt$; that is, averaging predicts nothing happens. Yet the slow x variable must fluctuate through its forcing by the fast y. Second, and similar to illuminating coordinate transforms explored in in this article, modify the x and y variables to new coordinates X and Y where

$$x = X - \epsilon Y + \epsilon \int_{-\infty}^{t} e^{\tau - t}\,dW_\tau \quad \text{and} \quad y = Y + \int_{-\infty}^{t} e^{\tau - t}\,dW_\tau\,. \tag{2}$$

In the X and Y coordinates the SDE system (1) decouples to simply

$$dX = \epsilon\,dW \quad \text{and} \quad dY = -Y\,dt\,. \tag{3}$$

In these new coordinates $Y \to 0$ exponentially fast. Thus in the long term the only significant dynamics occurs in the modified slow variable X which system (3) shows undergoes a random walk. The method of averaging completely misses this random walk: true, the mean $\bar{x}$ remains at zero; but the growing spread about the mean is missed by averaging. Stochastic coordinate transforms such as (2) decouple fast and slow variables to empower us to extract accurate models for a true slow variable X. They are called 'normal form' transformations because this decoupling of stochastic dynamics is analogous to corresponding simplifications in deterministic systems [23, 3, e.g.]. This article establishes useful properties for such stochastic normal form coordinate transformations in modelling multiscale nonlinear stochastic dynamical systems.

One great advantage of basing modelling upon coordinate transforms is that exactly transformed dynamics fully reproduce the original dynamics for all time and all state space. It is only when we approximate the transformed dynamics that errors occur. Consequently, modelling errors can be much better controlled.

Stochastic ODEs and PDEs have many important applications. Here we restrict attention to nonlinear SDEs when the dynamics of the SDE has both long lasting slow modes and decaying fast modes [4, e.g.]. The aim underlying all the exploration in this article is to derive normal forms useful for macroscopic modelling of stochastic systems when the systems are specified at a detailed microscopic level. Thus we endeavour to separate *all* fast time processes from *all* slow processes [10, 30, e.g.]. Such separation is especially intriguing in stochastic systems as white noise has fluctuations on *all* time scales. In contrast, almost all previous approaches have been content to derive normal forms that support reducing the dimensionality of the dynamics. Here we go further than other researchers and *additionally and systematically separate fast time processes from the slow modes.*

Arnold and Imkeller [4, 3] developed a rigorous body of theory to support stochastic coordinate transforms to a normal form. They comment that the normal form transformation involves anticipating the noise processes, that is, involving integrals of the noise over a fast time scale of the future. However, in contrast to the examples of Arnold and Imkeller [4] [3, corrected], Sections 2 and 3 argue that such anticipation can be always removed from the slow modes with the result that no anticipation is required after the fast transients decay. Furthermore, Sections 2 and 3 argue that on the stochastic slow manifold all noise integrals can be removed from terms linear in the noise to leave a slow mode system, such as the simple $dX = \epsilon\, dW$ of the normal form (3), in which there are no fast time integrals at all. The arguments demonstrate that, except for some effects nonlinear in the noise, all fast time processes can be removed from the slow modes of a normal form of stochastic systems.

The theory of Arnold and Imkeller [4, 3] applies only to finite dimensional stochastic systems. Similarly, Du and Duan [14]'s theory of invariant manifold reduction for stochastic dynamical systems also only applies in finite dimensions. But many applications are infinite dimensional; for example, the discretisation of stochastic PDEs approximates an inertial manifold of stochastic dynamics [30]. Following the wide recognition of the utility of inertial manifolds [37, e.g.], Bensoussan and Flandoli [6] proved the existence of attractive stochastic inertial manifolds in Hilbert spaces. The stochastic slow manifolds obtained in Sections 2–4 via stochastic normal forms are examples of such stochastic inertial manifolds, albeit still in finite dimensions.

To derive a normal form we have to implement a coordinate transformation that simplifies a stochastic system. But the term 'simplify' means different things to different people depending upon how they wish to use the 'simplified' stochastic system. Our aim throughout this article is to create stochastic models that may efficiently simulate the long term dynamics of multiscale stochastic systems. This aim is a little different to that of previous researchers and so the results herein are a little different. For example, Coullet, Elphick and Tirapegui [11] and Arnold and Imkeller [4, 3] do not avoid fast time integrals because their aim is different. Principles that we require are the following:

1. Avoid unbounded (secular) terms in the transformation and the evolution (ensures uniform asymptotic approximations);

2. Decouple all the slow processes from the fast processes (ensures a valid long term model);

3. Insist that the stochastic slow manifold is precisely the transformed fast modes being zero;

4. Ruthlessly eliminate as many as possible of the terms in the evolution (to simplify at least the algebraic form of the SDEs);

5. Avoid as far as possible fast time memory integrals in the evolution (to endeavour to remove all fast time processes from the slow modes).

In general we can meet all these principles, although the last two are are only phrased as 'as far/many as possible': Section 2 explores the issues in a particular example stochastic system; whereas Section 3 presents general theory for finite dimensional, nonlinear, stochastic differential systems. The broad applicability of coordinate transforms empowers a web service to construct for you such stochastic normal forms [32]: enter any suitable system of SDEs into the web page and it will provide you with the stochastic normal form constructed according to the above principles to separate slow and fast stochastic dynamics.

Sri Namachchivaya, Leng and Lin [35, 36] emphasise the importance of effects quadratic in the stochastic noise "in order to capture the stochastic contributions of the stable modes to the drift terms of the critical modes." Sections 2–4 also address such important quadratic effects. The generic result of this normal form approach is that not all the memory integrals can be removed from the evolution of the stochastic slow variables: some terms quadratic in the noise retain fast time scale memory integrals.

Section 4 explores the implications of these results for macroscale simulation of stochastic systems. The normal form approach empowers us to address the effect of anticipatory integrals, the influence of the noise on averages, especially noise induced drift, and the failure of averaging to provide a systematic basis for macroscale simulation.

Lastly, Section 5 discusses in detail a normal form of a stochastically forced Hopf bifurcation, not because it is a Hopf bifurcation, but instead because it is a generic example of stochastic effects interacting nonlinearly with oscillatory dynamics. A complex valued, time dependent, coordinate transform can, with considerable care, derive a model SDE that is valid for simulating the long term evolution of the stochastic oscillating dynamics. A future application could be to the modelling of atmospheric white noise forcing of oceanic modes: Pierce [27] discusses this situation from an oceanographer's perspective.

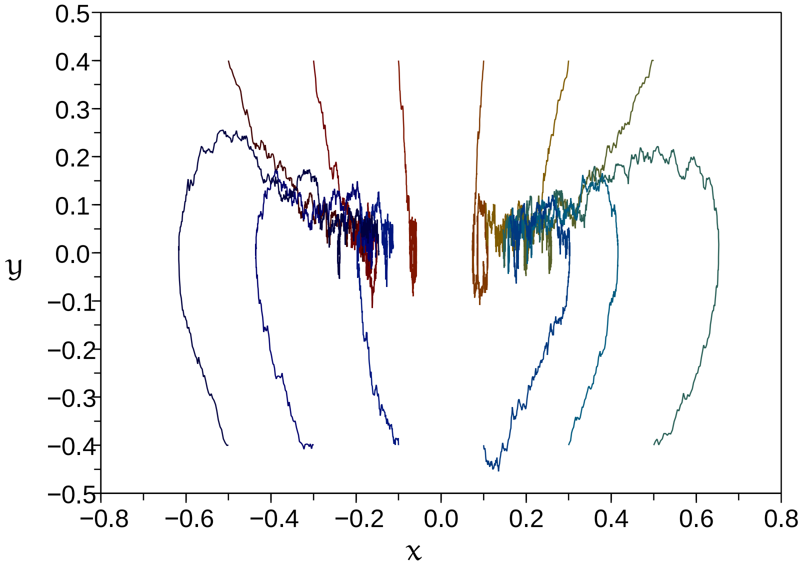

Figure 1: trajectories of the example stochastic system (5) from different initial conditions for different realisations of the noise, intensity $\sigma = 0.05$.

2 Explore in detail a simple nonlinear stochastic system

This section considers the dynamics of one of the most elementary, nonlinear, multiscale stochastic systems:

$$dx = -xy\,dt \quad \text{and} \quad dy = (-y + x^2 - 2y^2)dt + \sigma\,dW\,. \tag{4}$$

Figure 1 plots some typical trajectories of the SDE system (5). In this finite domain near the origin the y variable decays exponentially quickly to $y \approx x^2$; whereas the x variable evolves relatively slowly over long times. Our challenge is to separate, amid the noise, the slow $x(t)$ from the fast $y(t)$. The modelling issues raised, and their resolution, in this relatively simple stochastic system are generic as seen in Section 3. I emphasise that the detailed analysis of this simple example demonstrates the inevitability of the proposed methodology when you set out, as we do here, to systematically separate slow and fast processes in nonlinear stochastic systems.

Throughout this article I adopt the Stratonovich interpretation of SDEs, as does Arnold and Imkeller [4, 3], so that the usual rules of calculus apply. To ease asymptotic analysis I also adopt hereafter the notation of applied physicists and engineers. Thus I formally explore the SDE system (4) in the equivalent form of

$$\dot{x} = -xy \quad \text{and} \quad \dot{y} = -y + x^2 - 2y^2 + \sigma\phi(t)\,, \tag{5}$$

where overdots denote formal time derivatives and the 'white noise' $\phi(t)$ is the formal time derivative of the Wiener process $W(t)$. Both the Stratonovich interpretation and the adoption of this formal notation empowers the use of computer algebra to handle the multitude of details in examples.

The challenge is to adapt the deterministic normal form transformation, Section 2.1, to the stochastic system (5) in order to not only decouple the interesting slow modes, but to simplify them as far as possible, Section 2.2. The analysis and argument is detailed in order to demonstrate in a simple setting how Principles 1–5 are realised at the expense of having to anticipate future noise. Those familiar with the concept of stochastic normal forms could skip to Section 3 for generic arguments of the new results.

2.1 Decouple the deterministic dynamics

Initially consider the example toy system (5) when there is no noise, $\sigma = 0$. A deterministic, near identity, normal form coordinate transform decouples the deterministic slow and fast dynamics:

$$x = X + XY + \tfrac{3}{2}XY^2 - 2X^3Y + \tfrac{5}{2}XY^3 + \cdots\,, \tag{6}$$

$$y = Y + X^2 + 2Y^2 + 4Y^3 - 4X^2Y^2 + 8Y^4 + \cdots\,. \tag{7}$$

Figure 2 shows the coordinate curves of this (X, Y) coordinate system. In the new (X, Y) coordinate system, the evolution of the toy system (5) becomes

$$\dot{X} = -X^3 \quad \text{and} \quad \dot{Y} = -(1 + 2X^2 + 4X^4)Y + \cdots\,. \tag{8}$$

Observe the Y-dynamics are that of exponentially quick decay to the slow manifold $Y = 0$ at the X dependent rate $(1 + 2X^2 + 4X^4 + \cdots)$. Substituting $Y = 0$ into the transform (6) and (7) shows this slow manifold is the curve $x = X$ and $y = X^2$ [15]. The dynamics on this slow manifold, $\dot{X} = -X^3$ from (8), form the accurate, macroscopic, long term model.

The slow X dynamics are *also* independent of the Y variable and thus the initial value $Y(0)$ and subsequent $Y(t)$ are immaterial to the long term evolution. Thus to make accurate forecasts, project onto the slow manifold $Y = 0$ along the coordinate curves of constant X seen in Figure 2 [12]. Equivalently, because the slow X dynamics are independent of the Y variable, the dynamics of the system (5) map the curves of constant X in Figure 2 into other curves of constant X. Thus initial conditions on any one curve of constant X all evolve towards the same trajectory on the slow manifold.

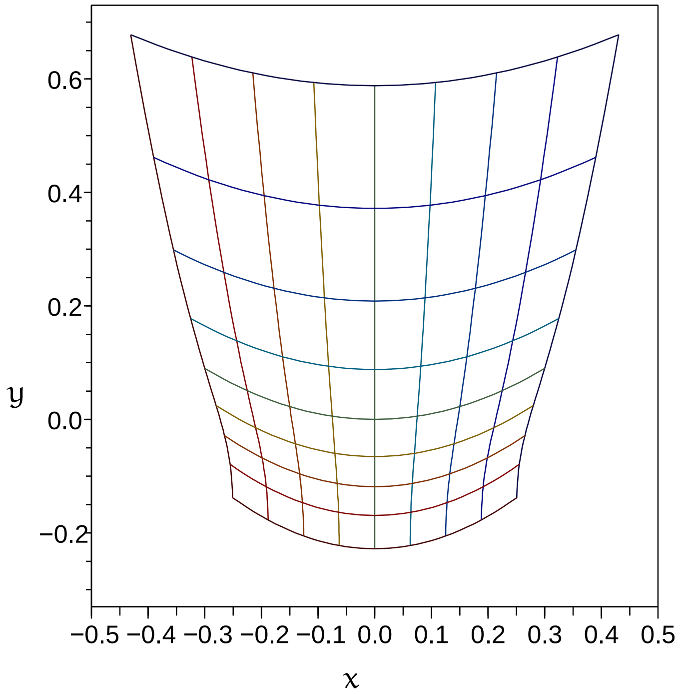

Figure 2: coordinate curves in the xy-plane of the (X, Y) coordinate system that simplifies to (8) the algebraic description of the dynamics of the deterministic ($\sigma = 0$) system (5).

But these comments are all for deterministic dynamics, $\sigma = 0$. The next subsection answers the question: how can we adapt this beautifully simplifying coordinate transform to cater for stochastic dynamics?

2.2 Simplify stochastic evolution as far as possible

Now explore the construction of a coordinate transform that decouples the fast and slow dynamics of the toy SDE (5) in the presence of its stochastic forcing. In order to cater for the stochastic fluctuations, the coordinate transform must be time dependent through dependence upon the realisation of the noise, as shown schematically in Figure 3. This subsection is detailed in order to argue that no alternatives go unrecognised. The method is to iteratively refine the stochastic coordinate transform based upon the residuals of the governing toy SDE (5).

Although our focus is on the case when $\phi(t)$ is a white noise, because we use the usual calculus of the Stratonovich interpretation, the algebraic results also apply to smoother processes $\phi(t)$. For two examples, the forcing $\phi(t)$

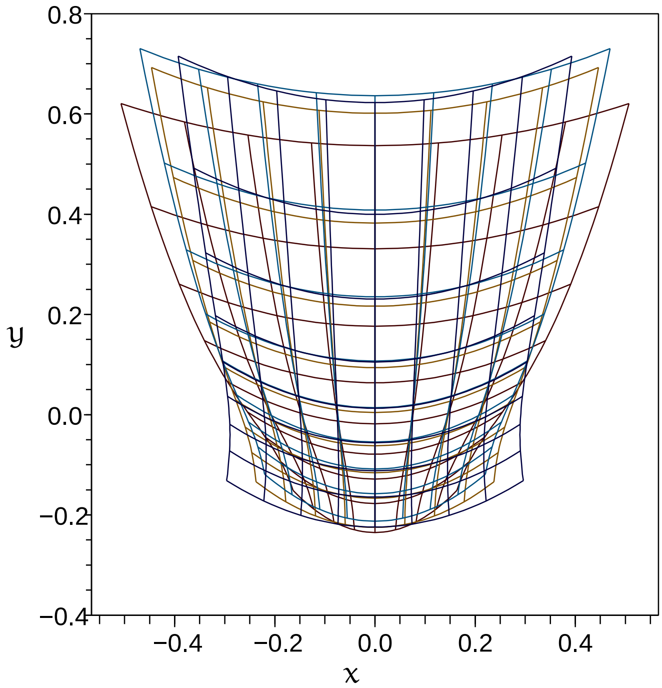

Figure 3: four different (coloured) meshes represent *either* four realisations sampled at one instant *or* one realisation sampled at four instants of the coordinate curves in the xy-plane of the *stochastic* (X, Y) coordinate system that simplifies to (26) and (23) the algebraic description of the dynamics of the stochastic ($\sigma = 0.2$) system (5).

could be the output of a deterministic chaotic system [19, e.g.], or the forcing $\phi(t)$ could be even as regular as a periodic oscillator. Thus the algebraic expressions derived herein apply much more generally than to just white noise ϕ. However, the *justification for the particular coordinate transform* often depends upon the peculiar characteristics of white noise. For forcing ϕ which is smoother than white noise, although our results herein apply, other coordinate transforms *may be preferable* in order to achieve other desirable outcomes in the transformation (outcomes not attainable when ϕ is white noise). These possibilities are not explored. Instead, throughout this article the forcing $\phi(t)$ denotes a white noise process in a Stratonovich interpretation of SDEs.

Let us proceed to iteratively develop a stochastic coordinate transform of the SDE (5) via stepwise refinement [28].

First, simplifying fast dynamics introduces memory With $x \approx X$ and $\dot X \approx 0$, seek a change to the y coordinate of the form

$$y = Y + \eta'(t,X,Y) + \cdots \quad \text{and} \quad \dot Y = -Y + G'(t,X,Y) + \cdots\,, \tag{9}$$

where η' and G' are small, $\mathcal{O}(\epsilon^2)$, corrections to the transform and the corresponding evolution. I introduce the parameter ϵ to provide a convenient ordering of the terms that arise in the algebra: formally set $\epsilon = |(X,Y,\sigma)|$ with the effect that ϵ counts the number of X, Y and σ factors in any one term. In this section all errors are measured in orders of ϵ: thus all asymptotic expressions are *local* to $(X,Y,\sigma) = (0,0,0)$, that is, valid in some neighbourhood of the origin; later sections give examples *globally* valid in one or more variables. Substitute (9) into the y SDE (5) and drop products of small corrections to recognise we need to solve

$$G' + \frac{\partial \eta'}{\partial t} + \eta' - Y\frac{\partial \eta'}{\partial Y} = \sigma\phi(t) + X^2 - 2Y^2 + \cdots\,; \tag{10}$$

partial derivatives are here done keeping constant the other two variables of X, Y and t.

First, keep the deterministic evolution as simple as possible (Principle 4) by not changing the evolution, $G' = 0$, and by modifying the coordinate transform by $\eta' = X^2 + 2Y^2$.

Second, consider the remaining stochastic term $\sigma\phi(t)$ in the right-hand side of (10). Keep the Y dynamics as simple as possible (Principle 4) by choosing the convolution $\sigma e^{-t}\star\phi$, defined in (12), to be part of the correction η' to the coordinate transform. Consequently the new approximation of the coordinate transform and the dynamics is

$$y = Y + X^2 + 2Y^2 + \sigma e^{-t}\star\phi + \cdots \quad \text{and} \quad \dot Y = -Y + \cdots\,. \tag{11}$$

In these leading order terms of the coordinate transform, see the stochastic slow manifold (SSM) $Y = 0$ corresponds to the vertically fluctuating parabola $y \approx X^2 + \sigma e^{-t}\star\phi$ as seen in the overall vertical displacements of the coordinate mesh in Figure 3.

The convolution For any non-zero parameter μ, and consistent with the convolution in the example transform (2), define the convolution, for sufficiently well behaved stochastic processes $V(t)$,

$$e^{\mu t}\star V = \begin{cases} \int_{-\infty}^{t} \exp[\mu(t-\tau)]V(\tau)\,d\tau\,, & \mu < 0\,, \\ \int_{t}^{+\infty} \exp[\mu(t-\tau)]V(\tau)\,d\tau\,, & \mu > 0\,, \end{cases} \tag{12}$$

so that the convolution is always with a bounded exponential (Principle 1). Such convolutions are used throughout this article. Five useful properties of this convolution are

$$e^{\mu t}\star 1=\frac{1}{|\mu|}\,, \tag{13}$$

$$\frac{d}{dt}e^{\mu t}\star V=-\operatorname{sgn}\mu\, V+\mu e^{\mu t}\star V\,, \tag{14}$$

$$E[e^{\mu t}\star V]=e^{\mu t}\star E[V]\,, \tag{15}$$

$$E[(e^{\mu t}\star \phi)^2]=\frac{1}{2|\mu|}\,, \tag{16}$$

$$e^{\mu t}\star e^{\nu t}\star=\begin{cases}\frac{1}{|\mu-\nu|}\left[e^{\mu t}\star+e^{\nu t}\star\right], & \mu\nu<0\,,\\ \frac{-\operatorname{sgn}\mu}{\mu-\nu}\left[e^{\mu t}\star-e^{\nu t}\star\right], & \mu\nu>0 \text{ and } \mu\neq\nu\,.\end{cases} \tag{17}$$

Also remember that although with $\mu<0$ the convolution $e^{\mu t}\star$ integrates over the past, with $\mu>0$, as we will soon need, the convolution $e^{\mu t}\star$ integrates into the future; both integrate over a time scale of order $1/|\mu|$.

Second, split noise to eliminate memory from slow dynamics Seek a correction to the slow component of the stochastic coordinate transform of the form

$$x=X+\xi'(t,X,Y)+\mathcal{O}\big(\epsilon^3\big)\quad\text{and}\quad \dot X=F'(t,X,Y)+\mathcal{O}\big(\epsilon^3\big)\,. \tag{18}$$

where ξ' and F' are $\mathcal{O}\big(\epsilon^2\big)$ corrections to the transform and the corresponding evolution. Substitute into the x equation of SDE (5) and omit small products to recognise we need to solve

$$F'+\frac{\partial\xi'}{\partial t}-Y\frac{\partial\xi'}{\partial Y}=-XY+\sigma Xe^{-t}\star\phi+\mathcal{O}\big(\epsilon^3\big)\,. \tag{19}$$

First, keep the deterministic evolution unchanged, $F'=0$, by choosing $\xi'=XY$ in the coordinate transform. Second, consider the stochastic part of the equation: $F'+\xi'_t-Y\xi'_Y=\sigma Xe^{-t}\star\phi$. The ξ'_Y cannot help us solve this stochastic part as there is no Y factor in the right-hand side term. We cannot integrate the forcing ϕ into the coordinate transform ξ' as then terms would grow like the Wiener process $W=\int\phi\,dt$ (Principle 1). To avoid a fast time convolution in the slow evolution F' (Principle 5), formally integrate by parts to split $e^{-t}\star\phi=-\phi+e^{-t}\star\dot\phi$ and hence choose components $F'=-\sigma X\phi$ and $\xi'=\sigma Xe^{-t}\star\phi$. Consequently,

$$x=X+XY+\sigma Xe^{-t}\star\phi+\mathcal{O}\big(\epsilon^3\big)\quad\text{and}\quad \dot X=-\sigma X\phi+\mathcal{O}\big(\epsilon^3\big)\,. \tag{20}$$

Observe that the slow dynamics being $\dot{X} \approx -\sigma X\phi$ is one example of additive noise appearing as multiplicative noise in the true slow variable.

In contrast, earlier approaches to stochastic bifurcation do not split the noise and consequently they derive more complicated normal forms that additionally have fast time memory processes. For example, Arnold [3] analyses a two variable fast/slow stochastic system, and presents in his equation (8.5.47) a normal form for the slow mode. Whereas his normal form is perfectly good for exploring bifurcations, it is inadequate for long time, macroscale modelling as it contains several fast time scale memory integrals of the sort we avoid (Principle 5).

Third, split noise to avoid memory in fast dynamics Seek corrections, η' and G', to the y transform and Y evolution driven by the updated residual of the y equation in SDE (5):

$$G' + \frac{\partial\eta'}{\partial t} + \eta' - Y\frac{\partial\eta'}{\partial Y} = -4\sigma Y e^{-t}\star\phi - 2\sigma^2(e^{-t}\star\phi)^2 + \mathcal{O}\big(\epsilon^3\big)\,. \tag{21}$$

Separately consider the two stochastic forcing terms on the right-hand side.

- To solve $G' + \eta'_t + \eta' - Y\eta'_Y = -4\sigma Y e^{-t}\star\phi$ we must seek G' and η' proportional to Y, whence $\eta' - Y\eta'_y = 0$. Integration by parts enables us to choose $G' = -4\sigma Y\phi$ and $\eta' = 4\sigma Y e^{-t}\star\phi$ to avoid secular terms in η' (Principle 1), and to avoid fast time convolution in the Y evolution (Principle 5).
- The quadratic noise term on the right-hand side is no problem: keep evolution unchanged, $G' = 0$ (Principle 4), and then the convolution $\eta' = -2\sigma^2 e^{-t}\star(e^{-t}\star\phi)^2$ corrects the coordinate transform.[1]

Consequently, the fast time transform and dynamics are more accurately

$$\begin{aligned} y &= Y + X^2 + 2Y^2 + \sigma\big[e^{-t}\star\phi + 4Ye^{-t}\star\phi\big] \\ &\quad - 2\sigma^2 e^{-t}\star(e^{-t}\star\phi)^2 + \mathcal{O}\big(\epsilon^3\big)\,, \end{aligned} \tag{22}$$

$$\dot{Y} = -Y - 4\sigma Y\phi + \mathcal{O}\big(\epsilon^3\big)\,. \tag{23}$$

[1]The right-hand side of this correction η' has non-zero mean. We could assign the mean, $-\sigma^2$, into the Y evolution as a mean (downwards) forcing, but then this destroys $Y = 0$ as being the slow manifold, contradicting Principle 3.

Fourth, quadratic noise normally appears in the slow dynamics Seek corrections to the transform and evolution, ξ' and F', driven by the updated residual of the x equation of SDE (5):

$$F' + \frac{\partial \xi'}{\partial t} - Y\frac{\partial \xi'}{\partial Y} = -X^3 - 3XY^2 + \sigma XY(5\phi - 6e^{-t}\star\phi) + \sigma^2 X\left[\phi e^{-t}\star\phi - (e^{-t}\star\phi)^2 + 2e^{-t}\star(e^{-t}\star\phi)^2\right] + \mathcal{O}\big(\epsilon^4\big)\,. \quad (24)$$

Consider the right-hand side term by term:

- We choose $F' = -X^3$ and $\xi' = \frac{3}{2}XY^2$ in the traditional deterministic approach.
- To match the term linear in noise, $F' + \xi'_t - Y\xi'_Y = \sigma XY(5\phi - 6e^{-t}\star\phi)$, we must seek F' and ξ' proportional to XY, whence $\xi'_t - Y\xi'_y \mapsto \xi'_t - \xi'$. Consequently foreknowledge, anticipation, of the noise appears. Consider the two cases:
  - allowing anticipation (implementing Principle 4) and in accord with Arnold and Imkeller [4, 3], we assign all of this term to the coordinate transformation with $F' = 0$ and $\xi' = \sigma XY(3e^{-t}\star\phi - 2e^{+t}\star\phi)$;
  - disallowing anticipation, we must assign all of this term into the X evolution by assigning $F' = \sigma XY(5\phi - 6e^{-t}\star\phi)$ and $\xi' = 0$—the difficulty here being that the evolution to the SSM then depends undesirably upon Y, contradicting Principle 2. We have to allow anticipation in the coordinate transformation.

  Many more anticipatory convolutions appear in higher order terms, namely those with Y factors. However, they need never occur in the evolution on the SSM where $Y = 0$ (Proposition 2).
- For the quadratic noise term in (24), seek contributions to the solution which are proportional to X; consequently, on the left-hand side $-Y\xi'_Y = 0$. At least part of the fluctuations cannot be assigned into the transform ξ' as the integral of noise is a Wiener process which almost surely is secular (Principle 1).

  Now, as in the earlier integration by parts, separate these quadratic noise terms into

$$\begin{aligned}(e^{-t}\star\phi)^2 &= \phi e^{-t}\star\phi - \tfrac{1}{2}\tfrac{d}{dt}[(e^{-t}\star\phi)^2]\,,\\ e^{-t}\star(e^{-t}\star\phi)^2 &= (e^{-t}\star\phi)^2 - \tfrac{d}{dt}[e^{-t}\star(e^{-t}\star\phi)^2]\end{aligned}$$

$$= \phi e^{-t}\star\phi - \tfrac{d}{dt}[\tfrac{1}{2}(e^{-t}\star\phi)^2 + e^{-t}\star(e^{-t}\star\phi)^2]\,,$$

and so these contribute corrections $F' = \sigma^2 X\phi e^{-t}\star\phi$ and $\xi' = -\sigma^2 X(\frac{1}{2} + 2e^{-t}\star)(e^{-t}\star\phi)^2$.

The upshot is that the x transformation and X evolution is more accurately

$$\begin{aligned} x &= X + XY + \tfrac{3}{2}XY^2 + \sigma\big[Xe^{-t}\star\phi + XY(3e^{-t}\star\phi - 2e^{+t}\star\phi)\big] \\ &\quad - \sigma^2 X(\tfrac{1}{2} + 2e^{-t}\star)(e^{-t}\star\phi)^2 + \mathcal{O}\big(\epsilon^4\big)\,, && (25) \\ \dot X &= -X^3 - \sigma X\phi + 2\sigma^2 X\phi e^{-t}\star\phi + \mathcal{O}\big(\epsilon^4\big)\,. && (26) \end{aligned}$$

Higher order modelling Further algebra constructs higher order stochastic coordinate transform from the original (x, y) variables to the new (X, Y) variables. For later discussion, the web service [32] informs us that the dynamics of the example SDE (5) is

$$\begin{aligned} \dot X &= -X^3 - \sigma X\phi + 2\sigma^2 X\phi e^{-t}\star\phi \\ &\quad - 4\sigma^2 X^3\phi e^{-t}\star e^{-t}\star\phi + \mathcal{O}\big(\epsilon^6, \sigma^3\big)\,, && (27) \\ \dot Y &= -(1 + 2X^2 + 4X^4)Y - 4\sigma(1 + X^2)Y\phi + 8\sigma^2 Y\phi e^{-t}\star\phi \\ &\quad + 4\sigma^2 X^2 Y\phi\big[3e^{-t}\star\phi - e^{+t}\star\phi - 2e^{-t}\star e^{-t}\star\phi\big] \\ &\quad + \mathcal{O}\big(\epsilon^6, \sigma^3\big)\,. && (28) \end{aligned}$$

Expect that the series solutions we generate for a coordinate transform and normal form are divergent. It is rare to find convergence of these sort of asymptotic expansions of nonlinear dynamics. Nonetheless, there exist a coordinate transform and normal form in a finite neighbourhood to which they are the asymptotic approximations, see Section 3.

Irreducible fast time convolutions generate drift As is generally true, the X and Y evolution equations, (27) and (28), contain algebraically irreducible nonlinear noise such as $\phi e^{-t}\star\phi$, in defiance of Principle 5. Over long times I recommend such irreducible noise be replaced by $\frac{1}{2} + \frac{1}{\sqrt{2}}\phi^{(1)}$ for some effectively new white noise $\phi^{(1)}(t)$ [10]. Such replacement was also justified by Khasminskii (1996) as described by Sri Namachchivaya and Leng [35]. Importantly, such quadratic noise, in effect, generates a mean deterministic drift term in the slow dynamics [35, 36, e.g.]. In applications such drifts can be vital.

The average SSM is not the deterministic slow manifold For the toy SDE (5), Section 2.1 shows the deterministic slow manifold is $y = x^2$. In general the SSM fluctuates about a mean location which is different to this deterministic slow manifold. From (25) and (22) with fast variable $Y = 0$, the SSM is

$$x = X + \sigma X e^{-t}\star\phi - \sigma^2 X(\tfrac{1}{2} + 2e^{-t}\star)(e^{-t}\star\phi)^2 + \mathcal{O}(\epsilon^4), \quad (29)$$

$$y = X^2 + \sigma e^{-t}\star\phi - 2\sigma^2 e^{-t}\star(e^{-t}\star\phi)^2 + \mathcal{O}(\epsilon^3), \quad (30)$$

Take expectations, and using (15) and (16),

$$E[x] = (1 - \tfrac{5}{4}\sigma^2)X + \mathcal{O}(\epsilon^4) \quad \text{and} \quad E[y] = X^2 - \sigma^2 + \mathcal{O}(\epsilon^3). \quad (31)$$

Observe $E[y] \neq E[x]^2$, instead $E[y] \approx (1 + \tfrac{5}{2}\sigma^2)E[x]^2 - \sigma^2$ so that the average SSM is a displaced steeper parabola shape than the deterministic slow manifold. It is quadratic noise processes that deform the average SSM from the deterministic.

2.3 Forecast from initial conditions

Suppose at time $t = 0$ we observe the state (x_0, y_0), what forecast can we make with the SSM SDE (27)? Revert the asymptotic expansion of the stochastic coordinate transform (25) and (22) to deduce

$$X = x + x^3 - xy + \tfrac{3}{2}xy^2 + 2\sigma xy e^{+t}\star\phi - 2\sigma^2 x(e^{+t}\star\phi)(e^{-t}\star\phi) + \mathcal{O}(\epsilon^4), \quad (32)$$

$$Y = y - x^2 - 2y^2 - \sigma e^{-t}\star\phi + \sigma^2(1 + e^{-t}\star)(e^{-t}\star\phi)^2 + \mathcal{O}(\epsilon^3). \quad (33)$$

Then the correct initial condition for the long term dynamics on the SSM, governed by the SDE (27), is the X component of this reversion, (32), evaluated at the observed state, namely

$$X(0) = x_0 + x_0^3 - x_0y_0 + \tfrac{3}{2}x_0y_0^2 + 2\sigma x_0y_0 e^{+t}\star\phi - 2\sigma^2 x_0(e^{+t}\star\phi)(e^{-t}\star\phi) + \mathcal{O}(\epsilon^4). \quad (34)$$

This is a projection of the observed initial state onto the SSM to provide an initial condition $X(0)$ for the slow mode. However, this projection involves both memory and anticipatory convolutions of the noise. There are at least three interesting issues with computing this initial $X(0)$. First, at the initial instant we do not know either the future nor the past, so the terms involving the noise ϕ are unknown. Using the expectations (15) and (16), the projection $X(0)$ has known mean

$$E[X(0)] = x_0 + x_0^3 - x_0y_0 + \tfrac{3}{2}x_0y_0^2 + \mathcal{O}(\epsilon^4),$$

with known variance

$$\operatorname{Var}[X(0)] \approx 2\sigma^2 x_0^2 y_0^2 + \sigma^4 x_0^2 .$$

That is, a given observed state (x_0, y_0) corresponds to a stochastic state for the evolution of the slow mode model on the SSM.

Second, but if this state $X(0)$ for the slow mode is to be used in a simulation to make forecasts of the future, then we know the future of the noise ϕ. The future values of noise ϕ are just those we use in integrating the slow mode SDE (27). Thus for simulation, we do eventually know the anticipatory convolutions $e^{+t}\star\phi$ in (34), but not the memory convolution $e^{-t}\star\phi$. In this case the mean of the projection

$$E[X(0)] = x_0 + x_0^3 - x_0 y_0 + \tfrac{3}{2} x_0 y_0^2 + 2\sigma x_0 y_0 e^{+t}\star\phi + \mathcal{O}\big(\epsilon^4\big) ,$$

with variance $\operatorname{Var}[X(0)] \approx 2\sigma^4 x_0^2 (e^{+t}\star\phi)^2$.

Lastly, if we made additional observations for times $t < 0$, then the additional information could partially determine the past history of the noise ϕ and hence help us estimate the memory convolution $e^{-t}\star\phi$. These three cases emphasise that the initial state $X(0)$ of the slow variable depends upon more than just the observed state (x_0, y_0) at an initial instant.

Avoiding anticipation is less useful Alternatively, suppose we disallow anticipatory convolutions. Arguments show that we can reasonably abandon only Principle 2, the requirement to completely decouple the slow modes from the fast modes. But abandoning this principle also means we are no longer able to use the slow model to make high accuracy forecasts from every initial condition.

Suppose we specify some initial state (X_0, Y_0), either deterministic or stochastic. What forecast can we easily make with the SSM model (27)? In general, none. The reason is that in the evolution to the SSM, the slow X dynamics are coupled to the fast Y dynamics. But the point of deriving a slow model, for most purposes, is to avoid resolving the details of the fast dynamics; thus we cannot rationally project from (X_0, Y_0) onto the SSM. In contrast, the normal form of a deterministic system empowers rational projection from nearby initial conditions onto the slow model for accurate forecasts [12]. Abandoning Principle 2 means we cannot make accurate forecasts.

Because it adheres to Principle 2, the stochastic normal form of Section 2.2, similarly to the deterministic normal form, empowers rational projection from nearby initial conditions onto the SSM. But there is a catch: in order to do the projection we need to anticipate the noise. Since we generally will not know the future noise, the stochastic normal form of Section 2.2 also cannot be used for accurate forecasting. In this sense the two stochastic normal forms would have equivalent power. However, there is a difference. The anticipatory stochastic normal form of Section 2.2 has explicit convolutions for the projection: we may not know what they are, but we could certainly use the convolutions to estimate bounds and distributions for the projection of initial conditions. In contrast, stochastic normal forms that couple the slow and the fast modes keeps such information encrypted in the coupled fast and slow dynamics. Consequently, *maintaining Principle 2, decoupling the slow modes from the fast, appears more powerful than avoiding anticipatory convolutions.*

3 Normal forms of SDEs for long term modelling

This section uses formal arguments to establish a couple of key generic properties of stochastic normal forms seen in the example SDE system of the previous section. We establish firstly that a stochastic coordinate transform can decouple slow modes from fast modes, to make the stochastic slow manifold (SSM) easy to see, and secondly that although anticipation of the noise may be necessary in the full transform no anticipation need appear on the SSM.

Consider a general system of Stratonovich SDEs in $m+n$ dimensions for variables $x(t)\in\mathbb{R}^m$ and $y(t)\in\mathbb{R}^n$:

$$\dot x = Ax + f(x,y,t)\,, \tag{35}$$

$$\dot y = By + g(x,y,t)\,, \tag{36}$$

where

- the spectrum of A is zero and for simplicity we assume A is upper triangular with all elements zero except possibly $A_{i,j}$ for $j>i$ (such as in the Jordan form appropriate for position and velocity variables of a mechanical system);
- for simplicity assume matrix B has been diagonalised with diagonal elements $\beta_1,\ldots,\beta_n$, possibly complex, with $\Re\beta_j<0$;[2]

[2] If matrix B is in Jordan form, rather than diagonalisable, then extensions of the arguments lead to the same results.

- $\mathbf{f}$ and $\mathbf{g}$ are stochastic functions that are "nonlinear", that is, f and g and their gradients in $\mathbf{x}$ and $\mathbf{y}$ are all zero at the origin;
- the stochastic nature of the system of SDEs arises through the dependence upon the time t in the nonlinearity $\mathbf{f}$ and $\mathbf{g}$—assume the time dependence is implicitly due to some number of independent white noise processes $\phi_k(t)$ (which are derivatives of independent Wiener processes).

For such systems, Boxler [8] guarantees the existence, relevance and approximability of a stochastic centre manifold for (35–36) in some finite neighbourhood of the origin. We call this a stochastic slow manifold (SSM) because we assume matrix A does not have complex eigenvalues (an extension to oscillatory dynamics is explored in Section 5).

For example, the toy SDE system (5) takes the form (35–36) with variables $\mathbf{x} = (\sqrt{\sigma}, x)$ and $\mathbf{y} = y$, then

$$A = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad B = -1\,, \quad \mathbf{f} = \begin{bmatrix} 0 \\ -xy \end{bmatrix}, \quad \mathbf{g} = x^2 - 2y^2 + \sqrt{\sigma}^2\phi(t)\,.$$

In principle, the matrices A and B could also depend upon the realisation of the noise. When the Lyapunov exponents of the corresponding linear dynamics are zero and negative respectively, then a stochastic centre manifold still exists and has nice properties [8, 3]. However, here I restrict attention to the algebraically more tractable case when the basic linear operators A and B are deterministic.

Stochastic singular perturbation systems such as those explored by Berglund and Gentz [7], are a subset of the systems encompassed by (35–36). For example, let us transform the deterministic singular perturbation system

$$\dot{x} = f(x,y)\,, \quad \dot{y} = \frac{1}{\epsilon} g(x,y)\,, \tag{37}$$

into the form (35–36). First, change to the fast time $\tau = t/\epsilon$ so that

$$\frac{dx}{d\tau} = \epsilon f(x,y)\,, \quad \frac{dy}{d\tau} = g(x,y)\,.$$

Then change to a coordinate system ξ and η, where $\eta = 0$ is the curve $g(x,y) = 0$, in which the system takes the form

$$\frac{d\xi}{d\tau} = \epsilon F(\xi,\eta)\,, \quad \frac{d\eta}{d\tau} = \beta(\xi)\eta + G(\xi,\eta)\,.$$

Consequently, in variables $x = (\sqrt{\epsilon}, \xi)$ and $y = \eta$, the curve $(x, y) = (0, \xi, 0)$ is a set of equilibria, at each equilibria the dynamics are of the form (35–36). Consequently there exists a slow manifold around each point of the curve, which as a whole forms a slow manifold in a neighbourhood of the curve [9]: consequently, an asymptotic approximation to a normal form coordinate transform is global in variable ξ and local in parameter $\sqrt{\epsilon}$ and variable η. The analysis of this section also applies to singular perturbation problems by a change in time scale and coordinate system.

A stochastic coordinate transform We transform the SDE (35–36) in (x, y) to a new (X, Y) coordinate system by a stochastic, near identity, coordinate transform

$$x = X + \xi(X, Y, t) \quad \text{and} \quad y = Y + \eta(X, Y, t) . \tag{38}$$

This stochastic coordinate transform is to be chosen such that the SDE (35–36) transforms to a 'simpler' form for multiscale modelling. Based upon the experience of Section 2.2, we seek to simplify the SDEs according to Principles 1–5, and allowing anticipation.

3.1 Transform the fast dynamics

Suppose (38) is some approximation to the desired coordinate transform. Iteratively we seek corrections ξ' and η' to the transform, namely

$$x = X + \xi(X, Y, t) + \xi'(X, Y, t) \quad \text{and} \quad y = Y + \eta(X, Y, t) + \eta'(X, Y, t) . \tag{39}$$

Find corrections such that the corresponding updates to the evolution, say F' and G' in

$$\dot{X} = AX + F(X, Y, t) + F'(X, Y, t) , \tag{40}$$

$$\dot{Y} = BY + G(X, Y, t) + G'(X, Y, t) , \tag{41}$$

are as simple as possible (Principle 4).

For the fast dynamics, the iteration is to substitute the corrected transform (39) and evolution (40)–(41) into the governing SDE (36) for the fast variables. Then drop products of corrections as being negligible, and approximate coefficients of corrections by their leading order term. Then the equation for the jth component of the correction to the transform of the fast variable and the new fast dynamics is the stochastic version of the usual homological equation

$$G'_j + \frac{\partial \eta'_j}{\partial t} - \beta_j \eta'_j + \sum_{\ell=1}^{n} \beta_\ell Y_\ell \frac{\partial \eta'_j}{\partial Y_\ell} = \text{Res}_{36,j} , \tag{42}$$

where $\text{Res}_{36,j}$ denotes the residual of the jth component of the SDE (36). In constructing a coordinate transform we repeatedly solve equations of this form to find corrections.

We find what sort of terms may be put into the transformation η and what terms have to remain in the $\mathbf{Y}$ evolution by considering the possibilities for the right-hand side. The transform is constructed as a multivariate asymptotic expansion about the origin in $(\mathbf{X}, \mathbf{Y})$ space. Suppose the right-hand side, the residual $\text{Res}_{36,j}$, has, potentially among many others, a term of the multinomial form

$$c(t)\mathbf{X}^{\mathbf{p}}\mathbf{Y}^{\mathbf{q}} = c(t)\prod_{i=1}^{m} X_i^{p_i}\prod_{j=1}^{n} Y_j^{q_j}\,,$$

for some vectors of integer exponents $\mathbf{p}$ and $\mathbf{q}$. Because of the special form of the homological operator on the left-hand side of (42), seek contributions to the corrections of $G_j' = a(t)\mathbf{X}^{\mathbf{p}}\mathbf{Y}^{\mathbf{q}}$ and $\eta_j' = b(t)\mathbf{X}^{\mathbf{p}}\mathbf{Y}^{\mathbf{q}}$. Then this component of (42) becomes

$$a + \dot{b} - \mu b = c \quad \text{where} \quad \mu = \beta_j - \sum_{\ell=1}^{n} q_\ell\beta_\ell\,. \tag{43}$$

Three cases arise.

1. In the resonant case $\mu = 0$, we need to satisfy $a + \dot{b} = c$ where we want to put as much into b as possible (Principle 4). Neither of the mean and stochastically fluctuating components of the forcing $c(t)$ can be integrated into b as they both give rise to secular terms (Principle 1): the stochastically fluctuating part of c almost surely generates square-root growth.[3] Thus the generic solution is $a = c$ and $b = 0$, that is, assign $c(t)\mathbf{X}^{\mathbf{p}}\mathbf{Y}^{\mathbf{q}}$ to the $\mathbf{Y}$ evolution and nothing into the coordinate transform $\boldsymbol{\eta}$.

   In general, since $\Re\beta_\ell$ are *all* negative,[4] this case of $\mu = 0$ can only arise when at least one of the exponents $\mathbf{q}$ is positive in order for the sum in (43) to be zero. Hence, there will be at least one Y_ℓ factor in updates $\mathbf{G}'$ to the $\mathbf{Y}$ evolution, and so we maintain that $\mathbf{Y} = \mathbf{0}$ is the SSM.

[3] In contrast, when the forcing $c(t)$ is periodic, instead of stochastic, then the the forcing $c(t)$ may be integrated into the coordinate transform b, instead of being assigned to the evolution a.

[4] More generally, provided the eigenvalues β_j are all non-zero whether real or complex, the argument still holds. Thus we can maintain Principle 3 over a very wide range of circumstances.

2. When $\Re\mu < 0$, a solution of (43) is to place all the forcing into the SSM, $b = e^{\mu t} \star c$, and do not introduce a component into the $\mathbf{Y}$ evolution, $a = 0$. As $\Re\mu < 0$, the convolution is over the past history of the noise affected forcing $c(t)$; the convolution represents a memory of the forcing over a time scale of $1/|\Re\mu|$.

   However, for large enough exponents $\mathbf{q}$, that is for high enough order in Y, the rate $\Re\mu$ must eventually become positive. In the transition from negative to positive, the rate $\Re\mu$ may become close to zero. Then the time scale $1/|\Re\mu|$ becomes large and may be as large as the macroscopic time scale of the slow dynamics of interest. In that case set the transform $b = 0$ and assign this term in the forcing into the $\mathbf{Y}$ evolution with $a = c$. The intended use of a macroscopic model defines a slow time scale and consequently affects which terms appear in the model.

3. When $\Re\mu > 0$, and accepting anticipation in the transform, we simply set $b = e^{\mu t} \star c$, and do not change the $\mathbf{Y}$ evolution, $a = 0$.

Consequently, *we are always able to find a coordinate transform which maintains that $\mathbf{Y} = \mathbf{0}$ is the* SSM.

You may have noticed that I omit a term in (42): the term $\frac{\partial\eta_j'}{\partial X_\ell} A_{\ell,i} X_i$ should appear in the left-hand side. However, my omission is acceptable when the matrix A is upper triangular, as specified earlier, as then any term introduced which involves X_ℓ only generates extra terms which are lower order in X_ℓ. Although such extra terms increase the order of X_i for $i > \ell$, successive iterations generate new terms involving only fewer factors of X_ℓ and so iteration steadily accounts for the introduced terms. Similarly for the $\mathbf{Y}$ variables when the linear operator B is in Jordan form due to repeated eigenvalues. Discussing equation (42) for corrections is sufficient. Analogous comments apply to the the slow dynamics to which we now turn.

3.2 Transform the slow dynamics

For the slow dynamics, each iteration towards constructing a stochastic coordinate transform substitutes corrections to the transform (39) and the evolution (40–41) into the governing SDE (35) for the slow variables. Analogous to the fast dynamics, the equation for the jth component of the correction to the transform of the slow variable is the homological equation

$$F_j' + \frac{\partial\xi_j'}{\partial t} + \sum_{\ell=1}^{n} \beta_\ell Y_\ell \frac{\partial\xi_j'}{\partial Y_\ell} = \text{Res}_{35,j}\,, \tag{44}$$

where $\mathrm{Res}_{35,j}$ denotes the residual of the jth component of the SDE (35) evaluated at the current approximation. The crucial difference with the previous discussion of the fast variables is that the left-hand side of (44) does not have an analogue of the $-\beta_j\eta_j'$ term.

Consider the range of possibilities for the right-hand side. In general, the right-hand side residual $\mathrm{Res}_{35,j}$ is a sum of terms of the form $c(t)\mathbf{X}^{\mathbf{p}}\mathbf{Y}^{\mathbf{q}}$ for some vectors of integer exponents $\mathbf{p}$ and $\mathbf{q}$. Because of the special form of the 'homological' operator on the left-hand side of (44), seek corresponding corrections $F_j' = a(t)\mathbf{X}^{\mathbf{p}}\mathbf{Y}^{\mathbf{q}}$ and $\xi_j' = b(t)\mathbf{X}^{\mathbf{p}}\mathbf{Y}^{\mathbf{q}}$. Then (44) becomes

$$a + \dot{b} - \mu b = c \quad \text{where} \quad \mu = -\sum_{\ell=1}^{n} q_\ell \beta_\ell\,. \tag{45}$$

Two cases typically arise.[5]

1. The resonant case $\mu = 0$ only arises when the $\mathbf{Y}$ exponents $\mathbf{q} = \mathbf{0}$ as the exponents have to be non-negative and $\Re\beta_\ell < 0$. We need to solve $a + \dot{b} = c$ where we want to put as much into b as possible (Principle 4). Since the forcing $c(t)$ generally has mean and stochastically fluctuating components, at first sight the generic solution is $a = c$ and $b = 0$, that is, assign $c(t)\prod_{i=1}^{m} X_i^{p_i}$ to the $\mathbf{X}$ evolution and nothing into the coordinate transform.

   But recall Principle 5: we want to avoid fast time integrals in the slow evolution. Consider the case when the forcing has the form of a fast time convolution $c = e^{\nu t}\star C(t)$ for some $C(t)$. From (14) deduce

   $$c = e^{\nu t}\star C = \frac{1}{|\nu|}C + \frac{1}{\nu}\frac{d}{dt}(e^{\nu t}\star C) = \frac{1}{|\nu|}C + \frac{1}{\nu}\frac{dc}{dt}\,.$$

   Hence to avoid fast time memory integrals in the slow $\mathbf{X}$ evolution (Principle 5), set $a = C/|\nu|$ and $b = c/\nu$. If $C(t)$ in turn is a fast time convolution, then continue the above separation. This separation corresponds to the integration by parts that Section 2.2 uses to avoid fast time, memory convolutions in the slow evolution.

   When the forcing c is a quadratic product of convolutions, then similar transformations and integration by parts eliminates all memory from the slow variables except terms of the form $c^{(1)}(t)e^{\nu t}\star c^{(2)}(t)$ where $c^{(1)}$ has no convolutions. Algebraic transformations cannot eliminate

[5] The case $\Re\mu < 0$ cannot arise as all the decay rates $-\Re\beta_j > 0$ when there are no fast unstable modes.

such terms; for now accept the violation of Principle 5 in such quadratic forcing terms.

Since the case $\mu = 0$ can only arises for terms in the residual with no $\mathbf{Y}$ dependence, we maintain that the slow evolution of the $\mathbf{X}$ variables are independent of $\mathbf{Y}$, and this holds both on and off the SSM.[6]

2. The remaining case when $\Re\mu > 0$ occurs when at least one of the exponents $\mathbf{q}$ is positive. Accepting anticipation in the transform, we simply assign $b = e^{\mu t} \star c$, and do not change the $\mathbf{X}$ evolution, $a = 0$.

By anticipating noise we are *always able to find a coordinate transform which maintains a slow* $\mathbf{X}$ *evolution that is independent of whether the system is on or off the* SSM. Thus the projection of initial conditions and the exponential approach to a solution of the slow variables, called asymptotic completeness by Robinson [34], is only assured by anticipation of the noise.

The preceding arguments are phrased in the context of an iteration scheme to construct the stochastic coordinate transform and the corresponding evolution. Each step in the iterative process satisfies the governing SDEs to higher order in the asymptotic expansions. By induction, we immediately deduce the following proposition.

Proposition 1 *with stochastic anticipation allowed, a near identity stochastic coordinate transformation exists to convert the stochastic system (35–36) into the normal form*

$$\dot{\mathbf{X}} \simeq A\mathbf{X} + \mathbf{F}(\mathbf{X}, t)\,, \tag{46}$$
$$\dot{\mathbf{Y}} \simeq \big[B + G(\mathbf{X}, \mathbf{Y}, t)\big]\mathbf{Y}\,, \tag{47}$$

where $\simeq$ *denotes that these are equalities to any power of* $(\mathbf{X}, \mathbf{Y})$ *in an asymptotic expansion about the origin.*

Note: $\mathbf{F}$ *and* G *may contain fast time memory integrals but these need only occur as products with other noise processes; for example, see* (27) *and* (28).

This proposition corresponds to the general Theorem 2.1 of Arnold and Xu Kedai [5] and to the general Theorem 8.4.1(i) of Arnold [3]; the crucial difference is they do not identify that memory integrals may be mostly eliminated. Proposition 1 becomes significant when placed within the context of other theorems: Theorem 8.4.1(ii–iii) by Arnold [3] asserts that the locally

[6] However, when the fast dynamics contain rapidly oscillating, non-decaying modes, the corresponding eigenvalues occur as complex conjugate pairs which typically interact to cause $\mu = 0$; among rapid oscillation we cannot completely decouple the slow modes from the fast oscillations [12]. Physically, waves do interact with mean flow.

invariant $\mathbf{Y}=0$ is indeed attractive and approximates the SSM of the original dynamics. This, in turn, builds upon the earlier Theorem 8.3.10 [3] that there exists a stochastic coordinate transform which, by a stochastic version of Borel's lemma [3, Lemma 8.2.12], has the same formal asymptotic series as can be constructed to any order. In its turn, this builds upon the existence of well-behaved invariant measures [3, Theorem 5.6.5]. Propositions 1 and 2 empower this significant body of theory to also support macroscale modelling.

3.3 Slow dynamics do not need to anticipate the noise

Despite the presence of anticipatory convolutions appearing in the stochastic coordinate transform, we here argue that none of them appear in the slow dynamics because the anticipatory convolutions always involve fast variables. Bensoussan and Flandoli [6] correspondingly show we do not need to anticipate noise on a stochastic inertial manifold.

In the previous sections, the anticipatory convolutions only occur when the rate $\Re\mu > 0$. But for both the slow and the fast components, this occurrence is only generated when at least one fast Y_j variable appears in the term under consideration. Moreover, there is no ordinary algebraic operation that reduces the number of $\mathbf{Y}$ factors in any term: potentially the time derivative operator might,

$$\frac{d}{dt} = \frac{\partial}{\partial t} + \sum_{\ell,k} Y_k B_{\ell,k} \frac{\partial}{\partial Y_\ell} + \sum_{\ell,k} X_k A_{\ell,k} \frac{\partial}{\partial X_\ell}\,,$$

but although for non-diagonal A and B, in the algebra X_ℓ variables may be replaced by X_k and Y_ℓ variables may be replaced by Y_k, nonetheless the same number of variables are retained in each term and a $\mathbf{Y}$ variable is never replaced by an $\mathbf{X}$ variable. The reason is that the x and y dynamics are *linearly* decoupled in the system (35–36). Consequently all anticipatory convolutions appear in terms with at least one component of the fast variables $\mathbf{Y}$.

Since the evolution (46) of the slow modes $\mathbf{X}$ is free of $\mathbf{Y}$ variables, the evolution is also free of anticipatory convolutions. However, as seen in examples, there generally are anticipatory convolutions in the evolution (46) of the fast modes $\mathbf{Y}$. Further, although the stochastic coordinate transform (38) has anticipatory convolutions, on the SSM $\mathbf{Y}\simeq 0$ there are none. Consequently the preceding formal analysis leads to the following proposition.

Proposition 2 *although stochastic anticipation may be invoked, there* need not *be any anticipation in the dynamics* (46) *of the slow modes in the stochastic normal form of the system (35–36). Moreover, on the* SSM, $\mathbf{Y}\simeq 0$, *the stochastic coordinate transform* (38) *need not have anticipation.*

In contrast, Arnold, Xu Kedai and Imkeller [5, 4] record anticipatory convolutions in the slow modes of their examples, respectively (12) and (4.6). Such anticipatory convolutions are undesirable in using the normal form to support macroscale models.

4 Implications for multiscale modelling

This section describes some of the generic consequences of the previous sections in modelling stochastic systems.

Anticipation All who write down and then use coarse scale models of stochastic dynamics implicitly are soothsayers. In writing down a coarse scale model, researchers neglect the many details of any quickly decaying insignificant ignored modes. Proposition 1 assures us that normally this neglect requires us to know aspects of the near future of the ignored modes in order to decouple the coarse modes from the uninteresting details. In particular, providing initial conditions for the coarse model requires looking into the future. Nonetheless, Proposition 2 assures us that non-anticipative coarse models do exist and may be accurate for all time.

4.1 Prefer coordinate transformation over averaging

Papavasiliou and Kevrekidis [25, §5] explored the multiscale, equation free, modelling of the simple, two variable, one noise, stochastic system

$$dx = -(y+y^2)\,d\tau\,, \tag{48}$$
$$dy = -\frac{1}{\epsilon}(y-x)\,d\tau + \frac{1}{\sqrt{\epsilon}}\,dW_\tau\,. \tag{49}$$

This system has two time scales for small parameter ϵ: for small ϵ the fast variable y decays quickly to $y \approx x$ on a τ time scale $\mathcal{O}(\epsilon)$; substituting this approximate balance into (48) gives, in the absence of noise, $dx \approx -(x+x^2)d\tau$ to predict the relatively slow x evolution. We compare the information provided by averaging to that provided by stochastic normal forms in multiscale modelling.

Many apply methods of singular perturbations to systems of the form (48)–(49). For example, Papavasiliou and Kevrekidis [25] use the method of averaging to deduce that

$$x = \bar{x} + \mathcal{O}\big(\sqrt{\epsilon}\big) \quad \text{where} \quad d\bar{x} = -(\bar{x} + \bar{x}^2 + \tfrac{1}{2})\,d\tau\,. \tag{50}$$

That is, solutions of (48)–(49) are modelled to an error $\mathcal{O}(\sqrt{\epsilon})$ by the deterministic ODE (50) which applies over τ times longer than $\mathcal{O}(\epsilon)$. The noise in the fast variable y generates the extra drift $-\frac{1}{2}\,d\tau$ in (50) through the quadratic nonlinearity in the slow equation (48). However, averaging gives no basis for improving the $\mathcal{O}(\sqrt{\epsilon})$ error: such errors are often large in applications as the scale separation may only be an order of magnitude or two; for example, Papavasiliou and Kevrekidis [25] simulate SDEs (48)–(49) with scale separation $\epsilon = 0.01$ implying errors are roughly $\sqrt{\epsilon} = 10\%$. Nor does averaging recognise the stochastic fluctuations induced in the slow variable x through fluctuations in the fast variable y. *Stochastic normal forms extract both effects, and more as well.*

Computer algebra behind a web service [32] readily derives a stochastic normal form for the system (48)–(49). But first we avoid the straightjacket of singular perturbations by simply rescaling time to $t = \tau/\epsilon$: that is, we adopt a time scale t where the rapid transients decay on a t time of $\mathcal{O}(1)$, and the slow variable x evolves on long times $\Delta t \sim 1/\epsilon$. Being simply a coordinate transform, the normal form we derive is valid for *all time*: the errors only arise through the controlled truncation of the asymptotic approximations. In principle one could invoke any finite truncation of the coordinate transform and retain all generated terms in the transformed dynamical system: then the transformed system in $\mathbf{X}$ and $\mathbf{Y}$ is still exact for all time over all state space (for which the coordinate transform is not degenerate). It is only when one truncates the transformed dynamics, the SDEs in the new coordinates $\mathbf{X}$ and $\mathbf{Y}$, that approximations occur. The dynamics are formally valid for all time provided any truncation error remains small enough for your purposes. Importantly, you can control the truncation error very flexibly through any reasonable truncation of the transformed dynamics.

To proceed with this particular example, diagonalise the linear dynamics through the transform $x = x_1$ and $y = x_1 + y_1$. The example system (48)–(49) is then identical to the Stratonovich system

$$\dot{x}_1 = -\epsilon\left[x_1 + y_1 + (x_1 + y_1)^2\right], \tag{51}$$

$$\dot{y}_1 = -y_1 + \epsilon\left[x_1 + y_1 + (x_1 + y_1)^2\right] + \sigma\phi(t), \tag{52}$$

when the new noise magnitude $\sigma = 1$. I introduce the noise magnitude σ in the SDE system (51)–(52) in order to control truncation of noise effects. We seek a near identity coordinate transform $x_1 \approx X$ and $y_1 \approx Y$ to separate the slow and fast dynamics that occur in this system for small parameter ϵ. The web service [32] derives that the coordinate transform

$$x_1 = X + \epsilon(Y + \tfrac{1}{2}Y^2 + 2XY)$$

$$
\begin{aligned}
&+\epsilon\sigma\left[(Y+2X)e^{-t}\star\phi+Ye^{+t}\star\phi\right] \\
&+\tfrac{1}{2}\epsilon\sigma^2(e^{-t}\star\phi)^2+\mathcal{O}\big(\epsilon^2,\sigma^3\big)\,, \qquad (53)\\
y_1 \;=\;& Y+\sigma e^{-t}\star\phi+\epsilon\big[-Y^2+X+X^2\big] \\
&+\epsilon\sigma\big[(1+2X)e^{-t}\star e^{-t}\star\phi-2Ye^{-t}\star\phi\big] \\
&+\epsilon\sigma^2e^{-t}\star(e^{-t}\star\phi)^2+\mathcal{O}\big(\epsilon^2,\sigma^3\big)\,, \qquad (54)
\end{aligned}
$$

maps the SDE system (51)–(52) into the following Stratonovich SDE system for the new variables X and Y:

$$
\begin{aligned}
\dot X &= -\epsilon(X+X^2)-\epsilon\sigma(1+2X)\phi-\epsilon\sigma^2\phi e^{-t}\star\phi+\mathcal{O}\big(\epsilon^2,\sigma^3\big)\,, \qquad (55)\\
\dot Y &= Y\left[(-1+\epsilon)+2\epsilon X+2\epsilon\sigma\phi\right]+\mathcal{O}\big(\epsilon^2,\sigma^3\big)\,. \qquad (56)
\end{aligned}
$$

As before, the utility of the coordinate transformation is that the SDE (56) shows that the transformed fast variable $Y\to 0$ exponentially quickly from a wide range of initial conditions. Moreover, the methodology may refine the approximation, through further iteration, to suit a wide range of specified finite scale separation ϵ.

This normal form transformation reaffirms that the new slow variable X evolves independently of the fast variable Y, see (55), both throughout the initial transient as well as thereafter: there are no initial transients in X. Furthermore, being just a transform of the original SDE (52), the SDE (55) for the slow variable $X(t)$ applies for all times, albeit to the truncation error; in contrast, the averaged model (50) generally only applies for a finite time span. Also, since the coordinate transform and normal form are exact when $\epsilon=\sigma=0$, the normal form (55)–(56) is globally valid in (X,Y) and hence in (x,y). Consequently, the expansions (53)–(56) are asymptotic in only the parameters ϵ and σ, as indicated by their orders of errors. The expansions are local in the parameters ϵ and σ, but are global in X and Y.

Although not immediately apparent, the leading approximation of the slow X evolution (55) is the averaged model (50). The quadratic noise term in (55) generates a mean drift and an effective new noise over long times: Chao and I [30, 10] argued that over long times

$$
\phi e^{-t}\star\phi\mapsto \tfrac{1}{2}+\tfrac{1}{\sqrt{2}}\phi^{(1)}\,, \qquad (57)
$$

where $\phi^{(1)}(t)$ is effectively a new 'white' noise process independent of the original noise process $\phi(t)$. Thus the slow variable SDE (55) is effectively the SDE

$$
\dot X=-\epsilon(\tfrac{1}{2}\sigma^2+X+X^2)-\epsilon\sigma(1+2X)\phi-\epsilon\sigma^2\tfrac{1}{\sqrt{2}}\phi^{(1)}\,. \qquad (58)
$$

Reverting to the original (slow) time τ, setting $\sigma = 1$ to match the original noise intensity, and re-expressing, the SDE (58) becomes

$$dX = -(\tfrac{1}{2} + X + X^2)d\tau - \sqrt{\epsilon}(1 + 2X)dW_\tau - \sqrt{\tfrac{\epsilon}{2}}\, dW_\tau^{(1)} . \qquad (59)$$

The deterministic part found at leading order, $dX = -(\frac{1}{2} + X + X^2)d\tau$, is the averaged model (50). However, the SDEs (58) and (59) also make explicit some of the errors in averaging. Firstly, the $\sqrt{\epsilon}$ error of the averaged model (50) comes from its neglect of the stochastic fluctuations: to leading order we can combine noise processes W and $W^{(1)}$ to determine that the slow variables are better modelled by the SDE $dX = -(\frac{1}{2} + X + X^2)d\tau + \sqrt{3\epsilon/2}\, dW_\tau^{(2)}$; although the two stochastic terms in the SDE (59) should be better. Secondly, higher orders in the coordinate transform, not recorded here, correct errors that may be significant at finite scale separation ϵ. Simple averaging misses all of these effects.

4.2 Avoiding homogenisation of stochastic dynamics

Pavliotis, Stuart and Hairer [26, §11.6.7] develop a combination of averaging and homogenisation for the modelling of stochastic dynamics. One of their examples that requires homogenisation is the following five variable system, upon scaling time by a factor of ϵ^2, for $i = 1, 2$, and using $i' = 3 - i$:

$$\dot{x}_i = \epsilon y_i , \qquad (60)$$
$$\dot{x}_3 = \epsilon(x_1 y_2 - x_2 y_1) , \qquad (61)$$
$$\dot{y}_i = -y_i + (-1)^i \alpha y_{i'} + \phi_i(t) . \qquad (62)$$

This stochastic system has two independent white noise sources $\phi_i(t)$, and, for small parameter ϵ, has three slow variables $x_i(t)$ and two fast variables $y_i(t)$. The spiralling decay of the fast variables, combined with the form of the slow dynamics of x_3 generate subtle effects that averaging does not resolve but homogenisation does. A normal form coordinate transform proceeds without any difficulty.

Using the web service [32], which requires a diagonal form for the linear dynamics, let us explore the case of weak spiralling, that is, the regime where parameter α is small. Because of the simplicity of this system, we do not insist on small forcing and so the parameter σ does not appear. The required coordinate transform to errors $\mathcal{O}(\epsilon^2 + \alpha^2)$ is

$$\begin{aligned} y_i &\approx Y_i + e^{-t} \star \phi_i + (-1)^i \alpha e^{-t} \star e^{-t} \star \phi_{i'} , \\ x_i &\approx X_i + \epsilon\big[- Y_i - e^{-t} \star \phi_i \big] , \end{aligned}$$

$$x_3 \approx X_3 + \epsilon\left[X_2Y_1 - X_1Y_2 + X_2e^{-t}\star\phi_1 - X_1e^{-t}\star\phi_2\right].$$

In the new variables and to errors $\mathcal{O}(\epsilon^2, \alpha^2)$, the stochastic dynamics become the normal form[7]

$$\dot{Y}_i \approx -Y_i + (-1)^i\alpha Y_{i'}\,, \tag{63}$$

$$\dot{X}_i \approx \epsilon\phi_i + (-1)^i\epsilon\alpha\phi_{i'}\,, \tag{64}$$

$$\begin{aligned}\dot{X}_3 \approx{}& \epsilon(-\phi_1X_2 + \phi_2X_1) + \epsilon\alpha(\phi_1X_1 + \phi_2X_2)\\ &+ \epsilon^2(\phi_1e^{-t}\star\phi_2 - \phi_2e^{-t}\star\phi_1)\\ &+ \epsilon^2\alpha\left[\phi_1(1+e^{-t}\star)e^{-t}\star\phi_1 + \phi_2(1+e^{-t}\star)e^{-t}\star\phi_2\right].\end{aligned} \tag{65}$$

The SDE for X_3 is the one of interest. The drift in X_3 identified by Pavliotis et al. is here due to the quadratic stochastic terms $\epsilon^2\alpha\phi_ie^{-t}\star\phi_i$, which over long times are $\approx \frac{1}{2} + \frac{1}{\sqrt{2}}\phi_i^{(1)}$ [10, 30]; the other quadratic stochastic terms have no drift. Retaining the $\mathcal{O}(\epsilon^2)$ drift but neglecting the $\mathcal{O}(\epsilon^2)$ fluctuations reduces the SDE (65) to the homogenised model of Pavliotis et al. [26, (11.6.28c)]. Coordinate transforms to decouple slow and fast dynamics extend the methodology of averaging and homogenisation.

4.3 Equation free simulation

Kevrekidis et al. [21] promote a framework for computer aided, equation free, multiscale analysis, which empowers systems specified at a microscopic level of description to perform modeling tasks at a macroscopic, systems level. When the microscopic simulator is stochastic (Monte Carlo) or effectively stochastic, such as molecular and discrete element simulators, then the in principle issues addressed in this article of the nature and extraction of long term dynamics from a stochastic system are crucial to the equation free methodology.

Equation free modelling is designed to solve specific multiscale systems with specific finite scale separations. Thus a challenge for future research is to maintain reasonable accuracy in estimating long term dynamics by extracting from numerical realisations the sort of information extracted by these algebraic normal form coordinate transforms and without knowing any algebraic representations of the systems of interest. The stochastic normal form transformation shows what might be achieved in principle. The challenge is to find out how to achieve it from a finite number of short bursts of realisations.

[7] As typical for examples sourced from singular perturbation problem, this coordinate transform and normal form SDEs appear globally valid in the five dynamic variables, and locally valid in the parameters ϵ and α.

On the macroscale the stochastic effects may be relatively small. However, a deterministic macroscale model is often structurally unstable: one example is the structural instability of the averaged model $d\bar{x} = 0\,dt$ for the SDE (1); instead we prefer the stochastic model $dX = \epsilon\,dW$ of the SDE (3). Moreover, even on the macroscale a deterministic model for some averaged slow variable is almost inevitably different from the average of the system with noise included. This difference follows from the same line of argument that establishes that the expectation of realisations is generally different from the expected position of the stochastic slow manifold, see (31). Noise induced mean drift must be recognised [35, 36].

As well as the drift, the fluctuations in the macroscopic quantities usually need modelling. Thus the macroscale integration should be that of a system of SDEs. Because the macroscale SDEs model microscale processes, I conjecture that the macroscale SDEs must be interpreted as Stratonovich SDEs. The challenge is to develop Stratonovich integrators that only use short bursts of realisations.

In equation free simulation one projects a macroscopic time step into the future, then executes a burst of microscale simulation in order to estimate the macroscopic rate of change [21]. Initial rapid transients must be ignored in each burst as the microscopic system attains the quasi-equilibrium of the SSM. In a stochastic system, the true SSM can only be identified via integrals over fast time scales, see Section 2.3. However, these are generally integrals of both the past and the future. Thus, to estimate macroscopic rates of change in a stochastic system, we must not only neglect initial transients, but also data from the end of a burst of microscopic simulation in order to be able to account for the integrals which anticipate the noise processes.

Lastly, the gap-tooth scheme empowers equation free modelling across space scales as well as time scales [16, e.g.]. For spatiotemporal stochastic systems we need theoretical support for the notion that SPDEs can be modelled by the gap-tooth scheme in the same way as deterministic PDEs [33]. Only then will we be assured that we can cross space scales as well as time scales.

5 Long time modelling of stochastic oscillations

Persistent oscillations are another vitally important class of dynamics. Hopf bifurcation is the example considered in this section, but many other cases occur, including wave propagation. The challenge addressed here is how to

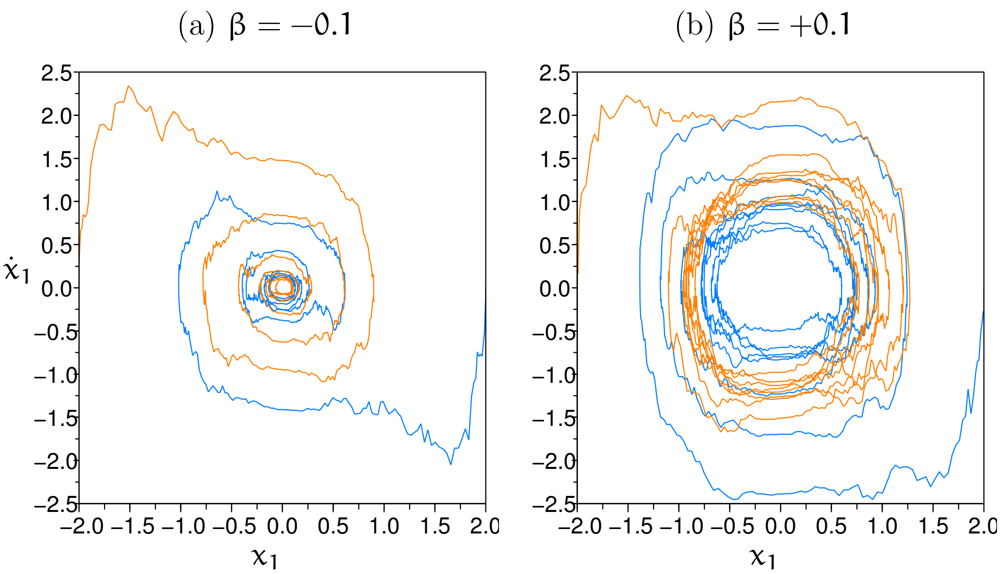

Figure 4: stochastic Hopf bifurcation in the Duffing–van der Pol oscillator (66) as parameter β crosses zero with noise amplitude $\sigma = 0.5$. Two realisations are plotted in each case.

consistently model the evolution of oscillations *over long time scales* when the oscillations are fast and in the presence of stochastic noise fluctuations over all time scales: we eliminate from the model *all* fast time dynamics.

As an example let us explore the stochastic Duffing–van der Pol dynamics also analysed by Arnold, Xu Kedai and Imkeller [5, 4]:

$$\ddot{x}_1 = (-1 + \sigma\phi(t))x_1 + \beta\dot{x}_1 - x_1^3 - x_1^2\dot{x}_1\,, \qquad (66)$$

where, as before, ϕ is some white noise process. Arnold and Xu Kedai [2] describe the importance of the stochastic system (66) in applications. In the absence of noise, $\sigma = 0$, this system exhibits a deterministic Hopf bifurcation as the parameter β crosses zero. This section explores the issues arising when constructing a normal form for the Duffing–van der Pol dynamics (66) near the stochastic Hopf bifurcation as the parameter β crosses zero with $\sigma > 0$. The aim is to explore the issues arising in systematically constructing a long time model of oscillatory dynamics.

5.1 Approaches to stochastic Hopf bifurcation

Figure 4 shows trajectories of the noisy Duffing–van der Pol oscillator (66) of Arnold and Imkeller [4] for parameter $\beta = \pm 0.1$. Figure 4 reaffirms that a noisy version of a Hopf bifurcation takes place. Coullet et al. [11] first explored a normal form of Hopf bifurcations with noise. Further research on such stochastic bifurcations elucidated some fascinating fine structure. For example, Keller and Ochs [20] explored the structure of the random 'limit cycle' attractor using a stochastic version of the subdivision algorithm of Dellnitz and Hohmann [13]; whereas Arnold and Imkeller [4] explored the structures using a normal form approach very close to that used here. However, the emphasis here is not on the stochastic Hopf bifurcation as such, but instead using it as the simplest prototype system with stochastic oscillatory dynamics. We look at the issues afresh to *explore the characteristics of a long term stochastic model* of such stochastic oscillatory dynamics. In the future, these considerations will underpin the multiscale modelling of general stochastic oscillations and waves.

Section 5.2 constructs a stochastic coordinate transform from which we may extract significant properties of a stochastic Hopf bifurcation. Solutions of the Duffing–van der Pol oscillator (66) are most conveniently represented in complex exponentials as

$$x_1 \approx a(t)e^{it} + b(t)e^{-it}\,, \qquad (67)$$

where for real solutions x_1, the amplitudes a and b are complex conjugates. Then Section 5.2 finds a stochastically forced Landau model governing the evolution of the complex amplitudes a and b:

$$\dot a \approx \tfrac{1}{2}\beta a - (\tfrac{1}{2} - \tfrac{3}{2}i)a^2b + \sigma\sqrt{\delta/2}(a\phi_0 - b\phi_{+2})\,, \qquad (68)$$

$$\dot b \approx \tfrac{1}{2}\beta b - (\tfrac{1}{2} + \tfrac{3}{2}i)ab^2 + \sigma\sqrt{\delta/2}(b\phi_0 - a\phi_{-2})\,, \qquad (69)$$

to errors $\mathcal{O}\big(\beta^2 + \sigma^2 + \epsilon^4\big)$ where here $\epsilon = |(a, b)|$ measures the size of the oscillations, and where $\phi_m(t)$ are independent 'white' noises arising from stochastic fluctuations near 'resonant' frequencies 0 and ±2 in the applied noise process $\phi(t)$; 'near' means within $\pm\delta$ of the specified frequency. This model resolves the slow evolution of the complex amplitudes near the Hopf bifurcation, small β, under the influence of the nonlinearity and a weak stochastic forcing, small σ. As the complex amplitudes a and b vary slowly in time, relative to the period of the oscillator, this model empowers long term simulations with efficient large time steps.[8]

View (67) as a time dependent coordinate transform of the $(x_1, \dot x_1)$ phase plane. In principle, any dynamics in the phase plane may be described by the evolution of the complex amplitudes a and b. The utility of the

[8]Note: the analysis also applies in the case of the deterministic forcing $\phi = \cos 2t$, for which $\phi_0 = 0$ and $\sqrt{2\delta}\phi_{\pm2} = \frac{1}{2}$. Then the above model, $\dot a \approx \frac{1}{2}\beta a + \frac{1}{4}\sigma b$ and its conjugate, successfully predicts the Mathieu-like instability with eigenvalues $\lambda = \frac{1}{2}\beta \pm \frac{1}{4}\sigma$.

coordinate transform (67) is that it empowers a simple description of oscillations with frequency near 1: namely (68)–(69) for the Duffing–van der Pol oscillator (66). Section 5.2 modifies the coordinate transform (67) through nonlinear and stochastic terms in order to simply describe nonlinear stochastic oscillations. That is, there is a time dependent, coordinate transform of the phase plane that leads to the normal form (68)–(69).

I emphasise this different view of (67). Many would view (67) as an approximation to $x(t)$ that can only resolve slowly varying oscillations. In contrast, I present (67) as the leading term in a coordinate transform, a reparametrisation, of the entire phase $(x_1, \dot{x}_1)$ plane that in principle encompasses *all* dynamics in the phase plane. The approximate model then arises by finding parameter regimes, in this new coordinate system, where the evolution of 'coordinates' a and b is slow and thus useful for long time modelling.

Amplitude/phase models do not decouple Arnold and Imkeller [4] analysed a Hopf bifurcation by transforming to real amplitude r and phase angle φ coordinates and deducing a model $\dot{r} = \cdots$ and $\dot{\varphi} = 1 + \cdots$. This approach is certainly effective for unforced deterministic problems [29, e.g.]. However, the presence of time dependent forcing, whether stochastic or deterministic, breaks time translation symmetry. Consequently, Arnold and Imkeller [4] must couple the phase φ back into the amplitude r evolution, as also seen in the normal forms of Arnold [3, p.446] and Xu Kedai [2, equation (39)]. Such coupling of the fast phase into the notionally slow amplitude confounds our aim to use the normal form for long time modelling.

Because of their different aim, Arnold and Xu Kedai [2] convert back to a pair of fast Cartesian variables to obtain a canonical system that is generic for the class of stochastic Hopf bifurcations; thus they establish that the pattern of behaviour they explore is generic for Hopf bifurcations. But our aim is different: we aim to construct models suitable for exploring long time evolution; our normal form is consequently different. We use complex amplitude coordinates, the a and b seen in (68) and (69), as originally proposed by Coullet et al. [11].

Stochastic averaging seems to suffer the same defect of not recognising the broken time symmetry [2, equations (16–20)]. Stochastic averaging also does not appear to detect the split in Lyapunov exponents present in stochastic Hopf bifurcations.

Prefer a strong model Olarrea and de la Rubia [24] comment that "When the reduction to the normal form is done ... only the deterministic part of the equations retain the characteristic radial symmetry.” and then assert “This makes it necessary to work with the two-dimensional probability distribution.” Thus they introduce early in their analysis some probability distributions governed by Fokker–Planck equations and hence derive only weak models. In contrast, here we maintain strong modelling of each realisation of the noise. We avoid weak models.

5.2 Construct a stochastic normal form

To construct the stochastic normal form for the Duffing–van der Pol oscillator (66), I use an iterative scheme to construct a useful nonlinear coordinate transform. The coordinate transform must be time dependent to adapt to both the oscillations and to the stochastic effects. The starting approximation to the linear time dependent coordinate transform is (67). Iterative modifications to (67) result in a description of the Duffing–van der Pol oscillator (66) which has only slow processes and is thus suitable for long time simulation.

The homological equation Each step in the iteration improves the normal form description of the dynamics. Suppose that at some step in the iteration, the coordinate transform and consequent evolution is $x_1 = \xi(a, b, t)$ where $\dot a = g(a, b, t)$ and $\dot b = h(a, b, t)$ for some specific functions ξ, g and h. Seek small corrections, denoted by dashes, to ξ, g and h so that

$$x_1 = \xi + \xi'(a, b, t) \quad \text{where} \quad \dot a = g + g'(a, b, t) \quad \text{and} \quad \dot b = h + h'(a, b, t) \tag{70}$$

better satisfies the Duffing–van der Pol oscillator (66). We measure how well the Duffing–van der Pol oscillator (66) is satisfied by its residual, Res_{66}. Substitute (70) into the Duffing–van der Pol oscillator (66), omit products of small corrections, approximate $\xi \approx ae^{it} + be^{-it}$ and $g \approx h \approx \beta \approx \sigma \approx 0$ whenever multiplied by a correction, and deduce that in the complex amplitude coordinates, an appropriate homological equation is

$$\xi'_{tt} + \xi' + (i2g' + g'_t)e^{it} + (-i2h' + h'_t)e^{-it} + \text{Res}_{66} = 0\,.$$

But there is one further refinement: we aim for $\dot a = g$ and $\dot b = h$ to only possess *slow* dynamics; thus, for parameter regimes where the evolution of a and b are slow, also omit the time derivatives g'_t and h'_t to give the homological equation

$$\xi'_{tt} + \xi' + i2g'e^{it} - i2h'e^{-it} + \text{Res}_{66} = 0\,. \tag{71}$$

This last simplification enables systematic algebraic construction, and is often implemented inconsistently, as commented later. The homological equation (71) governs corrections to the complex coordinate transform.

5.3 Linear noise effects

An iterative scheme to find a stochastic coordinate transform and corresponding evolution was coded into computer algebra [31]. Iterative improvements to the coordinate transform and the model continue until the residual of the Duffing–van der Pol oscillator (66) reaches a specified order of error. To effects linear in the noise magnitude σ the iteration finds the stochastic model (68) and (69) to the specified errors. In terms of the Fourier transform $\tilde\phi(\Omega)$ of the noise, $\phi(t) = \int_{-\infty}^{\infty} e^{i\Omega t}\tilde\phi(\Omega)\,d\Omega$, the corresponding stochastic complex coordinate transform is

$$\begin{aligned}
x_1 \;=\;& ae^{it} + be^{-it} + \tfrac{1}{8}\left[(1+i)a^3e^{i3t} + (1-i)b^3e^{-i3t}\right] \\
& - \sigma i a \int_D \frac{1}{\Omega(\Omega+2)} e^{i(\Omega+1)t}\tilde\phi(\Omega)\,d\Omega \\
& + \sigma i b \int_D \frac{1}{\Omega(\Omega-2)} e^{i(\Omega-1)t}\tilde\phi(\Omega)\,d\Omega \\
& + \sqrt{2\delta}\sigma\Big[\tfrac{i}{4}(a\phi_0 - b\phi_2)e^{it} - \tfrac{i}{4}(b\phi_0 - a\phi_{-2})e^{-it} \\
& \quad - \tfrac{i}{8}a\phi_2e^{i3t} + \tfrac{i}{8}b\phi_{-2}e^{-i3t}\Big] + \mathcal{O}\left(\beta^2 + \sigma^2 + \epsilon^4, \delta^{3/2}\right), \qquad (72)
\end{aligned}$$

where the integration domain D avoids singularities in the integrand as explained in Section 5.3.2.

5.3.1 Deterministic effects

The first line of (72) describes the well established deterministic shape of the limit cycle in the deterministic Hopf bifurcation. When the residual Res_{66} has terms with factors e^{imt} for some integer m, $|m| \neq 1$, and no other explicit time dependence, then as usual we update the complex coordinate transform by a correction ξ' proportional to $e^{imt}/(m^2-1)$, and do not change the evolution, $g' = h' = 0$.

Deterministic terms in the residual with factors $e^{\pm it}$, and no other explicit time dependence, such as the term $(i\beta a + 3a^2 b)e^{it}$, are resonant and as usual must be assigned to correct the evolution; see the deterministic nonlinear and β terms in the model (68)–(69).

5.3.2 Non-resonant fluctuations

The second two lines of the transform (72) describe how stochastic fluctuations non-resonantly perturb the oscillating dynamics. These arise from terms in the residual Res_{66} of the form

$$\phi(t)e^{\pm it} = \int_{-\infty}^{\infty} e^{i(\Omega\pm1)t}\tilde\phi(\Omega)\,d\Omega\,.$$

Away from resonance, namely in the domain $D = \mathbb{R}\backslash\cup_{m\in\{-2,0,2\}}[m-\delta, m+\delta]$, these terms in the residual generate the desingularised integrals in (72).[9] Rewriting these integrals as a convolution $f(t)\star\phi(t)$ recognise that formally $f = e^{\pm it}\int_D \frac{1}{\Omega(\Omega\pm2)}e^{i\Omega t}\,d\Omega$. This integral for the convolution kernel f may be written in terms of the Sine integral [1, §5.2] from which we deduce that the convolution kernel $f(t)$ decays like $1/(\delta|t|)$ for large $|t|$. Assuming that convolutions of $f(t)$ with stochastic white noise do converge in some sense, the complex transform appears to necessarily involves the entire past and future of the noise. In contrast to the pitchfork bifurcation, which only needs to look a little way into the future and the past, in the Hopf bifurcation we look far into the future and the past in order to construct the stochastic coordinate transform.

In contrast, Coullet et al. [11], in their equations (18) and (19), assign the entire integral to the evolution (68)–(69), rather than to the transformation, just because one frequency is resonant. This approach seems inconsistent in the neglect of the time derivatives g'_t and h'_t in the homological equation as such derivatives are large for 'white' noise. Their assignment to the evolution is consistent only when the noise $\phi(t)$ has a narrow band spectrum around the resonant frequencies.

5.3.3 Resonant fluctuations

The excised parts of the integrals in the transform (72) correspond to resonances. These resonances generate terms in the model (68)–(69) involving components of the (complex) noise process

$$\phi_m(t) = \frac{1}{\sqrt{2\delta}}\int_{m-\delta}^{m+\delta} e^{i(\Omega-m)t}\tilde\phi(\Omega)\,d\Omega\,, \tag{73}$$

[9] In analyses to higher order in the oscillation amplitude more resonant frequencies occur; for example, integrals arise with singularities at frequencies $\Omega = \pm4$ in some terms of $\mathcal{O}(\epsilon^2\sigma)$. In such higher order analyses the domain of integration D will have further intervals excised to avoid resonances.

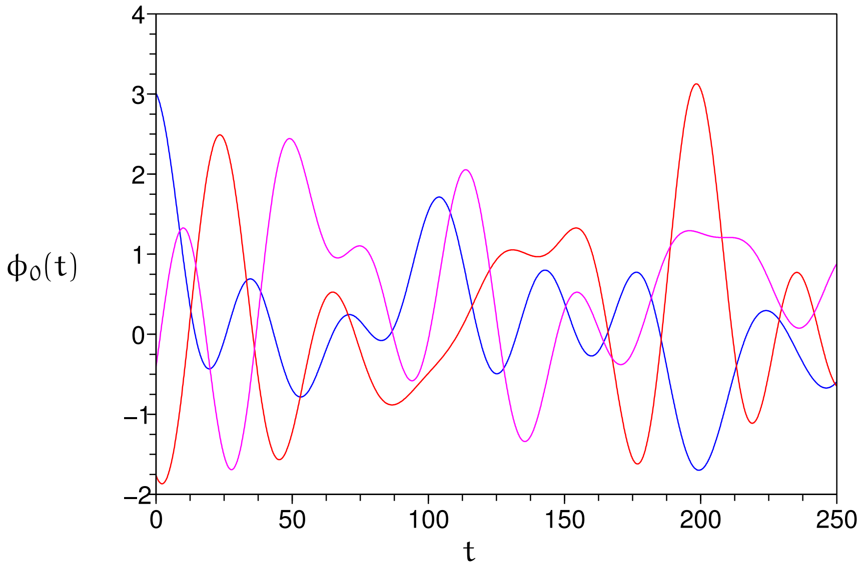

Figure 5: schematic plot of three realisations of the amplitude of resonant noise $\phi_0(t)$ for some mesoscale cutoff δ.

normalised so that $\operatorname{E}\left[|\phi_m|^2\right] = 1$ under the original white noise assumption that $\operatorname{E}\left[\tilde\phi(\Omega)^*\tilde\phi(\Omega^{(1)})\right] = \delta(\Omega - \Omega^{(1)})$ (here $\delta()$ denotes the Dirac delta function and $*$ the complex conjugate); Figure 5 plots three realisations. Being a narrow band integral (with the dominant frequency accounted for by the e^{-imt} factor) the $\phi_m(t)$ are slowly varying noise processes: Figure 5 shows $\phi_0(t)$ has slow variations on the fast times scale of the 2π-periodic oscillations. They are independent of each other as the domains of integration do not overlap (for small cutoff δ). Each $\phi_m(t)$ has autocorrelations which decay on a time scale of order $1/\delta$, roughly the width of the window in Figure 5, but for time scales $\gg 1/\delta$ the autocorrelation is zero and the ϕ_m look like white noise processes. Thus choose the 'cutoff' $1/\delta$ to be a mesoscopic time scale: one significantly longer than the period of the limit cycle; but significantly shorter than the long macroscopic time scale on which the model (68)–(69) is to be used. Then $\phi_m(t)$ are effectively independent white noise processes in the long term model.

Encouragingly, although the Fourier transform $\tilde\phi(\Omega)$ requires the entire history of the noise, the parts of it that appear in the model (68)–(69) are essentially local in time. That is, as for non oscillatory dynamics, *the long term model itself does not require anticipation of the noise.*

The fourth and fifth lines in the coordinate transform (72) arise through the excision of the resonant parts of the frequency domain from the integrals in the coordinate transform (72).

These resonant fluctuations also force the complex amplitudes a and b to change their meaning in the presence of noise. I do not precisely and explicitly define the complex amplitudes a and b; *implicitly* they are the component in $e^{\pm it}$ in the oscillations. However, whatever definition one may try to adopt, implicitly or explicitly, the noise changes the definition through the terms appearing on the fourth and fifth lines in the transform (72). Recall that in non-oscillatory systems noise also changes the presumed definition of slow variables: for two examples, the SSMs (2) and (29) show that we cannot parametrise a SSM in terms of the original slow variable x, but a new variable X which is necessarily different in the presence of noise. Similarly here: in the presence of noise, the coefficient of e^{it} in the stochastic coordinate transform is not just the complex amplitude a but instead is approximately $a + i\sigma\sqrt{2\delta}\frac{1}{4}(a\phi_0 - b\phi_2)$, and analogously for the coefficient of e^{-it}. Noise affects the meaning of the complex amplitudes.

These terms of the fourth and fifth lines in the transform (72), and the corresponding terms in the model (68)–(69), are proportional to $\sqrt{\delta}$ where δ is the small width of the domain excised from frequency space about the resonant terms. Can these terms be ignored as small? I contend it depends upon the use of the slow model (68)–(69). In a long term simulation we may use macroscopic time steps of large size Δt, say, when numerically integrating (68)–(69). In this numerical integration we would treat the $\phi_m(t)$ noises as white; thus their decorrelation time $1/\delta$ must be less than the numerical time step Δt. That is, a lower bound for the excised mesoscale cutoff is $\delta > 1/\Delta t$. Thus, a stochastic time integrator could treat these terms as $\mathcal{O}(1/\sqrt{\Delta t})$ but no smaller.

5.4 Quadratic noise effects

In many applications, quadratic noise effects generate important mean deterministic drifts [35, 36, e.g.], as seen in examples of Sections 4.1 and 4.2. Such mean drifts are important here too.

5.4.1 Double integrals of noise complicate

For oscillatory dynamics, as in the Hopf bifurcation of the Duffing–van der Pol oscillator (66), the outstanding complication is the appearance of double integrals across all frequencies in the stochastic fluctuations. Quadratic noise effects *not* involving such double integrals are straightforwardly handled as before and thus are not discussed here. Terms of $\mathcal{O}(\sigma^2)$ contain double integrals of the form $\int_D\int_D \cdot\, d\Omega\, d\bar\Omega$ where both Ω and $\bar\Omega$ represent noise

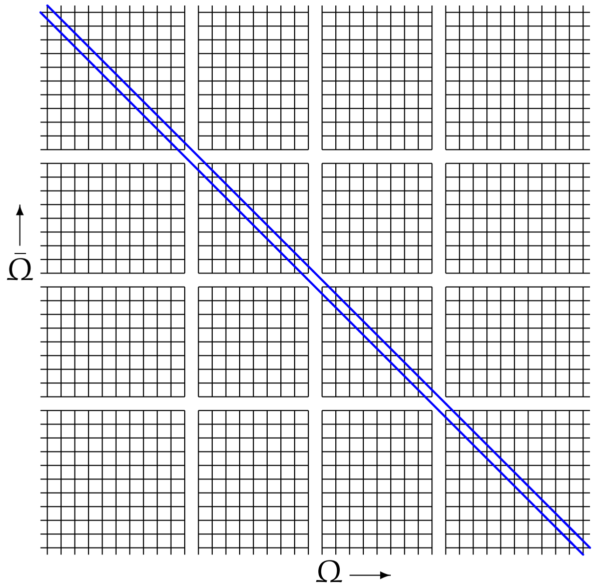

Figure 6: the integration domain $D \times D$, hatched, also has a further resonant region, the diagonal blue strip, excised to give the integration domain $\bar{D}$ for double integrals over the noise frequency.

frequencies. The (black) hatched region in Figure 6 shows this domain of integration. However, in the Hopf bifurcation of the Duffing–van der Pol oscillator (66), the kernel of such double integrals also has a singularity along the line $\Omega + \bar\Omega = 0$. Thus excise the (blue) diagonal strip shown in Figure 6 to remove the singularity to leave an integral over non-resonant effects in the domain $\bar{D}$. Then additionally analyse the excised strip as a resonant effect that directly influence the evolution of complex amplitudes a and b.

Recall we use the residual of an SDE system to drive corrections to the normal form stochastic coordinate transform. In the residual of the Duffing–van der Pol oscillator (66) quadratic noise terms arise of the form

$$\int_D \int_D e^{i(\Omega+\bar\Omega\pm 1)t} K_\pm(\Omega, \bar\Omega)\tilde\phi(\Omega)\tilde\phi(\bar\Omega)\, d\Omega\, d\bar\Omega\,,$$

where the integrand kernels are

$$K_\pm = -\frac{(\Omega + \bar\Omega \pm \Omega\bar\Omega)(\Omega + \bar\Omega \pm 2)}{2(\Omega \pm 2)(\bar\Omega \pm 2)\Omega\bar\Omega}\,. \tag{74}$$

Before excising the blue strip in Figure 6 to avoid the division by zero near $\Omega + \bar\Omega = 0$, change the parametrisation of the integration domain to $\omega = \frac{1}{2}(\Omega - \bar\Omega)$ and $\bar\omega = \Omega + \bar\Omega$ so that $\Omega = \omega + \frac{1}{2}\bar\omega$, $\bar\Omega = -\omega + \frac{1}{2}\bar\omega$, and the

Jacobian of the transform is one: parameter $\bar\omega$ measures the distance from resonance. In this new parametrisation, the integration kernels

$$\begin{aligned} K_\pm &= \frac{2(4\omega^2 \mp 4\bar\omega - \bar\omega^2)(2 \pm \bar\omega)}{(2\omega \pm 4 + \bar\omega)(2\omega \mp 4 - \bar\omega)(2\omega + \bar\omega)(2\omega - \bar\omega)} \\ &\to \frac{1}{(\omega+2)(\omega-2)} \quad \text{as } \bar\omega \to 0\,. \end{aligned} \tag{75}$$

Then the double integrals in the residual are split into non-resonant and resonant parts:

$$\begin{aligned} I_\pm &= \iint_{\bar D} e^{i(\Omega+\bar\Omega\pm1)t} K_\pm(\Omega,\bar\Omega)\tilde\phi(\Omega)\tilde\phi(\bar\Omega)\,d\Omega\,d\bar\Omega \\ &\quad + e^{\pm it}\int_{-\delta}^{\delta} e^{i\bar\omega t}\tilde\psi_\pm(\bar\omega)\,d\bar\omega\,, \end{aligned} \tag{76}$$

where

$$\tilde\psi_\pm(\bar\omega) = \int_D K_\pm\,\tilde\phi(\omega + \tfrac{\bar\omega}{2})\tilde\phi(-\omega + \tfrac{\bar\omega}{2})\,d\omega\,, \tag{77}$$

and where domain $\bar D = D \times D$ without the resonant strip as excised in Figure 6. The non-resonant double integral in the first line of (76) contributes components to the stochastic coordinate transform. The resonant integral on the second line of (76) contributes a component to the evolution in the new coordinates. Although the details will differ, the above integrals will appear in the analysis of general stochastic oscillations.

The stochastic dynamics in the normal form coordinates involves the integral (77). The integral (77) specifies the Fourier transforms of two complex conjugate components $\psi_\pm(t)$ that express a nonlinear combination of the original noise process $\phi(t)$. Here write these in terms of the real and imaginary parts

$$\psi_\pm(t) = c_r\psi_r(t) \pm ic_i\psi_i(t)\,, \tag{78}$$

where the constants c_r and c_i are chosen so the variances $E\left[\psi_r^2\right] = E\left[\psi_i^2\right] = 1$; these constants do not seem to vary significantly with mesoscale cutoff δ. Figure 7 shows one realisation of $\psi_\pm(t)$ illustrating that they vary slowly over one period of the microscale limit cycle, and that they look like white noise processes over the long time scales resolved by the complex amplitudes a and b. In the Hopf bifurcation of the Duffing–van der Pol oscillator (66) the processes $\psi_\pm(t)$ appear to have zero mean; this may not hold for other stochastic oscillations.

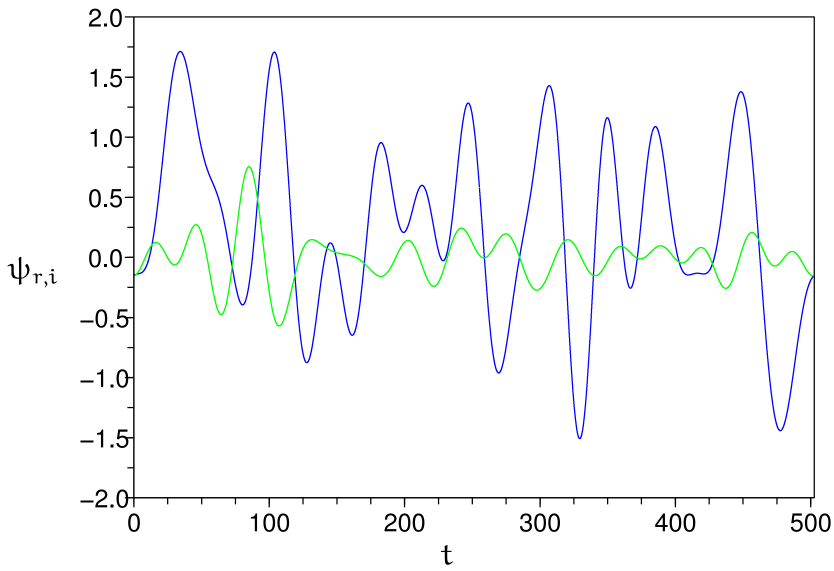

Figure 7: one realisation of the complex quadratically generated 'noise' $\psi_\pm(t) \approx 0.87\psi_r(t) \pm i0.20\psi_i(t)$ where the real part is the larger blue curve and the imaginary part is the smaller green curve. The resonant window size $\delta = 0.2$.

5.4.2 Refine the normal form transformation

Separating the double integrals as described, computer algebra [31] derives the following SDEs for the evolution of the complex amplitudes of the Duffing–van der Pol oscillator (66):

$$\begin{aligned} \dot a &\approx \tfrac{1}{2}\beta a - (\tfrac{1}{2} - \tfrac{3}{2}i)a^2 b + \sigma\sqrt{\delta/2}(a\phi_0 - b\phi_{+2}) \\ &\quad + i\tfrac{1}{2}\sigma^2(c_r\psi_r + ic_i\psi_i)a - i\delta\sigma^2(\tfrac{1}{4}\phi_0^2 + \tfrac{1}{8}\phi_2\phi_{-2})a\,, \end{aligned} \tag{79}$$

$$\begin{aligned} \dot b &\approx \tfrac{1}{2}\beta b - (\tfrac{1}{2} + \tfrac{3}{2}i)ab^2 + \sigma\sqrt{\delta/2}(b\phi_0 - a\phi_{-2}) \\ &\quad - i\tfrac{1}{2}\sigma^2(c_r\psi_r - ic_i\psi_i)b + i\delta\sigma^2(\tfrac{1}{4}\phi_0^2 + \tfrac{1}{8}\phi_2\phi_{-2})b\,. \end{aligned} \tag{80}$$

The order of error in these SDEs is $\mathcal{O}\big(\epsilon^4 + \sigma^3 + \beta^2, \delta^{3/2}\big)$. These errors include neglecting effects of: higher order resonances from frequencies $\Omega = \pm4, \pm6, \ldots$; higher order nonlinear terms in amplitudes a and b; higher order β measuring departure from the onset of oscillations; cubic and higher order noise interactions; and higher order effects of the time scale separation δ. Nonetheless, these SDEs account for more noise interactions than the lower order model (68)–(69) and thus should be more accurate.

For very small mesoscale cutoff δ, that is for simulations on very long time scales, the quadratic noise effects involving ψ_r and ψ_i are the dominant influences on the complex amplitudes a and b of the oscillations of the Duffing–van der Pol oscillator (66). These two noise processes, see the integral (77), arise as integrals of quadratic terms in the original noise process ϕ. Analogously to the quadratic noise processes analysed on stochastic slow manifolds [30, §5], as used in (57), I conjecture that ψ_r, ψ_i and ϕ are effectively independent when sampled over long time scales. Consequently, over very long time scales, one would model real dynamics of the stochastic Duffing–van der Pol oscillator (66) by the Stratonovich SDE

$$da \approx \left[\tfrac{1}{2}\beta a - (\tfrac{1}{2} - \tfrac{3}{2}i)|a|^2 a\right]dt + \tfrac{1}{2}\sigma^2 a\left(ic_r\, dW_r - c_i\, dW_i\right), \qquad (81)$$

where complex a measures the amplitude and phase of the oscillations, W_r and W_i denote independent Wiener processes, and $c_r \approx .87$ and $c_i \approx .20$ (Figure 7).

For medium mesoscale cutoff δ use the more complete SDE model (79). This model, with its effects in $\sqrt{\delta}$ and δ, will be needed when the desired time resolution of a numerical simulator, essentially the integrator's time step Δt, is within a few orders of magnitude of the natural 2π-period of oscillations. A challenge for future research is to construct special SDE numerical iteration schemes when, as here, *the* SDE *itself depends upon the chosen time step* Δt; I am only aware of SDE schemes which assume the SDE is independent of the time step [18, 22, e.g.]. Physically, the dependence upon the macroscopic time step is due to the difficulty in discerning what is and what is not a resonant forcing of the oscillations, see Sections 5.3.2–5.3.3. In multiscale modelling, as shown here, the macroscopic system, whether expressed as algebraic equations or solved using equation free methods [21], may depend upon the the length or time scale chosen for simulation.

The specific equations and formulae in the section are specific to the Duffing–van der Pol equation (66). Nonetheless, I contend that the nonlinear and stochastic nature of these Duffing–van der Pol oscillations are generic for most of the interesting issues discussed in this section. Consequently, I conjecture that almost all long time scale modelling of stochastic oscillations has to address and resolve the issues discussed in this section.

6 Conclusion

Stochastic coordinate transforms illuminate modelling of multiscale stochastic systems. Being a coordinate transform, a resultant 'stochastic normal form' describes the complete dynamics of the original system, Proposition 1. From the normal form we easily extract the stochastic slow dynamics that are of interest over macroscopic times, from the uninteresting fast dynamics

[3, §8.4, e.g.]. This approach is more powerful than averaging and homogenisation as the coordinate transform may be systematically refined, especially with the aid of computer algebra [31, 32], and so errors are more controlled.

In contrast to earlier work, this article argues that two modelling simplifications may always be achieved without sacrificing fidelity with the original stochastic system. Firstly, the stochastic slow manifold and the evolution thereon need not have any terms anticipating the original noise processes, Proposition 2. Secondly, effects linear in the noise processes in the evolution on the stochastic slow manifold need not involve any memory integrals either, Proposition 1. Section 2 explores the application of these principles for the example SDE system (4).

A challenge for future research is to let the algebraic techniques used herein inspire development of numerical techniques useful for multiscale computations. From a finite number of bursts of stochastic realisations we need to determine information to empower making macroscale time steps while remaining faithful to the underlying stochastic dynamics.

Section 5 explored oscillatory dynamics in the stochastic Duffing–van der Pol equation (66). It demonstrates that transforming the SDE to a slow model for the complex amplitude is a delicate process that requires careful treatment of noise integrals in order to form a consistent model of the long term evolution. The specific and formal analysis herein needs to be extended to generic oscillatory systems to discover general modelling principles.

Acknowledgements: I thank Phil Pollett and Yannis Kevrekidis for useful discussions on aspects of this article. This project receives support from the Australian Research Council.

References

[1] M. Abramowitz and I. A. Stegun, editors. *Handbook of mathematical functions*. Dover, 1965. 36

[2] L. Arnold, N. Sri Namachchivaya, and K. R. Schenk-Hoppé. Toward an understanding of stochastic Hopf bifurcation: a case study. *Intl. J. Bifurcation & Chaos*, 6:1947–1975, 1996. 31, 33

[3] Ludwig Arnold. *Random Dynamical Systems*. Springer Monographs in Mathematics. Springer, June 2003. 3, 4, 6, 12, 13, 18, 23, 24, 33, 43

[4] Ludwig Arnold and Peter Imkeller. Normal forms for stochastic differential equations. *Probab. Theory Relat. Fields*, 110:559–588, 1998. doi:10.1007/s004400050159. 3, 4, 6, 13, 25, 31, 32, 33

[5] Ludwig Arnold and Xu Kedai. Simultaneous normal form and center manifold reduction for random differential equations. In C. Perello, C. Simo, and J. Sola-Morales, editors, *Equadiff-91*, pages 68–80, 1993. 23, 25, 31

[6] A. Bensoussan and F. Flandoli. Stochastic inertial manifold. *Stochastics and Stochastics Rep.*, 53:13–39, 1995. 4, 24

[7] Nils Berglund and Barbara Gentz. Geometric singular perturbation theory for stochastic differential equations. *J. Diff. Equations*, 191:1–54, 2003. doi:10.1016/S0022-0396(03)00020-2. 18

[8] P. Boxler. A stochastic version of the centre manifold theorem. *Probab. Th. Rel. Fields*, 83:509–545, 1989. 18

[9] J. Carr. *Applications of centre manifold theory*, volume 35 of *Applied Math. Sci.* Springer–Verlag, 1981. 19

[10] Xu Chao and A. J. Roberts. On the low-dimensional modelling of Stratonovich stochastic differential equations. *Physica A*, 225:62–80, 1996. doi:10.1016/0378-4371(95)00387-8. 3, 14, 27, 29

[11] P. H. Coullet, C. Elphick, and E. Tirapegui. Normal form of a Hopf bifurcation with noise. *Physics Letts*, 111A(6):277–282, 1985. 4, 31, 33, 36

[12] S. M. Cox and A. J. Roberts. Initial conditions for models of dynamical systems. *Physica D*, 85:126–141, 1995. doi:10.1016/0167-2789(94)00201-Z. 2, 7, 16, 23

[13] M. Dellnitz and A. Hohmann. A subdivision algorithm for the computation of unstable manifolds and global attractors. *Numerische Mathematik*, 75:293–317, 1997. doi:10.1007/s002110050240. 32

[14] Aijun Du and Jinqiao Duan. Invariant manifold reduction for stochastic dynamical systems. Technical report, `http://arXiv.org/abs/math.DS/0607366`, 2006. 4

[15] C. Elphick, E. Tirapeugi, M. E. Brachet, P. Coullet, and G. Iooss. A simple global characterisation for normal forms of singular vector fields. *Physica D*, 29:95–127, 1987. doi:10.1016/0167-2789(87)90049-2. 2, 7

[16] C. W. Gear, Ju Li, and I. G. Kevrekidis. The gap-tooth method in particle simulations. *Phys. Lett. A*, 316:190–195, 2003. doi:10.1016/j.physleta.2003.07.004. 30

[17] Dror Givon, Ioannis G. Kevrekidis, and Raz Kupferman. Strong convergence of projective integration schemes for singularly perturbed stochastic differential systems. Technical report, Department of Mathematics, LBNL, Berkeley CA 94720, 2006. 2

[18] Desmond J. Higham and Peter E. Kloeden. Numerical methods for nonlinear stochastic differential equations with jumps. *Numerische Mathematik*, 101:101–119, 2005. doi:10.1007/s00211-005-0611-8. 42

[19] Wolfram Just, Holger Kantz, Christian Rodenbeck, and Mario Helm. Stochastic modelling: replacing fast degrees of freedom by noise. *J. Phys. A: Math. Gen.*, 34:3199–3213, 2001. doi:10.1088/0305-4470/34/15/302. 9

[20] Hannes Keller and Gunter Ochs. Numerical approximation of random attractors. In H. Crauel and M. Gundlach, editors, *Stochastic dynamics*. Springer, 1999. `http://www.math.uni-bremen.de/Math-Net/preprints/ohne-metadaten/Report431.ps.gz`. 32

[21] I. G. Kevrekidis, C. W. Gear, J. M. Hyman, P. G. Kevrekidis, O. Runborg, and K. Theodoropoulos. Equation-free, coarse-grained multiscale computation: enabling microscopic simulators to perform system level tasks. *Comm. Math. Sciences*, 1:715–762, 2003. 29, 30, 42

[22] P. E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*, volume 23 of *Applications of Mathematics*. Springer–Verlag, 1992. 42

[23] James Murdock. *Normal forms and unfoldings for local dynamical systems*. Springer Monographs in Mathematics. Springer, 2003. 3

[24] José Olarrea and F. Javier de la Rubia. Stochastic Hopf bifurcation: the effect of colored noise on the bifurcation interval. *Phys. Rev. E*, 53:268–271, 1996. 33

[25] Anastasia Papavasiliou and Ioannis G. Kevrekidis. Variance reduction for the equation-free simulation of multiscale stochastic systems. Technical report, Department of Statistics, University of Warwick, Coventry, CV4 7AL, UK., June 2006. 25, 26

[26] G. A. Pavliotis and A. M. Stuart. Multiscale methods: averaging and homogenization. Technical report, http://www.ma.ic.ac.uk/~pavl, 2007. 2, 28, 29

[27] David W. Pierce. Distinguishing coupled ocean-atmosphere interactions from background noise in the North Pacific. Technical report, http://meteora.ucsd.edu/~pierce/docs/Pierce_2001_Prog_Oceanog.pdf, 2001. 5

[28] A. J. Roberts. Low-dimensional modelling of dynamics via computer algebra. *Computer Phys. Comm.*, 100:215–230, 1997. 9

[29] A. J. Roberts. Create useful low-dimensional models of complex dynamical systems. Technical report, http://www.sci.usq.edu.au/staff/robertsa/Modelling/, 2006. 33

[30] A. J. Roberts. Resolving the multitude of microscale interactions accurately models stochastic partial differential equations. *LMS J. Computation and Maths*, 9:193–221, 2006. http://www.lms.ac.uk/jcm/9/lms2005-032. 3, 4, 27, 29, 42

[31] A. J. Roberts. Computer algebra derives normal forms of stochastic differential equations. Technical report, http://eprints.usq.edu.au/archive/00001873, 2007. 35, 41, 43

[32] A. J. Roberts. Normal form of stochastic or deterministic multiscale differential equations. Technical report, http://www.sci.usq.edu.au/staff/robertsa/sdenf.html, 2007. 5, 14, 26, 28, 43

[33] A. J. Roberts and I. G. Kevrekidis. General tooth boundary conditions for equation free modelling. *SIAM J. Scientific Computing*, 29(4):1495–1510, 2007. 30

[34] J. C. Robinson. The asymptotic completeness of inertial manifolds. *Nonlinearity*, 9:1325–1340, 1996. http://www.iop.org/EJ/abstract/0951-7715/9/5/013. 23

[35] N. Sri Namachchivaya and G. Leng. Equivalence of stochastic averaging and stochastic normal forms. *J. Appl. Mech.*, 57:1011–1017, 1990. 5, 14, 30, 38

[36] N. Sri Namachchivaya and Y. K. Lin. Method of stochastic normal forms. *Int. J. Nonlinear Mechanics*, 26:931–943, 1991. 5, 14, 30, 38

[37] R. Temam. Inertial manifolds. *Mathematical Intelligencer*, 12:68–74, 1990. 4

[38] Ferdinand Verhulst. *Methods and applications of singular perturbations: boundary layers and multiple timescales*, volume 50 of *Texts in Applied Maths*. Springer, 2005. 2