INSTITUTE OF PHYSICS PUBLISHING NONLINEARITY

Nonlinearity **17** (2004) R55–R127 PII: S0951-7715(04)34115-0

INVITED ARTICLE

Extracting macroscopic dynamics: model problems and algorithms

Dror Givon[1], **Raz Kupferman**[1] **and Andrew Stuart**[2]

[1] Institute of Mathematics, The Hebrew University, Jerusalem, 91904 Israel
[2] Mathematics Institute, Warwick University, Coventry, CV4 7AL, UK

E-mail: givon@math.huji.ac.il, raz@math.huji.ac.il and stuart@maths.warwick.ac.uk

Received 5 June 2003, in final form 7 April 2004
Published 20 August 2004
Online at stacks.iop.org/Non/17/R55
doi:10.1088/0951-7715/17/6/R01

Recommended by R Krasny

Abstract

In many applications, the primary objective of numerical simulation of time-evolving systems is the prediction of coarse-grained, or macroscopic, quantities. The purpose of this review is twofold: first, to describe a number of simple model systems where the coarse-grained or macroscopic behaviour of a system can be explicitly determined from the full, or microscopic, description; and second, to overview some of the emerging algorithmic approaches that have been introduced to extract effective, lower-dimensional, macroscopic dynamics.

The model problems we describe may be either stochastic or deterministic in both their microscopic and macroscopic behaviour, leading to four possibilities in the transition from microscopic to macroscopic descriptions. Model problems are given which illustrate all four situations, and mathematical tools for their study are introduced. These model problems are useful in the evaluation of algorithms. We use specific instances of the model problems to illustrate these algorithms. As the subject of algorithm development and analysis is, in many cases, in its infancy, the primary purpose here is to attempt to unify some of the emerging ideas so that individuals new to the field have a structured access to the literature. Furthermore, by discussing the algorithms in the context of the model problems, a platform for understanding existing algorithms and developing new ones is built.

PACS numbers: 05.45.−a, 05.10.−a

(Some figures in this article are in colour only in the electronic version)

Contents

1. Set-up

The general problem may be described as follows: let $\mathcal{Z}$ be a Hilbert space, and consider the noise-driven differential equation for $z \in \mathcal{Z}$:

$$\frac{\mathrm{d}z}{\mathrm{d}t} = h(z) + \gamma(z)\frac{\mathrm{d}W}{\mathrm{d}t}, \tag{1.1}$$

where $W(t)$ is a noise process, chosen so that $z(t)$ is Markovian. We will focus mainly on the case where $W(t)$ is a standard multivariate Brownian motion and (1.1) is an Itô stochastic differential equation (SDE). In addition, we will also touch on the case where $\gamma(z)\mathrm{d}W(t)/\mathrm{d}t$ is replaced by a Poisson counting process $\mathrm{d}W(z, t)/\mathrm{d}t$, inducing jumps in z, whose magnitude depend upon the current state. The problem (1.1) also reduces to an ordinary differential equation (ODE) if $\gamma \equiv 0$; this situation will be of interest to us in some cases too.

This overview is focused on situations where the dynamics of interest for (1.1) takes place in a subspace $\mathcal{X} \subset \mathcal{Z}$ and our objective is to find a self-contained description of this dynamics, without fully resolving the dynamics in $\mathcal{Z}\backslash\mathcal{X}$. In particular we are interested in cases where $\mathcal{Z}$ has large (perhaps infinite) dimension and the dimension of $\mathcal{X}$ is small (finite). Anticipating this, we introduce the projection $P : \mathcal{Z} \mapsto \mathcal{X}$ and the orthogonal complement of $\mathcal{X}$ in $\mathcal{Z}$, $\mathcal{Y} = (I - P)\mathcal{Z}$.

Employing coordinates x in $\mathcal{X}$ and y in $\mathcal{Y}$ we obtain from (1.1) the coupled SDEs

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= f(x, y) + \alpha(x, y)\frac{\mathrm{d}U}{\mathrm{d}t}, \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= g(x, y) + \beta(x, y)\frac{\mathrm{d}V}{\mathrm{d}t}, \end{aligned} \tag{1.2}$$

where U, V are again noise processes.

We will study situations where the y variables can be eliminated, and an approximate, effective equation for x can be derived. We will refer to the equations for $z = (x, y)$ as the full, or microscopic, description; and we will refer to the effective equations for x as the coarse-grained, or macroscopic, description.

In many cases we will be looking for an Itô stochastic differential equation for $X \in \mathcal{X}$:

$$\frac{\mathrm{d}X}{\mathrm{d}t} = F(X) + A(X)\frac{\mathrm{d}U'}{\mathrm{d}t}, \tag{1.3}$$

where $X(t)$ approximates $x(t)$, in a sense that is to be determined for each class of problems, and U' is a noise process. In other cases, the reduced dynamics require the introduction of additional auxiliary variables, so that the approximate solution $X(t)$ is a component of a problem which evolves in a space of dimension higher than the dimension of $\mathcal{X}$, but still smaller than the dimension of $\mathcal{Z}$. We consider cases where the original model (1.1) for z is either an autonomous ODE or a noise-driven differential equation, such as an SDE, and where the effective dynamics (1.3) for X is either an ODE or an SDE. The ideas we describe have

discrete time analogues, and some of the algorithms we overview extract a low-dimensional discrete time model, such as a Markov chain, rather than a continuous time model. We will also examine situations where the effective dimension reduction can be carried out in the space of probability densities propagated by the paths of (1.1); this requires taking into consideration the master equation for probability densities.

A primary motivation for this paper is to overview the wealth of recent work concerning algorithms which attempt to find the effective dynamics in $\mathcal{X}$. This work is, at present, not very unified and our aim is to highlight the similarities and differences among currently emerging approaches. It is important to realize, however, that there is no single algorithmic approach which will solve all problems, and this explains the range of algorithmic ideas overviewed here. In the long term these methods need to be compared and hence evaluated by their ability to solve particular classes of problems. It is too early to undertake such evaluations for many of the methods overviewed here, and hence we do not attempt this.

Another primary motivation for this review is to collect together examples and mathematical intuition about situations in which it is possible to find closed equations for X that adequately approximate the dynamics of $x \in \mathcal{X}$. Thus, much of the paper will be devoted to the development of model problems, and the underlying theoretical context in which they lie. Model problems are of central importance in order to make clear statements about the situations in which we expect the given algorithms to be of use, and in order to develop examples which can be used to test these algorithms. We do not state theorems nor give proofs—we present the essential ideas and refer to the literature for precise statement and rigorous analysis.

Section 2 contains an introduction to the master equation, first for countable state space Markov chains. On uncountable state spaces, and for $W(t)$ representing Brownian motion in (1.1), the master equation is a partial differential equation (PDE)—the Fokker–Planck equation—and its adjoint—the Chapman–Kolmogorov equation—propagates expectations; we describe these PDEs. In section 3, we outline the Mori–Zwanzig projection operator approach which describes the elimination of variables in a general setting. Sections 4–8 describe a variety of situations where an effective equation for the dynamics in $\mathcal{X}$ can be derived. Each section contains a discussion of the theoretical development of the subject, statement of a class (or classes) of model problems, together with one or more explicit examples. Section 9 describes a reduction principle somewhat different from those outlined above: it is concerned with the derivation of a small and finite state Markov chain from dynamics such as (1.3). It is hence an additional layer of variable elimination. Such situations arise frequently in applications, and for this reason the section is included.

A starting point for the evaluation of the algorithms in this paper would be to carefully compare their behaviour when applied to the model problems described here, and to more challenging problems with similar character drawn from the sciences and engineering. The discussion in this paper is neither comprehensive, nor do we claim to make any evaluation of the relative merits of the algorithms described. Sections 10–13 are devoted to a description of a variety of algorithms recently developed, or currently under development, which aim to find effective dynamics in $\mathcal{X}$, given the full evolution equation (1.1) in $\mathcal{Z}$. In all cases we overview the algorithms used, describe what is known about them analytically and through numerical experience, and show how the model problems of sections 4–8 are relevant to the evaluation of the algorithms used.

The following provides an overview of a number of important themes running throughout this paper.

(i) *Classification of model problems.* It is useful to classify the model problems according to whether or not the dynamics in $\mathcal{Z}$ and $\mathcal{X}$ are deterministic or stochastic. The situations

outlined in sections 4–8 are of the form D–D, D–D, S–S, D–S and S–D, in sequence, where D denotes *deterministic*, and S denotes *stochastic*, and the first (respectively second) letter defines the type of dynamics in $\mathcal{Z}$ (respectively $\mathcal{X}$). The Markov chain extraction in section 9 is of the form S–S.

(ii) *Scale-separation.* Sections 4–6 all rely on explicit time-scale separation to achieve the memoryless effect; in a different way, the work of section 9 also relies on a scale-separation. In contrast the examples in section 7 rely on the high dimensionality of $\mathcal{Y}$ relative to $\mathcal{X}$; the mean time-scale in $\mathcal{Y}$ then separates from that in $\mathcal{X}$, but there is no pure separation of scales. Thus, whilst scale-separation is a useful concept which unifies many of the underlying theoretical models in this subject area, the details of how one establishes rigorously a given dimension reduction differ substantially depending on whether there is a clear separation of scales, or instead a separation in a mean sense.

(iii) *Reduction principles.* In conceptualizing these algorithms it is important to appreciate that any algorithm aimed at extracting dynamics in $\mathcal{X}$, given the equations of motion (1.1) in $\mathcal{Z}$, has two essential components: (i) determining the projection P which defines $\mathcal{X}$ through $\mathcal{X} = P\mathcal{Z}$; (ii) determining the effective dynamics in $\mathcal{X}$. In some instances P is known *a priori* from the form of model (1.1) and/or from physical considerations; in others its determination may be the most taxing part of the algorithm. In all the model problems developed in sections 4–8 the definition of $\mathcal{X}$ is explicit in the problem statement. Section 9 is somewhat different as it identifies collective variables, within $\mathcal{X}$, which can be modelled by a finite state Markov chain. Section 13 describes algorithms designed to identify $\mathcal{X}$ as well as the dynamics within it; in contrast, the algorithms described in sections 10–12 all assume that $\mathcal{X}$ is known.

(iv) *Memory.* An important aim of any such algorithm is to choose P in such a way that the dynamics in $\mathcal{X}$ is memoryless. In principle, y can always be eliminated from (1.2) to obtain an equation for x alone but, in general, this equation will involve the past history of x; this is the idea of the Mori–Zwanzig formalism. In order to understand and improve algorithms it is therefore important to build up intuition about situations in which memory in $\mathcal{X}$ disappears or, alternatively, in which it can be modelled by a few degrees of freedom. Sections 4–8 are all devoted to situations where the effect of memory disappears completely, except section 7 which includes the description of situations where a memory effect remains, but can be modelled by adding a small number of extra degrees of freedom.

Although we describe this subject primarily through simple model problems in finite dimensions, the majority of applications are to infinite-dimensional problems; furthermore, the applied context provides an important component for the shaping of relevant questions in the area of extracting macroscopic behaviour. The paper concludes in section 14 with some brief comments about the literature in the computational PDE and various applied communities.

2. The master equation

In this section we describe dynamical systems of the form (1.1) within a probabilistic setting, by considering the evolution of probability measures induced by the dynamics of the paths of (1.1). There are several reasons for including this section. First, it provides a way of understanding the Fokker–Planck equation for SDEs, building on the pedagogically straightforward case of the master equation for countable state space Markov chains; the Fokker–Planck equation, and its adjoint the Chapman–Kolmogorov equation, play a central role in this paper. Second, for the birth–death processes described in section 8, the master equation is used to derive

a Liouville equation for the effective, non-stochastic, behaviour. Third, many of the basic theoretical ideas in this paper can be described simply by analogy with particular structures arising in the master equation for a finite or countable Markov chain.

There are two primary benefits which follow from considering a probabilistic description rather than a pathwise one, even when (1.1) is deterministic ($\gamma = 0$):

(i) Variable elimination is often related to uncertainty in initial data, hence to ensembles of solutions. The reduced initial data $x(0) = Pz(0)$ are *a priori* compatible with a large set of initial data $z(0)$ for the full evolution equation. Every initial datum $z(0)$ gives rise to a different solution $z(t)$ and to different projected dynamics $x(t)$. In many cases it is meaningless to consider how $x(0)$ evolves into $x(t)$ without specifying how the eliminated variables, $y(0)$, are initially distributed.

(ii) The evolution of the measure is governed by a *linear* PDE. In spite of the increased complexity owing to the infinite-dimensionality of the system, the linearity enables the use of numerous techniques adapted for linear systems, such as projection methods, and perturbation expansions.

A useful example illustrating the first point comes from statistical mechanics. It is natural to specify the temperature of (some components of) a molecular system, since it is a measurable macroscopic quantity, without specifying the exact position and velocity of every particle; this corresponds to specifying a probability measure on the positions and velocities, with the variance being determined by the temperature. A useful example illustrating the second point is passive tracer advection: the position of a particle advected in a velocity field and subject to molecular diffusion, can then be modelled by a nonlinear SDE; collections of such particles have a density satisfying a linear advection–diffusion equation. In the absence of noise this simply reflects the fact that the method of characteristics for a linear hyperbolic problem gives rise to nonlinear ODEs.

In section 2.1 we describe the derivation of the equation governing probability measures for countable state space Markov chains; in section 2.2 we generalize this to the case of Itô SDEs, which give rise to Markov processes on uncountable state spaces. Section 2.3 looks ahead to sections 4–9, motivating what is done in those sections by using the master equation and its variants.

2.1. Countable state space

Consider a continuous time Markov chain $z(t)$, $t \geqslant 0$, taking values in the state space $\mathcal{I} \subseteq \{0, 1, 2, \ldots\}$. Let $p_{ij}(t)$ be the transition probability from state i to j:

$$p_{ij}(t) = \mathbb{P}\{z(t) = j \mid z(0) = i\},$$

i.e. the probability that the process is in state j at time t, given that it was in state i at time zero. The Markov property implies that for all t, $\Delta t \geqslant 0$,

$$p_{ij}(t + \Delta t) = \sum_k p_{ik}(t) p_{kj}(\Delta t),$$

and so

$$\frac{p_{ij}(t + \Delta t) - p_{ij}(t)}{\Delta t} = \sum_k p_{ik}(t) \ell_{kj}(\Delta t),$$

where

$$\ell_{kj}(\Delta t) = \frac{1}{\Delta t} \times \begin{cases} p_{kj}(\Delta t) & k \neq j, \\ p_{jj}(\Delta t) - 1 & k = j. \end{cases}$$

Suppose that the limit $\ell_{kj} = \lim_{\Delta t \to 0} \ell_{kj}(\Delta t)$ exists. We then obtain, formally,

$$\frac{\mathrm{d}p_{ij}}{\mathrm{d}t} = \sum_k p_{ik}\ell_{kj}. \tag{2.1}$$

Because $\sum_j p_{ij}(\Delta t) = 1$ it follows that $\sum_j \ell_{ij}(\Delta t) = 0$, and we expect that

$$\sum_j \ell_{ij} = 0.$$

This implies that

$$\sum_j p_{ij} = 1$$

for the limit equation (2.1).

Introducing the matrices P, L with entries p_{ij}, ℓ_{ij}, respectively, $i, j \in \mathcal{I}$, equation (2.1) reads, in matrix notation,

$$\frac{\mathrm{d}P}{\mathrm{d}t} = PL, \qquad P(0) = I. \tag{2.2}$$

The matrix L is known as the *generator* of the process. Since $P(t) = \exp(Lt)$ solves this problem we see that P and L commute so that $P(t)$ also solves

$$\frac{\mathrm{d}P}{\mathrm{d}t} = LP, \qquad P(0) = I. \tag{2.3}$$

We refer to (2.2) and (2.3) as the *forward* and *backward* equations of the Markov chain. Equation (2.2) is also called the *master equation*.

Let $\mu(t) = (\mu_0(t), \mu_1(t), \ldots)^T$ be the ith row of $P(t)$, i.e. a column vector whose entries $\mu_j(t) = p_{ij}(t)$ are the probabilities that a system starting in state i will end up, at time t, in each of the states $j \in \mathcal{I}$. By virtue of (2.2),

$$\frac{\mathrm{d}\mu}{\mathrm{d}t} = L^T\mu, \qquad \mu(0) = e_i, \tag{2.4}$$

where e_i is the ith unit vector, zero in all entries except the ith one, in which it is one; this initial condition indicates that the chain is in state i at time $t = 0$. Equation (2.4) is the discrete version of the Fokker–Planck equation described below.

Let $w : \mathcal{I} \mapsto \mathbb{R}$ be a real valued function defined on the state space; it can be represented as a vector with entries w_j, $j \in \mathcal{I}$. Then let $v(t) = (v_0(t), v_1(t), \ldots)^T$ denote the vector with ith entry

$$v_i(t) = \mathbb{E}\{w_{z(t)} \mid z(0) = i\},$$

where $\mathbb{E}$ denotes expectation with respect to the Markov transition probabilities. The function $v_i(t)$ denotes the expectation value at time t of a function of the state space (an 'observable'), given that the process started in the ith state. This function can be written explicitly in terms of the transition probabilities:

$$v_i(t) = \sum_j p_{ij}(t)w_j. \tag{2.5}$$

If we set $w = (w_0, w_1, \ldots)^T$ then this can be written in vector form as $v(t) = P(t)w$. Differentiating with respect to time and using the backward equation (2.3), $v(t)$ satisfies the following system of ODEs:

$$\frac{\mathrm{d}v}{\mathrm{d}t} = Lv, \qquad v(0) = w. \tag{2.6}$$

Equation (2.6) is the discrete version of the Chapman–Kolmogorov equation described in section 2.2.

2.2. Fokker–Planck and Chapman–Kolmogorov equations

The concepts introduced in section 2.1 are now extended to continuous time Markov processes over uncountable state spaces, specifically, to diffusion processes defined by SDEs. Consider the case where $W(t)$ is a multi-dimensional Brownian motion and (1.1) is an Itô SDE. We assume that $\mathcal{Z}$ has dimension d and let ∇ and $\nabla\cdot$ denote gradient and divergence in $\mathbb{R}^d$. The gradient can act on both scalar valued functions ϕ, or vector valued functions v, as follows:

$$(\nabla\phi)_i = \frac{\partial\phi}{\partial z_i}, \qquad (\nabla v)_{ij} = \frac{\partial v_i}{\partial z_j}.$$

The divergence acts on vector valued functions v, or matrix valued functions A as follows:

$$\nabla\cdot v = \frac{\partial v_i}{\partial z_i}, \qquad (\nabla\cdot A)_i = \frac{\partial A_{ij}}{\partial z_j}.$$

In the preceding we are adopting the Einstein summation convention, whereby repeated indexes imply a summation. Below we will use ∇_x (respectively ∇_y) to denote gradient or divergence with respect to x (respectively y) coordinates alone.

With the functions $h(z)$,$\gamma(z)$ given in the SDE (1.1) we define

$$\Gamma(z) = \gamma(z)\gamma(z)^T,$$

and then the *generator* $\mathcal{L}$ by

$$\mathcal{L}\phi = h\cdot\nabla\phi + \tfrac{1}{2}\Gamma : \nabla(\nabla\phi), \tag{2.7}$$

where $\cdot$ denotes the standard inner product on $\mathbb{R}^d$, and $:$ denotes the inner product on $\mathbb{R}^{d\times d}$ which induces the Frobenius norm—$A : B = \mathrm{trace}(A^T B)$. We will also be interested in the operator $\mathcal{L}^*$ defined by

$$\mathcal{L}^*\rho = -\nabla\cdot(h\rho) + \tfrac{1}{2}\nabla\cdot[\nabla\cdot(\Gamma\rho)],$$

which is the adjoint of $\mathcal{L}$ with respect to the scalar product

$$\langle\phi,\rho\rangle = \int_{\mathcal{Z}} \phi(z)\rho(z)\,\mathrm{d}z,$$

i.e. $\langle\mathcal{L}\phi,\rho\rangle = \langle\phi,\mathcal{L}^*\rho\rangle$.

If we consider solutions of (1.1) with initial data distributed according to a measure with density $\rho_0(z)$ then, at time $t > 0$, $z(t)$ is distributed according to a measure with density $\rho(z,t)$ satisfying the Fokker–Planck equation

$$\begin{aligned} &\frac{\partial\rho}{\partial t} = \mathcal{L}^*\rho \qquad (z,t)\in\mathbb{R}^d\times(0,\infty),\\ &\rho = \rho_0 \qquad (z,t)\in\mathbb{R}^d\times\{0\}. \end{aligned} \tag{2.8}$$

This is the analogue of the master equation (2.4) in the countable state space case. We are implicitly assuming that the measure μ_t, defined by $\mu_t(A) = \mathbb{P}\{z(t)\in A\}$, has density $\rho(z,t)$ with respect to Lebesgue measure. Here $\mathbb{P}$ is the probability measure on paths of Brownian motion (Wiener measure), and we denote by $\mathbb{E}$ the expectation with respect to this measure. Whether or not the smooth density ρ exists depends on the (hypo-) ellipticity properties of $\mathcal{L}$. In the case where no noise is present, equation (2.8) is known as the Liouville equation, describing how solutions of ODEs with random data evolve this randomness in time.

The adjoint counterpart of the Fokker–Planck equation is the Chapman–Kolmogorov equation. Let $w(z)$ be a function on $\mathcal{Z}$ and consider the function $v(z_0,t) = \mathbb{E}[w(z(t)) \mid z(0) = z_0]$,

where the expectation is with respect to all Brownian driving paths satisfying $z(0) = z_0$. Then $v(z, t)$ solves the Chapman–Kolmogorov equation

$$\begin{aligned} &\frac{\partial v}{\partial t} = \mathcal{L}v \qquad (z,t) \in \mathbb{R}^d \times (0,\infty), \\ &v = w \qquad (z,t) \in \mathbb{R}^d \times \{0\}. \end{aligned} \tag{2.9}$$

This is the analogue of (2.6) in the countable state space case. If $\gamma \equiv 0$ in (1.1), i.e. the dynamics are deterministic, and φ^t is the flow on $\mathcal{Z}$ so that $z(t) = \varphi^t(z(0))$, then the Chapman–Kolmogorov equation (2.9) reduces to a hyperbolic equation, whose characteristics are the integral curves of the ODE (1.1), and its solution is $v(z, t) = w(\varphi^t(z))$.

We will adopt the semi-group notation, denoting the solution of (2.8) by $\rho(z, t) = \mathrm{e}^{\mathcal{L}^* t}\rho_0(z)$, and the solution of (2.9) by $v(z, t) = \mathrm{e}^{\mathcal{L}t}w(z)$. The connection between the two evolution operators is as follows:

$$\int_{\mathcal{Z}} \rho_0(z)\,(\mathrm{e}^{\mathcal{L}t}w)(z)\,\mathrm{d}z = \int_{\mathcal{Z}} (\mathrm{e}^{\mathcal{L}^* t}\rho_0)(z)w(z)\,\mathrm{d}z,$$

and, in particular, for $\rho_0(z) = \delta(z - y)$:

$$\mathrm{e}^{\mathcal{L}t}w(y) = \int_{\mathcal{Z}} (\mathrm{e}^{\mathcal{L}^* t}\rho_0)(z)w(z)\,\mathrm{d}z. \tag{2.10}$$

This is the analogue of (2.5) in the countable state space case. Both sides of (2.10) represent the expectation value at time t of $w(z(t))$ with respect to the distribution of trajectories that originate at the point y.

2.3. Structured forms for the generator

Several of the ideas developed in this paper can be described in a basic way by refering to equations (2.4)/(2.8) or (2.6)/(2.9).

(i) The reductions used in sections 4–6 all correspond to situations where the generator in (2.6) or (2.9) takes either the form

$$\mathcal{L} = \frac{1}{\epsilon}\mathcal{L}_1 + \mathcal{L}_2 \qquad \text{or} \qquad \mathcal{L} = \frac{1}{\epsilon^2}\mathcal{L}_1 + \frac{1}{\epsilon}\mathcal{L}_2 + \mathcal{L}_3 \tag{2.11}$$

for some $\epsilon \ll 1$, with $\mathcal{L}_1$ being the generator of an ergodic process. Systematic expansions in ϵ then simplify the problem. (The reductions in section 4 are described mainly in the pathwise context, however, not using the Chapman–Kolmogorov equation).

(ii) Although the reductions in section 7 are performed pathwise, they do have an interpretation in terms of (2.9). The starting point is a Chapman–Kolmogorov equation for a deterministic Hamiltonian problem ($\Gamma = 0$), with $\mathcal{L}$ split as the sum of two operators $\mathcal{L}_{\text{particle}}$ and $\mathcal{L}_{\text{bath}}$. The independent variables on which the second of these operators acts are eliminated, by averaging over their random initial data, and an effective Chapman–Kolmogorov equation (with $\Gamma \neq 0$) is found for the variables on which the first of these operators acts.

(iii) The reduction used in section 9 corresponds to a situation where the generator in (2.6) or (2.9) takes the form

$$\mathcal{L} = \mathcal{L}_1 + \epsilon\mathcal{L}_2.$$

In this context $\mathcal{L}_1$ is not ergodic, so that time-rescaling does not yield the situation described in (2.11). Nonetheless, the structure of the problem allows substantial simplifications, using the scale-separation implied by the form of $\mathcal{L}$.

(iv) If we set $w_j = j^k$ for $k = \{0, 1, \ldots\}$, and denote the solution of (2.6) by $v^{(k)}(t)$ (the kth moment), then

$$v_i^{(k)}(t) = \sum_l \mu_l(t) w_l = \sum_l \mu_l(t) l^k,$$

where $\mu(t)$ is defined as above and satisfies (2.2). Notice that, if the evolution is deterministic, then $\mu(t)$ will remain at all times a unit vector, $e_m(t)$, for some fixed integer m. Then

$$v_i^{(k)}(t) = [v_i^{(1)}(t)]^k,$$

so that the first moment characterizes the process completely. This idea is used in section 8 where we show, for certain birth–death processes, that the master equation (2.4) can be approximated by a Liouville equation ((2.8) with $\Gamma \equiv 0$).

(v) The preceding discussion suggests a more a general question: for a given Markov chain on $\mathcal{I}$ starting from state i, do there exist a small number of linear functions of $\mu(t)$ (i.e. expectation values of functions on $\mathcal{I}$) which evolve, at least approximately, as a closed system and hence approximately characterize the behaviour of (some) components of the process? This question is at the heart of the method of optimal prediction, the moment map and the transfer operator approach (see sections 12 and 13).

(vi) Some of the algorithms that we highlight (the moment map and the transfer operator approach) work in discrete time. Then the analogue of (2.4) is the iteration

$$\mu^{n+1} = T\mu^n, \qquad \mu^0 = e_i.$$

(Indeed if $\mu^n = \mu(n\tau)$ for some $\tau > 0$ then $T = \exp^{L^T\tau}$.) Again, if the state space $\mathcal{I}$ is large or infinite, it is natural to ask whether the expectation values of a small number of functions of the process can be used to approximate the whole process, or certain (physically relevant) aspects of its behaviour.

2.4. Discussion and bibliography

A good background reference on Markov chains is Norris [1]. For a discussion of SDEs from the Fokker–Planck viewpoint, see Risken [2] or Gardiner [3]. For a discussion of the generator $\mathcal{L}$, and the Chapman–Kolmogorov equation, see Oksendal [4]. For a discussion concerning ellipticity, hypo-ellipticity and smoothness of solutions to these equations see Rogers and Williams [5]. A comment about terminology: the Fokker–Planck equation is often refered to as Kolmogorov's forward equation in the mathematics literature, whereas the Chapman–Kolmogorov equation is called Kolmogorov's backward equation. The paper [6] provides an overview of variable elimination in a wealth of problems with scale-separation. The papers [7,8] describe a variety of reduction techniques related to problems whose generator has the forms (2.11); they also touch on the heat bath material of section 7.

3. Mori–Zwanzig projection operators

The Mori–Zwanzig formalism is a technique that has been developed in irreversible statistical mechanics to reduce, at least formally, the dimensionality of a system of ODEs. It forms a useful conceptual underpinning of much of what is carried out in this paper, both theoretically and algorithmically. Note, however, that it simply consists of a suggestive rewriting of

equations (1.1). Some structure must be imposed on the problem in order to build from it a useful dimension reduction. Sections 4–9 highlight structures which facilitate dimension reduction.

3.1. Forms of the projection operators

For a system of the form (1.2), with $\alpha = \beta \equiv 0$, the Mori–Zwanzig formalism yields an equation for $x(t)$ of the form

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = \bar{f}(x(t)) + \int_0^t K(x(t-s), s)\,\mathrm{d}s + n(x(0), y(0), t). \tag{3.1}$$

The first term on the right-hand side is only a function of the instantaneous value of x at time t, and therefore represents a 'Markovian' term. The second term depends on values of x at all times between 0 and t, and therefore represents a 'memory' effect. The function $K : \mathcal{X} \times [0, \infty) \mapsto \mathcal{X}$ is the memory kernel. The function $n(x(0), y(0), t)$ satisfies an auxiliary equation, known as the *orthogonal dynamics equation*, and depends on the full knowledge of the initial conditions. If the initial data for $y(0)$ is random then this becomes a random force. The Mori–Zwanzig formalism is a nonlinear extension of the method of undetermined coefficients for variable reduction in linear systems.

The reduction from (1.2) to an equation of the form (3.1) is not unique. It relies on the definition of an operator $\mathscr{P}$, the projection[3], which maps functions of (x, y) into functions of x only. The projection operator that is most appropriate in our context is the following. The state of the system is viewed as random, distributed with a probability density $\rho(x, y)$. Any function $w(x, y)$ then has an expected value, which we denote by $\mathbb{E}w$; the expected value is the best approximation of a function by a constant in an L^2 sense. If the value of the coordinate x is known, then the best approximation to $w(x, y)$ is the conditional expectation of w given x, usually denoted by $\mathbb{E}[w \mid x]$. The conditional expectation defines a mapping $w \mapsto \mathscr{P}w = \mathbb{E}[w \mid x]$ from functions of (x, y) to functions of x. Specifically,

$$(\mathscr{P}w)(x) = \frac{\int_y \rho(x, y) w(x, y)\,\mathrm{d}y}{\int_y \rho(x, y)\,\mathrm{d}y}.$$

With the initial value $y(0)$ being viewed as random, the function $n(x(0), y(0), t)$ is a random function, or a stochastic process. The equation (3.1) is derived such that $n(x(0), y(0), t)$ has zero expected value for all times, which makes it an unbiased 'noise'. In the original context of statistical mechanics, where the governing dynamics are Hamiltonian, the noise $n(x(0), y(0), t)$ and the memory kernel $K(x, t)$ satisfy what is known as the *fluctuation–dissipation relation*[4]. Equation (3.1) is often called a *generalized Langevin equation*. Analogous to the Fokker–Planck versus Chapman–Kolmogorov duality there exist two versions of the Mori–Zwanzig formalism: one for the expectation value of functions and one for probability densities.

3.2. Derivation

The derivation of (3.1) is quite involved, and we only present here a summary. If $\varphi^t(x, y)$ is the flow map induced by (1.2) with $\alpha = \beta = 0$, and P is the projection $(x, y) \mapsto x$, then the

[3] Not to be confused with the projection P defined in the introduction.

[4] When the projection $\mathscr{P}$ maps functions of (x, y) onto the subspace of functions of x that depend on x linearly, then the fluctuation–dissipation relation is of the form $K(x, t) = \kappa(t)x$, where the matrix $\kappa(t)$ is proportional to the auto-covariance of the noise. In the general case the fluctuation–dissipation relation takes a more complicated form (see [9] for details).

function $x(t)$ is more accurately written as $P\varphi^t(x, y)$, where (x, y) are the initial data. The x-equation in (1.2) is

$$\frac{\partial}{\partial t} P\varphi^t(x, y) = f(\varphi^t(x, y)), \tag{3.2}$$

and this, of course, is not a closed equation for $P\varphi^t(x, y)$.

The first step in the Mori–Zwanzig formalism is to replace f on the right-hand side by its best approximation given only its first argument. Thus (3.2) is rewritten in the following equivalent way:

$$\frac{\partial}{\partial t} P\varphi^t(x, y) = (\mathscr{P}f)(P\varphi^t(x, y)) + [f(\varphi^t(x, y)) - (\mathscr{P}f)(P\varphi^t(x, y))]. \tag{3.3}$$

The function $\mathscr{P}f$ is identified with $\bar{f}$ in (3.1).

The next stage is to rearrange the terms in the square brackets in (3.3). Defining the operator $\mathcal{L} = f(x, y) \cdot \nabla_x + g(x, y) \cdot \nabla_y$ (which is (2.7) in the case of no noise, $\Gamma = 0$), the noise function $n(x, y, t)$ is defined as the solution of the orthogonal dynamics equation:

$$\begin{aligned} &\frac{\partial n}{\partial t} = (I - \mathscr{P})\mathcal{L}n, \\ &n(x, y, 0) = f(x, y) - (\mathscr{P}f)(x). \end{aligned} \tag{3.4}$$

The memory kernel is defined as

$$K(x, t) = \mathscr{P}\mathcal{L}n(x, y, t).$$

One can then check explicitly that the residual terms in (3.3) can be written as

$$f(\varphi^t(x, y)) - (\mathscr{P}f)(P\varphi^t(x, y)) = n(x, y, t) + \int_0^t K(P\varphi^{t-s}(x, y), s)\, \mathrm{d}s, \tag{3.5}$$

hence (3.3) takes the form (3.1).

To understand the last identity, we note that the left-hand side can be written as

$$\mathrm{e}^{t\mathcal{L}}(I - \mathscr{P})\mathcal{L}w,$$

where $w = w(z) = Pz$, hence $\mathcal{L}w(z) = f(z)$; semi-group notation has been used for the solution operator of the flow map $\varphi^t(z)$ which is defined so that

$$\exp(t\mathcal{L})w(z) = w(\varphi^t(z)).$$

The noise can, likewise, be written in the form

$$n(z, t) = \exp[t(I - \mathscr{P})\mathcal{L}]\,(I - \mathscr{P})\mathcal{L}w,$$

so that the right-hand side of (3.5) reads

$$\epsilon^{t(I-\mathscr{P})\mathcal{L}}\,(I - \mathscr{P})\mathcal{L}w + \int_0^t \mathrm{e}^{(t-s)\mathcal{L}}\mathscr{P}\mathcal{L}\mathrm{e}^{s(I-\mathscr{P})\mathcal{L}}\,(I - \mathscr{P})\mathcal{L}w\, \mathrm{d}s.$$

The validity of (3.5) is a consequence of the operator identity,

$$\mathrm{e}^{t\mathcal{L}} = \mathrm{e}^{t(I-\mathscr{P})\mathcal{L}} + \int_0^t \mathrm{e}^{(t-s)\mathcal{L}}\mathscr{P}\mathcal{L}\mathrm{e}^{s(I-\mathscr{P})\mathcal{L}}\, \mathrm{d}s,$$

known in the physics literature as Dyson's formula [10].

It is important to point out that (3.1) is not simpler than the original problem. The complexity has been transfered, in part, to the solution of the orthogonal dynamics (3.4). The value of (3.1) is first, in its being conceptual, and second, that it constitutes a good starting point for asymptotic analysis and stochastic modelling. In particular, it suggests that deterministic problems with random data may be modelled by stochastic problems with memory. In the case where the memory can be well-approximated by the introduction of a small number of extra variables so that the whole system is then Markovian in time, this leads to a simple low-dimensional stochastic model for the dynamics in $\mathcal{X}$. This basic notion underlies many of the examples used in the following.

3.3. Discussion and bibliography

The original derivation of the Mori–Zwanzig formalism can be found in Mori [11] and Zwanzig [12]. Most of the statistical physics literature uses the Mori–Zwanzig formalism with a projection on the span of linear functions of the essential degrees of freedom. In [9], Zwanzig developed a nonlinear generalization, which is equivalent to the conditional expectation used here. All the above references use the 'Chapman–Kolmogorov' version of this formalism; the 'Fokker–Planck' version can be found in [13]. The existence of solutions to the orthogonal dynamics equation (3.4) turns out to be subtle; see Givon *et al* [14]. An alternative to the Mori–Zwanzig approach is to derive convolutionless, non-Markovian evolution equations; this approach, and an application in plasma physics, is outlined in [15] and a more recent application is contained in [16]. Recent uses of the Mori–Zwanzig formalism in the context of variable reduction can be found in Just *et al* [17, 18] and Chorin *et al* [19].

4. Scale-separation and invariant manifolds

In the classification of section 1, this section is devoted to problems of the type D–D—deterministic systems with lower-dimensional deterministic systems embedded within them. The key mathematical construct used is that of invariant manifolds, and scale-separation is the mechanism by which these may be constructed. These ideas typically can be applied for all initial conditions $y(0) \in \mathcal{B} \subseteq \mathcal{Y}$, and the effect of the initial condition in $\mathcal{B}$ becomes irrelevant after an initial transient.

In the space of probability densities we are showing that, after an initial transient, the Liouville equation (the hyperbolic PDE (2.8) with $\Gamma = 0$) for $\rho(x, y)$, has a solution that can be approximated by a function of the form

$$\rho(x, y, t) = \delta(y - \eta(x))\bar{\rho}(x, t), \tag{4.1}$$

where $\bar{\rho}(x, t)$ satisfies a Liouville equation of lower spatial dimension than that for $\rho(x, y, t)$.

4.1. Spectral gaps

Consider equations (1.2) in the deterministic setting when $\alpha, \beta \equiv 0$ and let $f(x, y)$ and $g(x, y)$ be written in the following form:

$$\begin{aligned} f(x, y) &= L_1 x + \hat{f}(x, y), \\ g(x, y) &= L_2 y + \hat{g}(x, y). \end{aligned} \tag{4.2}$$

Assume for simplicity that the operators L_1, L_2 are self-adjoint and that the maximum eigenvalue of L_2 is less than the minimum of L_1. If the gap in the spectra of L_1 and L_2 is large then for the purely linear problem, found by dropping $\hat{f}$ and $\hat{g}$, the dynamics is dominated, relatively speaking, by the dynamics in $\mathcal{X}$. In the fully nonlinear problem, if the gap is assumed to be large relative to the size of $\hat{f}$ and $\hat{g}$ (and the argument can be localized by the use of cut-off functions), then the existence of an exponentially attractive invariant manifold

$$y = \eta(x)$$

may be proved. Thus, after an initial transient, the effective dynamics in $\mathcal{X}$ is governed by the approximating equation

$$\begin{aligned} &\frac{\mathrm{d}X}{\mathrm{d}t} = L_1 X + \hat{f}(X, \eta(X)), \\ &X(0) = x(0). \end{aligned}$$

Specifically, under suitable conditions on the spectra of L_1 and L_2, $X(t)$ and $x(t)$ remain close, at least on bounded time intervals. In many cases, the validity of this approximation requires that the initial value for the eliminated variables $y(0)$ is taken from within some bounded set $\mathcal{B}$.

4.2. Model problem

A useful model problem arises when $\mathcal{X} = \mathbb{R}^n$, $\mathcal{Y} = \mathbb{R}$ and (1.2) has the form, for $\epsilon \ll 1$,

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= f(x, y), \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= -\frac{y}{\epsilon} + \frac{\tilde{g}(x)}{\epsilon}. \end{aligned} \tag{4.3}$$

In reference to the spectral properties of L_1, L_2 in (4.2), the unique eigenvalue of L_2 is $-1/\epsilon$, which for small enough ϵ is less than the minimum eigenvalue of the linear component of $f(x, y)$ which is ϵ independent.

Assume that f, $\tilde{g}$ are smooth and bounded, with all derivatives bounded. Then, seeking an approximate invariant manifold in the form

$$y = \tilde{g}(x) + O(\epsilon) \tag{4.4}$$

gives, up to errors of $O(\epsilon)$, the reduced dynamics

$$\frac{\mathrm{d}X}{\mathrm{d}t} = f(X, \tilde{g}(X)), \tag{4.5}$$

with $X(0) = x(0)$. Note that for the derivation to be consistent the initial value $y(0)$ should be close to the invariant manifold, $|y(0) - \tilde{g}(x(0))| = O(\epsilon)$, otherwise, an initial layer forms near time $t = 0$; after this initial layer the initial condition in $\mathcal{Y}$ is essentially forgotten.

Example 4.1. Consider the equations

$$\begin{aligned} \frac{\mathrm{d}x_1}{\mathrm{d}t} &= -x_2 - x_3, \\ \frac{\mathrm{d}x_2}{\mathrm{d}t} &= x_1 + \frac{1}{5}x_2, \\ \frac{\mathrm{d}x_3}{\mathrm{d}t} &= \frac{1}{5} + y - 5x_3, \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= -\frac{y}{\epsilon} + \frac{x_1 x_3}{\epsilon}, \end{aligned} \tag{4.6}$$

so that $\mathcal{X} = \mathbb{R}^3$ and $\mathcal{Y} = \mathbb{R}$. The expression (4.4) with $\epsilon = 0$ indicates that, for small ϵ, the equations have the invariant manifold

$$y = \eta(x), \qquad \eta(x) = x_1 x_3 + \mathcal{O}(\epsilon). \tag{4.7}$$

Thus, $y \approx x_1 x_3$, so that the solution for $x = (x_1, x_2, x_3)$ should be well approximated by $X = (X_1, X_2, X_3)$ solving the Rössler system [20]

$$\begin{aligned} \frac{\mathrm{d}X_1}{\mathrm{d}t} &= -X_2 - X_3, \\ \frac{\mathrm{d}X_2}{\mathrm{d}t} &= X_1 + \frac{1}{5}X_2, \\ \frac{\mathrm{d}X_3}{\mathrm{d}t} &= \frac{1}{5} + X_3(X_1 - 5). \end{aligned} \tag{4.8}$$

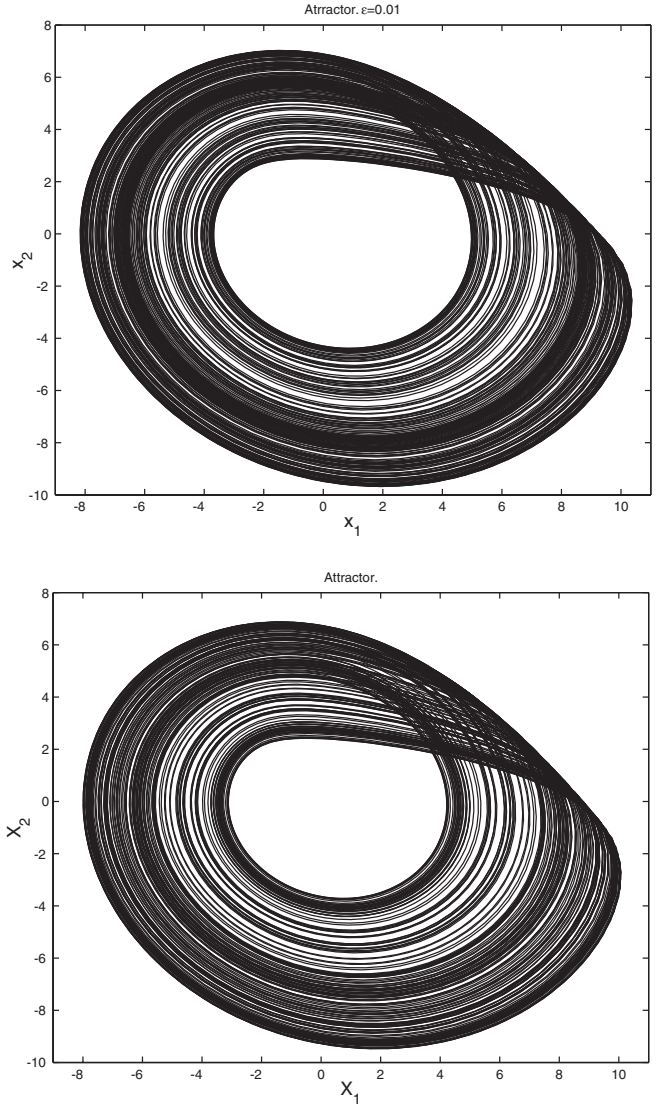

Figure 1. Comparison between the attracting sets for (4.6) with $\epsilon = 0.01$ (top) and (4.8) (bottom), projected on the (x_1, x_2) and (X_1, X_2) planes, respectively.

Figure 1 shows the attractor for (x, y) at $\epsilon = 0.01$, projected onto (x_1, x_2), compared with the attractor for X projected onto (X_1, X_2). Notice the clear similarities between the two. Figure 2 compares the histograms for x_1 and X_1 over 10^5 time units; the two histograms are also, clearly, closely related. ■

4.3. Discussion and bibliography

Early studies of the reduction of ODEs with an attracting slow manifold into differential-algebraic equations (DAE) includes the independent work of Levinson and of Tikhonov (see O'Malley [21] and Tikhonov *et al* [22]). The use of a spectral gap that is sufficiently large

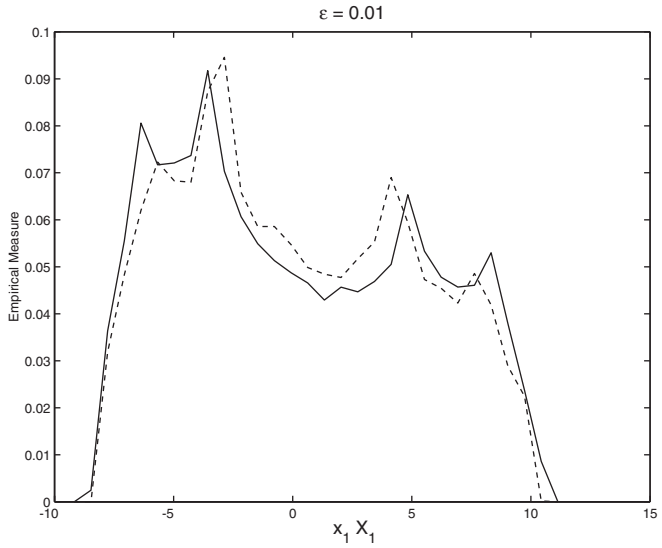

Figure 2. Comparison between the empirical measures of x_1 solving (4.6) with $\epsilon = 0.01$ (——), and X_1 solving (4.8) (- - - -). The empirical measure was computed from a trajectory over a time interval of 10^5 units.

relative to the size of the nonlinear terms is used in the construction of local stable, unstable and centre manifolds (e.g. Carr [23], Wiggins [24]), slow manifolds (Kreiss [25]) and inertial manifolds (Constantin *et al* [26]). A variety of methods of proof exist, predominantly the Lyapunov–Perron approach (Hale [27], Temam [28]) and the Hadamard graph transform (Wells [29]). Important work in this area is due to Fenichel [30, 31] who sets up a rather general construction of normally hyperbolic invariant manifolds.

5. Scale-separation and averaging

There is a vast literature on systems, which in the classification of section 1 are of type D–D, that can be unified under the title of 'averaging methods'. Averaging methods have their early roots in celestial mechanics, but apply to a broad range of applications.

Averaging methods are concerned with situations where, for fixed x, the trajectories of the y-dynamics do not tend to a fixed point, as happened in the previous section. Instead, the fast dynamics affect the slow dynamics through the empirical measure that its trajectories induce on $\mathcal{Y}$. The simplest such situation is where the fast dynamics converge to a periodic solution; other possibilities are convergence to quasi-periodic solutions, or chaotic solutions.

In the space of probability densities we are showing that, after an initial transient, the Liouville equation (the hyperbolic PDE (2.8) with $\Gamma = 0$) for $\rho(x, y)$, has solution that can be approximated by a function of the form

$$\rho(x, y, t) = \rho_\infty(y; x)\bar{\rho}(x, t), \tag{5.1}$$

where $\bar{\rho}(x, t)$ satisfies a Liouville equation of lower spatial dimension than that for $\rho(x, y, t)$. This general picture subsumes the case of the previous section, where ρ_∞ is a delta measure on $y = \eta(x)$.

5.1. The averaging method

For concreteness, we limit our discussion to systems with scale-separation of the form

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= f(x, y), \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= \frac{1}{\epsilon} g(x, y), \end{aligned} \tag{5.2}$$

where $\epsilon \ll 1$.

The starting point is to analyse the behaviour of the fast dynamics, with x being a parameter. In the previous section we considered systems in which the fast dynamics converge to an x-dependent fixed point. This gives rise to a situation where the y variables are 'slaved' to the x variables. Averaging generalizes this idea to situations where the dynamics in the y variable, with x fixed, is more complex.

We start by discussing the case when the dynamics for y is ergodic. A general theorem on averaging, due to Anosov, applies in this case. Let $\varphi_x^t(y)$ be the solution operator of the fast dynamics with x a fixed parameter. Then

$$\frac{\mathrm{d}}{\mathrm{d}t}\varphi_x^t(y) = g(x, \varphi_x^t(y)), \qquad \varphi_x^0(y) = y \tag{5.3}$$

(the $1/\epsilon$ rate factor has been omitted because time can be rescaled when the fast dynamics is considered alone). The fast dynamics is said to be ergodic, for given fixed x, if for all (sufficiently well-behaved) functions $\psi : \mathcal{Y} \to \mathbb{R}$ the limit of the time-average

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T \psi(\varphi_x^t(y))\, \mathrm{d}t,$$

exists and is independent of y. In particular, ergodic dynamics define an ergodic measure, μ_x, on $\mathcal{Y}$, which is invariant under the fast dynamics; note that the invariant measure depends, in general, on x. The measure is defined by

$$\mu_x(A) = \lim_{T\to\infty} \frac{1}{T} \int_0^T I_A(\varphi_x^t(y_0))\, \mathrm{d}t, \qquad A \subseteq \mathcal{Y},$$

and is independent of y_0 by assumption; I_A is the indicator function of the set A. Anosov's theorem states that, under ergodicity in $\mathcal{Y}$ for fixed x, the slow variables $x(t)$ converge uniformly on any bounded time interval to the solution $X(t)$ of the averaged equation,

$$\begin{aligned} \frac{\mathrm{d}X}{\mathrm{d}t} &= F(X), \\ F(\zeta) &= \int_{\mathcal{Y}} f(\zeta, y)\, \mu_\zeta(\mathrm{d}y). \end{aligned} \tag{5.4}$$

Similar ideas prevail if the invariant measure generated by the y-dynamics depends upon the initial data in $\mathcal{Y}$. The results are complex to state in general. The next subsection contains an example where the initial data in $\mathcal{Y}$ enter the averaged equation.

5.2. Stiff Hamiltonian systems

An application area where averaging techniques are of current interest is Hamiltonian mechanics. This is hence a natural source of model problems. One encounters Hamiltonian systems with strong potential forces, responsible for fast, small amplitude, oscillations around a constraining sub-manifold. An important goal is to describe the evolution of the slowly evolving degrees of freedom by averaging over the rapidly oscillating variables. The study of

such problems was initiated by Rubin and Ungar [32]. More recently the ideas of Neistadt [33], based on normal form theory, have been applied to such problems [34]; this approach is very powerful, yielding very tight, exponential, error estimates between the original and limiting variables. A recent approach to the problem, using the techniques of time-homogenization [35], is the paper [36].

The general setting is a Hamiltonian of the form

$$H(z, p) = \frac{1}{2}\sum_j \frac{p_j^2}{2m_j} + V(z) + \frac{1}{\epsilon^2}U(z), \tag{5.5}$$

where $z = (z_1, \ldots, z_{n+m})$ and $p = (p_1, \ldots, p_{n+m})$ are the coordinates and momenta, $V(z)$ is a 'soft' potential, whereas $\epsilon^{-2}U(z)$ is a 'stiff' potential. It is assumed that $U(z)$ attains a global minimum, 0, on a smooth n-dimensional manifold, $\mathcal{M}$. The limiting behaviour of the system, as $\epsilon \to 0$, depends crucially on the choice of initial conditions. The setting appropriate to molecular systems is where the total energy E (which is conserved) is assumed independent of ϵ. Then, as $\epsilon \to 0$, the states of the system are restricted to a narrow band in the vicinity of $\mathcal{M}$; the goal is to approximate the evolution of the system by a flow on $\mathcal{M}$.

Example 5.1. The following simple example, taken from [36], shows how problems with this form of Hamiltonian can be cast in the general set-up of equations (5.2). Consider a two-particle system with Hamiltonian,

$$H(x, p, y, v) = \frac{1}{2}(p^2 + v^2) + V(x) + \frac{\omega(x)}{2\epsilon^2}y^2,$$

where (x, y) and (p, v) are the respective coordinates and momenta of the two particles, $V(x)$ is a non-negative potential and $\omega(x)$ is assumed to be uniformly bounded away from zero, to ensure a strict separation of scales: $\omega(x) \geqslant \bar{\omega} > 0$ for all x. The corresponding equations of motion are

$$\begin{aligned}
\frac{\mathrm{d}x}{\mathrm{d}t} &= p,\\
\frac{\mathrm{d}p}{\mathrm{d}t} &= -V'(x) - \frac{\omega'(x)}{2\epsilon^2}y^2,\\
\frac{\mathrm{d}y}{\mathrm{d}t} &= v,\\
\frac{\mathrm{d}v}{\mathrm{d}t} &= -\frac{\omega(x)}{\epsilon^2}, y.
\end{aligned}$$

The assumption that the energy E does not depend on ϵ implies that $y^2 \leqslant 2\epsilon^2 E/\bar{\omega}$ and hence that the solution approaches the sub-manifold $y = 0$ as $\epsilon \to 0$. Note, however, that y appears in the combination y/ϵ in the x equations. Thus it is natural to make the change of variables $\eta = y/\epsilon$. The equations then read

$$\begin{aligned}
\frac{\mathrm{d}x}{\mathrm{d}t} &= p,\\
\frac{\mathrm{d}p}{\mathrm{d}t} &= -V'(x) - \frac{\omega'(x)}{2}\eta^2,\\
\frac{\mathrm{d}\eta}{\mathrm{d}t} &= \frac{1}{\epsilon}v,\\
\frac{\mathrm{d}v}{\mathrm{d}t} &= -\frac{\omega(x)}{\epsilon}\eta.
\end{aligned} \tag{5.6}$$

In these variables we recover a system of the form (5.2) with 'slow' variables, $x \leftarrow (x, p)$, and 'fast' variables, $y \leftarrow (\eta, v)$. The fast equations represent an harmonic oscillator whose frequency, $\omega^{1/2}(x)$, is modulated by the x variables.

The limiting solution of a fast modulated oscillator can be derived using a WKB expansion [37], but it is also instructive to consider the following heuristic approach. Suppose that the slow variables (x, p) are given. Then the energy available for the fast variables is

$$H_\eta(x, p) = E - \tfrac{1}{2}p^2 - V(x).$$

Harmonic oscillators satisfy an equipartition property, whereby, on average, the energy is equally distributed between its kinetic and potential contributions (the virial theorem). Thus the time-average of the kinetic energy of the fast oscillator is

$$\left\langle \frac{\omega(x)}{2}\eta^2 \right\rangle = \frac{1}{2}\left[E - \frac{1}{2}p^2 - V(x)\right],$$

where (x, p) are viewed as fixed parameters and the total energy E is specified by the initial data. The averaging principle states that the rapidly varying η^2 in the equation for p can be approximated by its time-average, giving rise to a closed system of equations for $(X, P) \approx (x, p)$,

$$\begin{aligned} \frac{\mathrm{d}X}{\mathrm{d}t} &= P, \\ \frac{\mathrm{d}P}{\mathrm{d}t} &= -V'(X) - \frac{\omega'(X)}{2\omega(X)}\left[E - \frac{1}{2}P^2 - V(X)\right], \end{aligned} \tag{5.7}$$

with initial data E, $X_0 = x_0$ and $P_0 = p_0$. It may be verified that (X, P) satisfying (5.7) conserve the following invariant,

$$\frac{1}{\omega^{1/2}(X)}\left[E - \frac{1}{2}P^2 - V(X)\right].$$

Thus, (5.7) reduces to the simpler form

$$\begin{aligned} \frac{\mathrm{d}X}{\mathrm{d}t} &= P, \\ \frac{\mathrm{d}P}{\mathrm{d}t} &= -V'(X) - J[\omega^{1/2}(X)]', \end{aligned} \tag{5.8}$$

where the adiabatic invariant J is given by

$$J = \frac{1}{\omega^{1/2}(X_0)}\left[E - \frac{1}{2}P_0^2 - V(X_0)\right].$$

This means that the influence of the stiff potential on the slow variables is to replace the potential $V(x)$ by an effective potential,

$$V_{\mathrm{eff}}(x) = V(x) + J\omega^{1/2}(x).$$

Note that the limiting equation contains a memory of the initial conditions for the fast variables through the constant, J. Thus the situation differs slightly from the Anosov theorem described previously. The heuristic derivation we have given here is made rigorous in [36], using time-homogenization techniques, and it is also generalized to higher dimension. Resonances become increasingly important as the co-dimension, m, increases, limiting the applicability of the averaging approach to such two-scale Hamiltonian systems (Takens [38]).

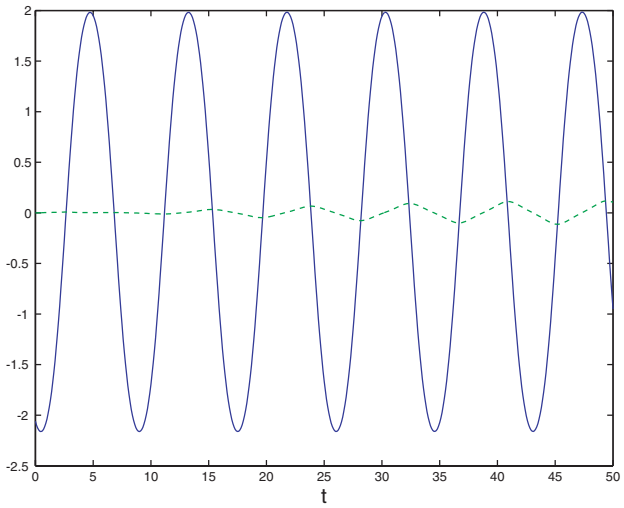

Figure 3. Time evolution of $x(t)$ solving (5.6) with $\epsilon = 0.1$ (——) and $X(t) - x(t)$, with $X(t)$ solving (5.8) (- - - -). We used $V(x) = -\cos x$ and $\omega(x) = 1 + 0.5 \sin x$.

Figure 3 shows a comparison of the solution of x solving (5.6) with $\epsilon = 0.1$ and X solving (5.8); we took $V(x) = -\cos x$ and $\omega(x) = 1 + 0.5 \sin x$. The time traces are of x and $X - x$. The results illustrate the foregoing analysis since $X - x$ is small. ■

5.3. Discussion and bibliography

A detailed account of the averaging method, as well as numerous examples can be found in Sanders and Verhulst [39] (see also [40]). An English language review of the Russian literature can be found in Lochak and Meunier [41]. An overview of the topic of slow manifolds, especially in the context of Hamiltonian problems, may be found in [42]. The averaging method applied to equations (5.2) is analysed in an instructive manner in [43], where the Liouville equation is used to construct a rigorous proof of the averaged limit.

Anosov's theorem requires the fast dynamics to be ergodic. Often ergodicity fails due to the presence of 'resonant zones'—regions in $\mathcal{X}$ for which the fast dynamics is not ergodic. Arnold and Neistadt [41] extended Anosov's result to situations in which the ergodicity assumption fails on a sufficiently small set of $x \in \mathcal{X}$. Those results were further generalized and extended to the stochastic framework by Kifer, who also studied the diffusive and large deviation character of the discrepancy between the effective and exact solution [44–47].

The situations in which the fast dynamics tend to fixed points, periodic solutions or chaotic solutions can be treated in a unified manner by the introduction of Young measures. Artstein and co-workers considered a class of singularly perturbed system of type (5.2), with attention given to the limiting behaviour of both slow and fast variables. In all the above cases the pair (x, y) can be shown to converge to (X, μ_X), where X is the solution of

$$\frac{\mathrm{d}X}{\mathrm{d}t} = \int_{\mathcal{Y}} f(X, y)\, \mu_X(\mathrm{d}y),$$

and μ_X is the ergodic measure on $\mathcal{Y}$; the convergence of y to μ_X is in the sense of Young measures. (In the case of a fixed point the Young measure is concentrated at a point.) A general theorem along these lines is proved in [48].

There are many generalizations of this idea. The case of non-autonomous fast dynamics, as well as a case with infinite dimension are covered in [49]. Moreover, these results still make

sense even if there is no unique invariant measure μ_x, in which case the slow variables can be proved to satisfy a (non-deterministic) differential inclusion [50].

In the context of SDEs, an interesting generalization of (5.2) is to consider systems of the form

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= f(x, y), \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= \frac{1}{\epsilon} g(x, y) + \frac{1}{\sqrt{\epsilon}} \frac{\mathrm{d}V}{\mathrm{d}t}. \end{aligned} \tag{5.9}$$

If the y-dynamics, with x frozen at ζ, is ergodic, then the analogue of the Anosov result holds with μ_ζ the invariant measure of this y-dynamics. This gives rise to a set-up of type S–D. The Fokker–Planck equation is, in this case, a degenerate parabolic equation—with diffusion only in y—and we seek an approximation of the form (5.1) where $\bar{\rho}(x, t)$ solves a hyperbolic PDE (Liouville equation) when ϵ is small. This idea is generalized in section 6.1, where the chosen scaling leads not only to an averaged deterministic vector field in x, but also to additional stochastic fluctuations.

6. Scale-separation and white noise approximation

In this section we consider how to use the PDEs for propagation of expectations and probability densities to study stochastic dimension reduction when there is a clear scale-separation between the x- and y-dynamics. The effective dynamics in $\mathcal{X}$ is stochastic; the original dynamics in $\mathcal{Z}$ may be deterministic or stochastic. Thus we study problems of the form D–S or S–S in the classification of section 1. The ideas will be developed here for problems with a *skew-product* structure: the dynamics for y evolves independently of the dynamics for x. This simplifies the analysis, but is not necessary; in the final subsection we review the literature in which full back-coupling between the variables is present.

In the skew-product case, when the inverse scale-separation is small, we will look for an approximate probability density of the form

$$\rho(x, y, t) \approx \rho_\infty(y)\bar{\rho}(x, t),$$

with $\rho_\infty(\cdot)$ a smooth probability density on $\mathcal{Y}$, invariant under the y-dynamics. Note that $\rho(x, y, t)$ solves a (possibly degenerate) parabolic equation (Fokker–Planck equation (2.8)) in the case where the $\mathcal{Z}$-dynamics is stochastic, and a hyperbolic PDE (the Liouville equation (2.8) with $\Gamma = 0$) when it is deterministic. The function $\bar{\rho}(x, t)$ satisfies a Fokker–Planck equation. The ansatz that we use here assumes that the distribution of y reaches equilibrium on a time-scale much shorter than the time-scale over which x evolves. This is the probabilistic analogue of the slaving and averaging techniques of the previous two sections.

When full back-coupling is present, the approximate solution will take the form

$$\rho(x, y, t) \approx \rho_\infty(y; x)\bar{\rho}(x, t), \tag{6.1}$$

where $\rho_\infty(x, y)$ is a density invariant under the y-dynamics, with x viewed as a fixed parameter.

We start by studying the approach based on the Chapman–Kolmogorov picture and then study the problem again using the Fokker–Planck approach. The two approaches are each illustrated by a simple example, accompanied by numerical results. The final subsection overviews the literature and describes a variety of extensions of the basic idea.

6.1. Chapman–Kolmogorov picture

Consider the case of (1.2) where $\alpha \equiv 0$ and f, g, β are of the form

$$\begin{aligned} f(x, y) &= \frac{1}{\epsilon} f_0(x, y) + f_1(x, y), \\ g(x, y) &= \frac{1}{\epsilon^2} g_0(y), \\ \beta(x, y) &= \frac{1}{\epsilon} \beta_0(y). \end{aligned} \tag{6.2}$$

This leads to model problems of the form

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= \frac{1}{\epsilon} f_0(x, y) + f_1(x, y), \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= \frac{1}{\epsilon^2} g_0(y) + \frac{1}{\epsilon} \beta_0(y) \frac{\mathrm{d}V}{\mathrm{d}t}. \end{aligned} \tag{6.3}$$

Both the x and y equations contain fast dynamics, but the dynamics in y is an order of magnitude faster than x (note that white noise scales differently from regular time derivatives and that, in the Fokker–Planck picture, the contributions from both the drift term, g, and the diffusion term, β, are of order $1/\epsilon^2$). Then the variable y induces fluctuations in the equation for x, (which we will see below are formally of order $1/\epsilon$). We are going to assume that $f_0(x, y)$ averages to zero under the y-dynamics, but that $f_1(x, y)$ does not necessarily do so. In certain situations it can then be shown that both terms in f contribute at the same order. The term f_0 will give the effective stochastic contribution and, together with f_1, the effective drift. One way to see this is by using the Chapman–Kolmogorov equation and we now develop this idea. The underlying theory in this area is developed by Kurtz [51]. We follow the presentation given in [7], where the perturbative structure of the techniques is highlighted; the work of [7] was exploited recently in [52].

Recall that $v(x, y, t)$ satisfying the Chapman–Kolmogorov equation (2.9), namely

$$\frac{\partial v}{\partial t} = \mathcal{L} v, \qquad v(x, y, 0) = w(x, y),$$

is the expected value at time t of $w(\cdot)$ over all solutions starting at the point (x, y); the probability space is induced by the Brownian motion in the y variables. If w is only a function of x, then $v(x, y, t)$ (which remains a function of both x and y) describes the expected evolution of a property pertinent to the essential dynamics on $\mathcal{X}$.

Substituting (6.2) into the Chapman–Kolmogorov equation (2.9) with $w = w(x)$ gives,

$$\frac{\partial v}{\partial t} = \frac{1}{\epsilon^2} \mathcal{L}_1 v + \frac{1}{\epsilon} \mathcal{L}_2 v + \mathcal{L}_3 v, \qquad v(x, y, 0) = w(x), \tag{6.4}$$

where

$$\mathcal{L}_1 v = g_0 \cdot \nabla_y v + \tfrac{1}{2} (\beta_0 \beta_0^T) : \nabla_y (\nabla_y v), \tag{6.5}$$

$$\mathcal{L}_2 v = f_0 \cdot \nabla_x v, \tag{6.6}$$

$$\mathcal{L}_3 v = f_1 \cdot \nabla_x v. \tag{6.7}$$

We seek an expansion for the solution with the form

$$v = v_0 + \epsilon v_1 + \epsilon^2 v_2 + \cdots.$$

Substituting this expansion into (6.4) and equating powers of ϵ gives a hierarchy of equations, the first three of which are

$$\mathcal{L}_1 v_0 = 0, \tag{6.8}$$

$$\mathcal{L}_1 v_1 = -\mathcal{L}_2 v_0, \tag{6.9}$$

$$\mathcal{L}_1 v_2 = \frac{\partial v_0}{\partial t} - \mathcal{L}_2 v_1 - \mathcal{L}_3 v_0. \tag{6.10}$$

The initial conditions are that $v_0 = w$ and $v_i = 0$ for $i \geqslant 1$.

Note that $\mathcal{L}_1$, given by (6.5), is the Chapman–Kolmogorov operator constrained to the y-dynamics, and that constants (in y) are in the null-space of $\mathcal{L}_1$. Assume that there is a unique density $\rho_\infty(y)$ in the null-space of $\mathcal{L}_1^*$ (i.e. a unique density invariant under the y-dynamics), and denote by $\langle\cdot\rangle$ averaging with respect to this density. Assume further that the dynamics is ergodic in the sense that any initial density $\rho_0(y)$, including a Dirac mass, tends, as $t \to \infty$, to the unique invariant density $\rho_\infty(y)$; the system 'returns to equilibrium'. By (2.10)

$$\lim_{t\to\infty} \mathrm{e}^{\mathcal{L}_1 t}\phi(y_0) = \lim_{t\to\infty} \int_{\mathcal{Y}} \phi(y) \left(\mathrm{e}^{\mathcal{L}_1^* t}\rho_0\right)(y)\,\mathrm{d}z = \int_{\mathcal{Y}} \phi(y)\rho_\infty(y)\,\mathrm{d}y = \langle\phi\rangle, \tag{6.11}$$

where $\rho_0(y) = \delta(y - y_0)$ and the limit is attained for any y_0. Later we will also assume that the operator $\mathcal{L}_1$ is negative definite on the inner product space weighted by the invariant density and excluding constants; this kind of spectral gap is characteristic of many ergodic systems.

Note that constants in y are in the null-space of $\mathcal{L}_1$. We now argue that all functions ϕ satisfying $\mathcal{L}_1\phi = 0$ are independent of y. Indeed,

$$(\mathcal{L}_1\phi)(y) = 0 \;\Rightarrow\; \frac{\mathrm{d}}{\mathrm{d}s}(\mathrm{e}^{\mathcal{L}_1 s}\phi)(y) = \mathrm{e}^{\mathcal{L}_1 s}\mathcal{L}_1\phi(y) = 0,$$

so that $\phi(y) = \mathrm{e}^{\mathcal{L}_1 s}\phi(y)$ for all s. Letting $s \to \infty$ and using (6.11) we get

$$\phi(y) = \langle\phi\rangle,$$

and the latter is independent of y. Thus, the first equation (6.8) in the hierarchy implies that $v_0 = v_0(x, t)$.

Consider next the v_1 equation (6.9). For it to be solvable, $\mathcal{L}_2 v_0$ has to be orthogonal to the kernel of $\mathcal{L}_1^*$, which by assumption contains only $\rho_\infty(y)$. Thus, orthogonality to the kernel of $\mathcal{L}_1^*$ amounts to averaging to zero under the y-dynamics. The solvability condition is then $\langle\mathcal{L}_2 v_0\rangle = 0$, or substituting (6.6):

$$\langle f_0\rangle \cdot \nabla_x v_0(x, t) = 0.$$

Thus, for the expansion to be consistent it suffices that $\langle f_0\rangle \equiv 0$; this means that the leading-order x-dynamics averages to zero under the invariant measure of the y-dynamics. It follows that the equation for v_1 is solvable and we may formally write

$$v_1 = -\mathcal{L}_1^{-1}\mathcal{L}_2 v_0.$$

Similarly, considering (6.10) the solvability condition for v_2 becomes

$$\frac{\partial v_0}{\partial t} = -\langle\mathcal{L}_2\mathcal{L}_1^{-1}\mathcal{L}_2 v_0\rangle + \langle\mathcal{L}_3 v_0\rangle. \tag{6.12}$$

In view of the fact that $\mathcal{L}_2$ and $\mathcal{L}_3$ are first-order differential operators in x, that $\mathcal{L}_1$ involves only y, and $\langle\cdot\rangle$ denotes y averaging, this is a Chapman–Kolmogorov equation for an Itô SDE in $\mathcal{X}$:

$$\frac{\mathrm{d}X}{\mathrm{d}t} = F(X) + A(X)\frac{\mathrm{d}U}{\mathrm{d}t}, \tag{6.13}$$

U being standard Brownian motion. That is, (6.12) is of the form

$$\frac{\partial v_0}{\partial t} = F(x)\cdot\nabla_x v_0 + \frac{1}{2}[A(x)A(x)^T] : \nabla_x(\nabla_x v_0),$$

To see how $F(X)$ and $A(X)$ are determined note that, by virtue of the linearity of $\mathcal{L}_1$ and structure of $\mathcal{L}_2$,

$$\mathcal{L}_1^{-1}\mathcal{L}_2 v_0 = \mathcal{L}_1^{-1} f_0 \cdot \nabla_x v_0 = r \cdot \nabla_x v_0,$$

where $r = r(x, y)$ solves the *cell problem*

$$\mathcal{L}_1 r(x, y) = f_0(x, y).$$

Hence

$$\mathcal{L}_2\mathcal{L}_1^{-1}\mathcal{L}_2 v_0 = f_0 \cdot \nabla_x(r \cdot \nabla_x v_0) = (f_0 r^T) : \nabla_x(\nabla_x v_0) + \nabla_x v_0 \cdot (\nabla_x r) f_0, \tag{6.14}$$

and (6.12) takes the explicit form

$$\frac{\partial v_0}{\partial t} = \nabla_x v_0 \cdot \langle f_1 - (\nabla_x r) f_0 \rangle + \frac{1}{2}\langle -2 r f_0^T \rangle : \nabla_x(\nabla_x v_0).$$

Thus $A(x)$ satisfies

$$A(x)A(x)^T = -2\langle f_0 r^T \rangle.$$

In order to be able to extract a non-singular matrix root $A(x)$ from $A(x)A(x)^T$ it is necessary to show that the right-hand side, $-2\langle f_0 r \rangle$, is positive definite. Notice that, for all constant vectors a,

$$a^T f_0 r^T a = (a \cdot r)(a \cdot f_0) = (a \cdot r)\mathcal{L}_1(a \cdot r).$$

If, as mentioned above when discussing ergodicity, $\mathcal{L}_1$ is negative definite in the inner product space weighted by the invariant density and excluding constants, we see that

$$a^T(A(x)A(x)^T)a > 0 \qquad \forall a \neq 0$$

and hence the diffusion coefficient is well-defined. Finally,

$$F(x) = \langle -(\nabla_x r) f_0 \rangle + \langle f_1 \rangle = F_0(x) + F_1(x).$$

The explicit extraction of $A(X)$, $F(X)$ may not be possible in general since it requires the inversion of $\mathcal{L}_1$ to find r from the cell problem. Thus we describe an alternative way to define these vector fields, useful for practical estimation procedures. Let $h = h(y)$ be orthogonal to the kernel of $\mathcal{L}_1^*$, i.e. $\langle h \rangle = 0$. If $H = H(y)$ given by

$$H = -\int_0^\infty \mathrm{e}^{\mathcal{L}_1 t} h \, \mathrm{d}t$$

is well-defined, then it is an integral representation of $\mathcal{L}_1^{-1} h$. This is because

$$\mathcal{L}_1 H = -\int_0^\infty \mathcal{L}_1 \mathrm{e}^{\mathcal{L}_1 t} h \, \mathrm{d}t = -\int_0^\infty \frac{\partial}{\partial t}\mathrm{e}^{\mathcal{L}_1 t} h \, \mathrm{d}t = h - \lim_{t\to\infty} \mathrm{e}^{\mathcal{L}_1 t} h = h.$$

Thus

$$-r = -\mathcal{L}_1^{-1} f_0 = \int_0^\infty \mathrm{e}^{\mathcal{L}_1 t} f_0 \, \mathrm{d}t.$$

Substituting this expression into (6.14) gives

$$\begin{aligned} -\langle \mathcal{L}_2\mathcal{L}_1^{-1}\mathcal{L}_2 v_0 \rangle &= \left\langle \int_0^\infty f_0 \mathrm{e}^{\mathcal{L}_1 t} f_0^T \, \mathrm{d}t \right\rangle : \nabla_x(\nabla_x v_0) + \left\langle \int_0^\infty f_0 \nabla_x \{\mathrm{e}^{\mathcal{L}1 t} f_0\} f_0 \, \mathrm{d}t \right\rangle \cdot \nabla_x v_0 \\ &= \frac{1}{2}[A(x)A(x)^T] : \nabla_x(\nabla_x v_0) + F_0(x) \cdot \nabla_x v_0, \end{aligned}$$

where

$$\begin{aligned}\frac{1}{2}[A(x)A(x)^T] &= \left\langle \int_0^\infty \mathbb{E} f_0(x, y) f_0^T(x, y(t))\,\mathrm{d}t \right\rangle, \\ F_0(x) &= \left\langle \int_0^\infty \{\nabla_x f_0(x, y(t))\} f_0(x, y)\,\mathrm{d}t \right\rangle.\end{aligned} \tag{6.15}$$

Here $y(t)$ denotes the solution to the y-equation in (6.3) with $\epsilon = 1$ and initial condition y; $\mathbb{E}$ denotes the expectation with respect to the Brownian trajectories, whereas $\langle\cdot\rangle$ denotes, as before, averaging over $y = y(0)$.

Since, by assumption, the y-dynamics is ergodic, the expectation over the Brownian trajectories and the initial data may be replaced by a single time-average. We now illustrate this fact, informally, freely exchanging the order of integration without justification. For example, the auto-correlation of $y(t)$ reduces as follows:

$$\langle \mathbb{E} y y^T(t)\rangle = \lim_{T\to\infty} \frac{1}{T}\int_0^T \mathbb{E} y(s) y^T(s+t)\,\mathrm{d}s = \mathbb{E} \lim_{T\to\infty} \frac{1}{T}\int_0^T y(s) y^T(s+t)\,\mathrm{d}s,$$

and the expectation $\mathbb{E}$ is now superfluous since the increments of B are stationary. By similar arguments we obtain

$$\begin{aligned}\frac{1}{2}[A(x)A(x)^T] &= \int_0^\infty \left\{ \lim_{T\to\infty} \frac{1}{T}\int_0^T f_0(x, y(s)) f_0^T(x, y(s+t))\mathrm{d}s \right\} \mathrm{d}t, \\ F_0(x) &= \int_0^\infty \left\{ \lim_{T\to\infty} \frac{1}{T}\int_0^T \nabla_x f_0(x, y(s+t)) f_0(x, y(s))\,\mathrm{d}s \right\} \mathrm{d}t.\end{aligned} \tag{6.16}$$

Similarly,

$$\langle \mathcal{L}_3 v_0\rangle = F_1(x)\cdot\nabla_x v_0,$$

where,

$$F_1(x) = \langle f_1(x, y)\rangle = \lim_{T\to\infty} \frac{1}{T}\int_0^T f_1(x, y(t))\,\mathrm{d}t. \tag{6.17}$$

Example 6.1. The Ornstein–Uhlenbeck (OU) process is a linear SDE driven by additive white noise, and with a linear drift term inducing returns to the centre of the phase space. In one dimension it reads

$$\frac{\mathrm{d}y}{\mathrm{d}t} = -y + \frac{\mathrm{d}V}{\mathrm{d}t}, \tag{6.18}$$

where V is standard Brownian motion on $\mathbb{R}$. This is the simplest ergodic SDE on $\mathbb{R}$. It has a Gaussian invariant measure $\mathcal{N}(0, \frac{1}{2})$. (For later developments it is important to note that generalizations to different constants in the coefficients, and to higher dimensions, are also referred to as OU processes.) We consider a simple skew-product example where the fast variable y is a speeded up OU process, leading to the following equations in $\mathcal{X} = \mathcal{Y} = \mathbb{R}$:

$$\begin{aligned}\frac{\mathrm{d}x}{\mathrm{d}t} &= -\lambda x + \frac{1}{\epsilon} y x, \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= -\frac{1}{\epsilon^2} y + \frac{1}{\epsilon}\frac{\mathrm{d}V}{\mathrm{d}t}.\end{aligned} \tag{6.19}$$

Here V is standard Brownian motion on $\mathbb{R}$. We have

$$\begin{aligned}\mathcal{L}_1 v &= -y\frac{\partial v}{\partial y} + \frac{1}{2}\frac{\partial^2 v}{\partial y^2}, \\ \mathcal{L}_2 v &= x y \frac{\partial v}{\partial x}, \\ \mathcal{L}_3 v &= -\lambda x \frac{\partial v}{\partial x}.\end{aligned}$$

From the definition of $\mathcal{L}_1$ it is easily verified that the only density invariant under $\mathcal{L}_1^*$ is $\pi^{-1/2}\exp(-y^2)$, so that the averaging $\langle\cdot\rangle$ is with respect to Gaussian measure $\mathcal{N}(0, \frac{1}{2})$. Now, $\mathcal{L}_1 v_1 = -\mathcal{L}_2 v_0$ reads

$$-y\frac{\partial v_1}{\partial y} + \frac{1}{2}\frac{\partial^2 v_1}{\partial y^2} = -xy\frac{\partial v_0}{\partial x},$$

which has the solution

$$v_1(x, y, t) = -\mathcal{L}_1^{-1}\mathcal{L}_2 v_0 = xy\frac{\partial v_0}{\partial x}.$$

Finally,

$$-\langle \mathcal{L}_2\mathcal{L}_1^{-1}\mathcal{L}_2 v_0\rangle = \left\langle xy\frac{\partial}{\partial x}\left(xy\frac{\partial v_0}{\partial x}\right)\right\rangle = \frac{x}{2}\frac{\partial}{\partial x}\left(x\frac{\partial v_0}{\partial x}\right)$$

and

$$\langle \mathcal{L}_3 v_0\rangle = -\lambda x\frac{\partial v_0}{\partial x},$$

so that (6.12) yields the following equation for $v_0 = v_0(x, t)$:

$$\frac{\partial v_0}{\partial t} = \left(\frac{x}{2} - \lambda x\right)\frac{\partial v_0}{\partial x} + \frac{x^2}{2}\frac{\partial^2 v_0}{\partial x^2}.$$

Comparing with (2.9) we see that this equation arises as the Chapman–Kolmogorov equation of the (effective) Itô SDE

$$\frac{\mathrm{d}X}{\mathrm{d}t} = \left(\frac{1}{2} - \lambda\right)X + X\frac{\mathrm{d}U}{\mathrm{d}t}. \tag{6.20}$$

Recall the Itô formula whereby a function $Y = g(t, U)$ satisfies the SDE:

$$\frac{\mathrm{d}Y}{\mathrm{d}t} = \left(\frac{\partial g}{\partial t} + \frac{1}{2}\frac{\partial^2 g}{\partial U^2}\right) + \frac{\partial g}{\partial U}\frac{\mathrm{d}U}{\mathrm{d}t}.$$

Using this, it is immediately verified (see, for example, [4]) that equation (6.20) has the exact solution

$$X(t) = X(0)\exp[-\lambda t + U(t)].$$

In order to test the theory we compare the behaviour of x against known theoretical properties of X solving the limiting SDE (6.20). From the exact solution $X(t)$ and the properties of Brownian motion (see Mao [53], p 105) it follows that:

$$\begin{aligned}
&\lambda > 0 \quad \Leftrightarrow \quad \lim_{t\to\infty} X(t) = 0 \quad \text{a.s.}\\
&\lambda = 0 \quad \Leftrightarrow \quad \limsup_{t\to\infty} |X(t)| = \infty \qquad \text{and} \qquad \liminf_{t\to\infty} |X(t)| = 0 \quad \text{a.s.}\\
&\lambda < 0 \quad \Leftrightarrow \quad \lim_{t\to\infty} |X(t)| = \infty \quad \text{a.s.}
\end{aligned}$$

In figure 4 we show three trajectories of $\log x(t)$ for $\lambda = -1, 0$ and 1 respectively. The value of ϵ is 0.1 . In figure 5 we repeat this experiment with smaller $\epsilon = 0.01$. Notice the agreement with theoretical predictions from the SDE, although for $\lambda = 0$ the wild oscillations appear to stop at a finite time, rather than persisting indefinitely, and then $x(t)$ dies out, decaying to 0 ($\log x(t)$ tends to $-\infty$). ■

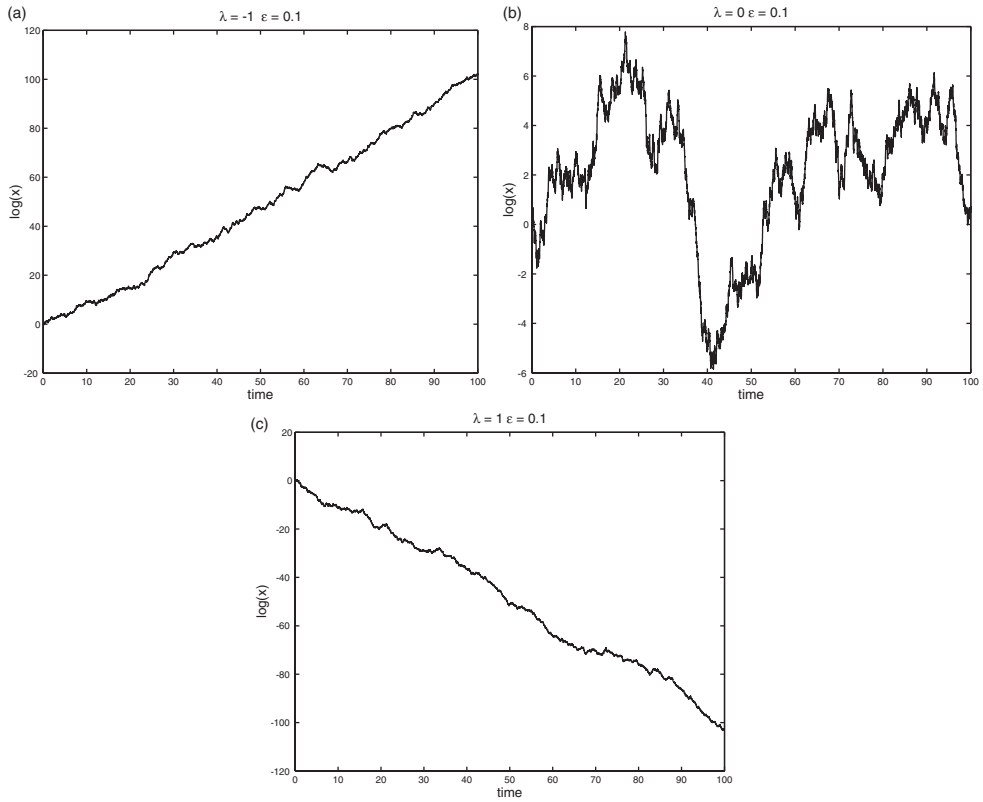

Figure 4. Time evolution of $\log x(t)$ for $\epsilon = 0.1$ and (*a*) $\lambda = -1$, (*b*) $\lambda = 0$ and (*c*) $\lambda = 1$.

6.2. The Fokker–Planck picture

We study the model problem (6.3) in the limit where β_0 is a constant, eventually set to zero. If $\beta_0 \equiv 0$ then the papers [7, 43, 54] show that the basic ideas outlined in section 6.1 still apply formally, provided the dynamics in z is mixing in a sufficiently strong sense[5]. The formulae (6.16), (6.17) are used to define the drift and diffusion terms, rather than (6.15), because there is no longer any expectation over Brownian paths. The formulae show that a decay of correlations is required in order for the outer integration in t to be well-defined.

We find it convenient to illustrate this situation by carrying out the programme of the previous subsection in the Fokker–Planck picture rather than its adjoint, the Chapman–Kolmogorov one. For rigorous proofs the Chapman–Kolmogorov approach is usually preferred because it allows the treatment of deterministic data in a straightforward way; note that deterministic data leads to a Dirac mass as initial data for the Fokker–Planck equation, something which is hard to handle analytically. On the other hand physicists tend to think in terms of the probability density function, and so the Fokker–Planck picture is natural in that context. At a formal level of perturbation expansions there is little to choose between the two approaches. Our approach will be to take $\beta_0 \neq 0$, derive the limiting Fokker–Planck equation in x, and then set $\beta_0 = 0$.

[5] Examples of deterministic dynamics with provably strong mixing properties are few, but include geodesic flow on manifolds with negative curvature [55]. Empirically, however, there are many interesting systems which appear to obey this condition, including the Lorenz equations, for example, and the 'Burgers' bath' of Majda and Timofeyev [56].

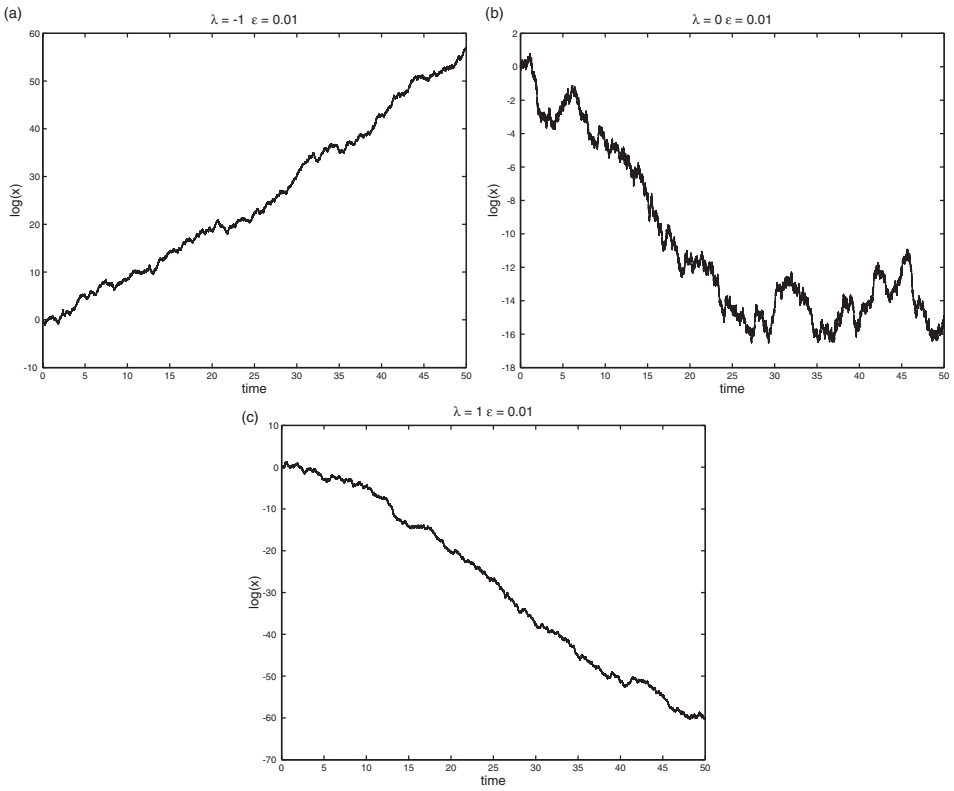

Figure 5. Time evolution of $\log x(t)$ for $\epsilon = 0.01$ and (*a*) $\lambda = -1$, (*b*) $\lambda = 0$ and (*c*) $\lambda = 1$.

With f, g and β still of the form (6.2), the Fokker–Planck equation (2.8) becomes

$$\frac{\partial \rho}{\partial t} = \frac{1}{\epsilon^2}\mathcal{L}_1^*\rho + \frac{1}{\epsilon}\mathcal{L}_2^*\rho + \mathcal{L}_3^*\rho,$$

where

$$\mathcal{L}_1^*\phi = -\nabla_y \cdot (g_0\phi) + \tfrac{1}{2}\nabla_y \cdot [\nabla_y \cdot (\beta_0\beta_0^T\phi)],$$
$$\mathcal{L}_2^*\phi = -\nabla_x \cdot (f_0\phi),$$
$$\mathcal{L}_3^*\phi = -\nabla_x \cdot (f_1\phi).$$

We seek an expansion for ρ in the form

$$\rho = \rho_0 + \epsilon\rho_1 + \epsilon^2\rho_2 + \cdots,$$

substitute it into the Fokker–Planck equation, and equate powers of ϵ to obtain

$$\mathcal{L}_1^*\rho_0 = 0,$$
$$\mathcal{L}_1^*\rho_1 = -\mathcal{L}_2^*\rho_0,$$
$$\mathcal{L}_1^*\rho_2 = \frac{\partial \rho_0}{\partial t} - \mathcal{L}_2^*\rho_1 - \mathcal{L}_3^*\rho_0.$$

We assume that the y-dynamics is ergodic so that $e^{\mathcal{L}_1 t}\phi \to \langle\phi\rangle$ as $t \to \infty$, with $\langle\cdot\rangle$ denoting expectation with respect to an invariant measure, possibly restricted to some sub-manifold, in $\mathcal{Y}$. Thus, $\mathcal{L}_1^*\rho_\infty(y) = 0$ for some density $\rho_\infty(y)$. The solution for $\rho_0(x, y, t)$ is then of the form

$$\rho_0(x, y, t) = \rho_\infty(y)\bar{\rho}(x, t).$$

Thus, to leading-order, the distribution of solutions is a product measure—the x and y components of the solution are independent. The density of the slow variables, $\bar{\rho}(x,t)$, is the quantity of interest.

To solve the equation $\mathcal{L}_1^*\rho = r$ we require that r be orthogonal to the null-space of $\mathcal{L}_1$, i.e. that it integrates to zero against constants (in y). If $\langle f_0 \rangle = 0$, so that the leading-order x-dynamics averages to zero under the invariant measure for y, the equation for ρ_1 is solvable, and

$$\rho_1 = -(\mathcal{L}_1^*)^{-1}\mathcal{L}_2^*\rho_\infty(y)\bar{\rho}(x,t).$$

A similar solvability condition applied to the equation for ρ_2 leads to the following equation for $\bar{\rho}(x,t)$:

$$\frac{\partial \bar{\rho}}{\partial t} = -\int_y \mathcal{L}_2^*(\mathcal{L}_1^*)^{-1}\mathcal{L}_2^*\rho_\infty\bar{\rho}\,\mathrm{d}y + \int_y \mathcal{L}_3^*\rho_\infty\bar{\rho}\,\mathrm{d}y. \tag{6.21}$$

In view of the fact that $\mathcal{L}_2^*$ is a first-order differential operator in x, and the averaging is over y, this is the Fokker–Planck equation for an SDE in $\mathcal{X}$:

$$\frac{\mathrm{d}X}{\mathrm{d}t} = F(X) + A(X)\frac{\mathrm{d}U}{\mathrm{d}t},$$

U being standard Brownian motion. The arguments showing this are similar to those in the previous subsection.

Example 6.2. Consider the equations

$$\begin{aligned}
\frac{\mathrm{d}x}{\mathrm{d}t} &= x - x^3 + \frac{4}{90\epsilon}y_2,\\
\frac{\mathrm{d}y_1}{\mathrm{d}t} &= \frac{10}{\epsilon^2}(y_2 - y_1),\\
\frac{\mathrm{d}y_2}{\mathrm{d}t} &= \frac{1}{\epsilon^2}(28y_1 - y_2 - y_1y_3),\\
\frac{\mathrm{d}y_3}{\mathrm{d}t} &= \frac{1}{\epsilon^2}\left(y_1y_2 - \frac{8}{3}y_3\right).
\end{aligned} \tag{6.22}$$

Note that the vector $y = (y_1, y_2, y_3)^T$ solves the Lorenz equations, at parameter values where the solution is chaotic [24]. Thus the equation for x is a scalar ODE driven by a chaotic signal with characteristic time ϵ^2. We will show how, for small ϵ, the x-dynamics may be approximated by the SDE

$$\frac{\mathrm{d}X}{\mathrm{d}t} = X - X^3 + \sigma\frac{\mathrm{d}W}{\mathrm{d}t}, \tag{6.23}$$

where σ is a constant. Although the asymptotics of the previous subsection cannot be rigorously justified in this case without the addition of a white noise term to the equations for y, we, nonetheless, proceed to find an SDE in the small ϵ limit, showing by means of numerical experiment that the fit between x and X is a good one. We interpret (6.21) by taking $\rho_\infty(y)$ to be the density generated by the empirical measure of the Lorenz equations.

Here $f_1 = f_1(x) = x - x^3$ and $f_0 = f_0(y) = 4y_2/90$. Since $\mathcal{L}_1$ is independent of x we deduce that

$$(\mathcal{L}_1^*)^{-1}\mathcal{L}_2^*\rho_\infty\bar{\rho} = -(\mathcal{L}_1^*)^{-1}\frac{\partial}{\partial x}(f_0\rho_\infty\bar{\rho}) = r\frac{\partial\bar{\rho}}{\partial x},$$

where $r = r(y)$ solves the equation

$$\mathcal{L}_1^* r(y) = -f_0(y)\rho_\infty(y).$$

(It is for this step that the regularization of the y-dynamics, by addition of white noise, is required; otherwise $\mathcal{L}_1^*$ may not have a unique inverse on the appropriate subspace, and $r(y)$ will not be well-defined.) Proceeding with this expression we find that

$$-\int_y \mathcal{L}_2^*(\mathcal{L}_1^*)^{-1}\mathcal{L}_2^*\rho_\infty\bar{\rho}\,\mathrm{d}y = \int_y \frac{\partial}{\partial x}\left(f_0 r\frac{\partial\bar{\rho}}{\partial x}\right)\mathrm{d}y = \frac{\sigma^2}{2}\frac{\partial^2\bar{\rho}}{\partial x^2},$$

where

$$\sigma^2 = \frac{8}{90}\int_y y_2 r(y)\,\mathrm{d}y.$$

Also

$$\int_y \mathcal{L}_3^*\rho_\infty\bar{\rho}\,\mathrm{d}y = -\frac{\partial}{\partial x}[(x-x^3)\bar{\rho}].$$

Thus, the limiting equation for probability densities is

$$\frac{\partial\bar{\rho}}{\partial t} + \frac{\partial}{\partial x}[(x-x^3)\bar{\rho}] = \frac{\sigma^2}{2}\frac{\partial^2\bar{\rho}}{\partial x^2},$$

which is the Fokker–Planck equation for the SDE (6.23).

However, we do not know $r(y)$ explicitly (indeed it is only well-defined if we add noise to the Lorenz equations) and thus do not know σ explicitly. To circumvent this difficulty, we estimate σ from a sample path of $x(t)$, calculated with a small (compared to ϵ^2) time-step Δt. We study the time-series γ_n defined by

$$\gamma_n = h^{-1/2}\{x_{n+1} - x_n - h[x_n - x_{n+1}x_n^2]\},$$

for $x_n = x(nh)$ and h small (typically chosen as some multiple of Δt so that interpolation of numerically generated data is not necessary). If x were governed by the SDE (6.23) then γ_n should be an approximately i.i.d. sequence distributed as $\mathcal{N}(0,\sigma^2)$ and this fact can be used to estimate σ [6].

Figure 6 shows the estimate of σ^2 calculated from this data, using $\epsilon = \Delta t = 10^{-3}$. The left figure shows the dependence of the estimate on the time interval for $h = 0.05$; notice that the estimate converges very fast in time. The right figure shows how this estimate varies with the sampling interval h. For $h \in [0.05, 0.4]$ we obtain $\sigma^2 = 0.126 \pm 0.003$.

To verify that the fit with the SDE at the predicted value of σ is a good one, we compare the empirical density of the data in figure 7, generated from $x(t)$ over a time interval of length 10^4, with the exact invariant measure for the SDE (6.23), at the estimated value of σ. The agreement is very good. ■

6.3. Discussion and bibliography

The basic perturbation expansion outlined in the Chapman–Kolmogorov case can be rigorously justified and weak convergence of x to X proved as $\epsilon \to 0$; see Kurtz [51]. The perturbation expansion which underlies the approach is clearly exposed in [7]. Applications to climate models, where the atmosphere evolves quickly relative to the slow oceanic variations, are surveyed in Majda *et al* [52]; we have followed the presentation in [7, 52] quite closely here. Further applications to the atmospheric sciences may be found in [58, 59].

[6] This method of parameter estimation for the stochastic model is quite general but does not exploit the scale-separation to optimize the computational work. Other methods could be used which do exploit scale-separation, such as the method introduced in [57], which is based on (6.16) and described in section 10.

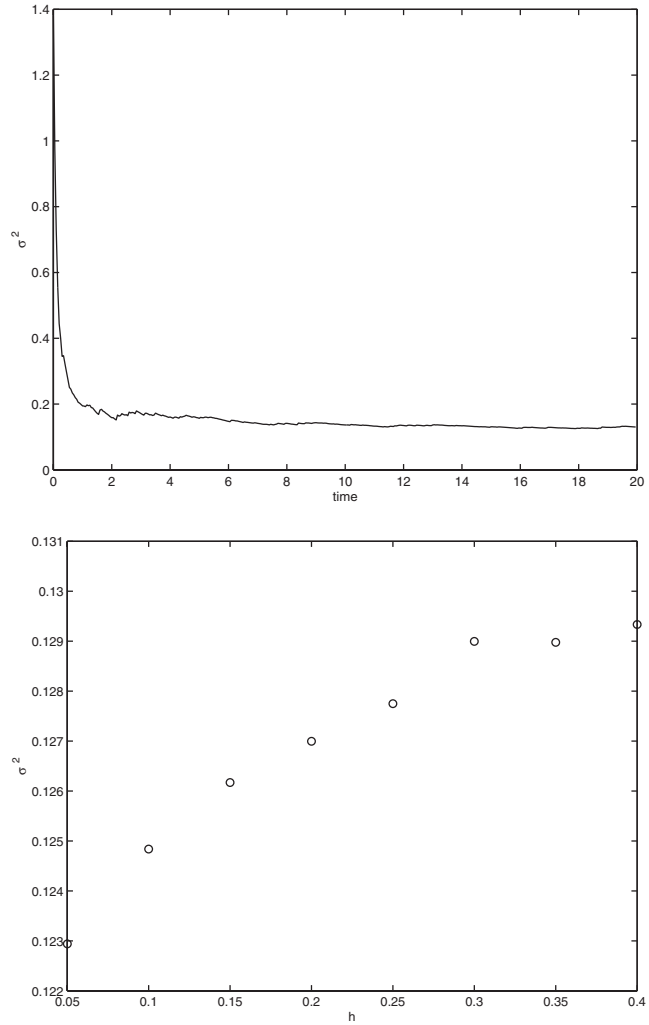

Figure 6. Top: estimated value of σ as function of the size of the time interval for $\epsilon^2 = 0.001$ and $h = 0.05$. Bottom: estimated value of σ as function of the sampling interval h.

The use of representation (6.15) is discussed in [7]. The representations (6.16) and (6.17) for the effective drift and diffusion can be used in the design of coarse time-stepping algorithms (see Vanden-Eijnden [57] and section 10.3).

Studying the derivation of effective stochastic models when the variables being eliminated do not necessarily come from an Itô SDE, as we did in the Fokker–Planck picture, is a subject investigated in some generality in [54]. The idea outlined here is carried out in discrete time by Beck [60] who also uses a skew-product structure to enable the analysis; the ideas can then be rigorously justified in some cases. In the paper [61] the idea that fast chaotic motion can introduce noise in slow variables is pursued for an interesting physically motivated problem where the fast chaotic behaviour arises from the Burgers' bath of [56]. Furthermore, by means of numerical experiments, a connection is made between driving the system by the chaotic, but deterministic, Burgers' bath and a fully stochastic model where the bath is represented by an OU process (a generalization of (6.18)).

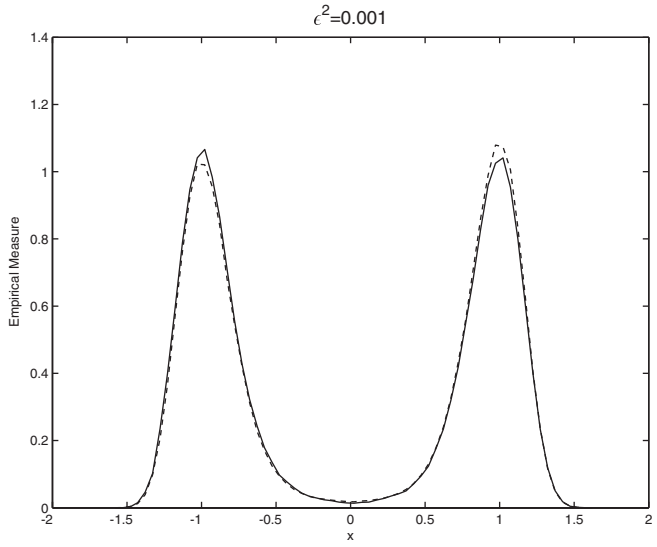

Figure 7. Empirical measure of $x(t)$ solving (6.22) for $\epsilon^2 = 0.001$ (——) compared with the empirical measure of $X(t)$ solving (6.23) ($\cdots\cdots$).

Related work can be found in [62] and similar ideas in continuous time are addressed in [17, 18] for differential equations; however, rather than developing a systematic expansion in powers of ϵ, they find the exact solution of the Fokker–Planck equation, projected into the space $\mathcal{X}$, by use of the Mori–Zwanzig formalism (see section 3) [19], and then make power series expansions in ϵ of the resulting problem.

There are many variants on the basic themes introduced in the previous two sections. Here, we briefly discuss two of them. The first is *fast deterministic dynamics*. The set-up is as in section 6.1, but we do not assume that the vector field $f_0(x, y)$ averages to zero under $\langle\cdot\rangle$ and, as a consequence, there is additional fast dynamics not present in section 6.1. Consequently we introduce a new time variable

$$s = \epsilon^{-1}t$$

and seek a two-time-scale expansion of the Chapman–Kolmogorov equation, setting

$$\frac{\partial}{\partial t} \to \frac{\partial}{\partial t} + \frac{1}{\epsilon}\frac{\partial}{\partial s}.$$

Having performed this expansion and converting back from the Chapman–Kolmogorov picture, combining to give one time variable yields

$$\frac{\mathrm{d}X}{\mathrm{d}t} = \frac{1}{\epsilon}F_0(X) + F_1(X) + A(X)\frac{\mathrm{d}U}{\mathrm{d}t},$$

U being standard Brownian motion.

In the Fokker–Planck picture we are seeking an approximation of the form

$$\rho(x, y, t) \approx \rho_\infty(y)\bar{\rho}(x, t, s).$$

This situation and more general, related ones are covered in a series of papers by Papanicolaou and co-workers—see [7, 43, 54, 63], building on the original work of Khasminkii [64, 65]. See also [17, 18, 52, 60].

The second generalization is to include *back-coupling*. We again consider a set-up similar to section 6, but now allow back-coupling of the x-variable into the equation for y. We consider (1.2) with $\alpha = 0$

$$f(x, y) = \frac{1}{\epsilon} f_0(x, y), \qquad g(x, y) = \frac{1}{\epsilon^2} g_0(x, y), \qquad \beta(x, y) = \frac{1}{\epsilon} \beta_0(x, y).$$

The equation for y is thus, in place of (6.3),

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \frac{1}{\epsilon^2} g_0(x, y) + \frac{1}{\epsilon} \beta_0(x, y) \frac{\mathrm{d}V}{\mathrm{d}t}.$$

Since x evolves more slowly than y it is natural to study the equation

$$\frac{\mathrm{d}Y}{\mathrm{d}t} = g_0(\zeta, y) + \beta_0(\zeta, y) \frac{\mathrm{d}V}{\mathrm{d}t}, \tag{6.24}$$

where ζ is viewed as a parameter. If this equation is ergodic, for each fixed ζ, with invariant measure μ_ζ, then it is natural to try and generalize the studies of the previous sections, replacing $\langle \cdot \rangle$ by averaging with respect to μ_x, since the slower time-scale of x relative to y means that it will be effectively frozen in the y-dynamics. In the Fokker–Planck picture we are seeking a solution of the form (6.1), where $\rho_\infty(y; \zeta)$ is the invariant density for (6.24). Such ideas can be developed systematically; see [17, 18, 43, 52, 54, 63–65] for details. An approximation of the form (6.1) also underlies the averaging techniques of section 5.

7. White and coloured noise approximations of large systems

In the previous section we showed how effective low-dimensional SDEs can arise from either higher-dimensional SDEs or from ODEs, when a separation of time-scales occurs. We worked with the Chapman–Kolmogorov or Fokker–Planck equations, rather than the paths of (1.1) itself. In this section we describe an alternative situation where effective low-dimensional SDEs can arise. This is achieved by coupling a small problem weakly to a heat bath, a large Hamiltonian system. We are studying problems of the form D–S in the classification of section 1.

Here we will study the system from a pathwise perspective, rather than using the Chapman–Kolmogorov or Fokker–Planck picture. However, it is of interest to give an interpretation for probability densities. The systems we study are large ODEs with random data. We study situations where the Liouville equation (the hyperbolic PDE (2.8) with $\Gamma = 0$) for $\rho(x, y, t)$ can be approximated by

$$\rho(x, y, t) \approx \rho(y; x) \bar{\rho}(x, t)$$

and $\bar{\rho}(x, t)$ satisfies a Fokker–Planck (parabolic) PDE (2.8). The analysis in this section works in a variety of situations, but in all cases the dimension of the space $\mathcal{Y}$ is tending to infinity.

The basic building block for the analysis is the trigonometric approximation of Gaussian processes, which we describe in the next subsection. In the two subsequent sections we study a skew-product system and then the more physically interesting case of a Hamiltonian system for a particle interacting with a heat bath.

7.1. Trigonometric approximation of Gaussian processes

Mean zero Gaussian processes $Z(t)$ have the property that given any sequence of times $t_1, t_2, \ldots, t_k$, the vector

$$(Z(t_1), Z(t_2), \ldots, Z(t_k))$$

is a mean zero Gaussian random vector in $\mathbb{R}^k$. It is *stationary* if the statistics are unchanged when the $\{t_i\}_{i=1}^k$ are all translated by a single time s. Subject to some continuity properties on paths (see, e.g. Karlin and Taylor [66]) a mean zero stationary Gaussian process is completely characterized by its auto-covariance function

$$R(\tau) := \mathbb{E}Z(t+\tau)Z(t).$$

The basic building block in this section is the trigonometric series for Gaussian processes (see Kahane [67]). We consider the approximation of Gaussian processes by a finite series of the form

$$Z_N(t) = \frac{1}{N^b}\sum_{j=1}^{N} F(\omega_j)[\xi_j \cos(\omega_j t) + \eta_j \sin(\omega_j t)], \tag{7.1}$$

where the ξ_j and η_j are mutually independent i.i.d. sequences with $\xi_1, \eta_1 \sim \mathcal{N}(0,1)$. The sequence of frequencies ω_j may or may not be random. The parameter b will be chosen differently depending on the choice of frequencies $\{\omega_j\}$; see below for details. The process (7.1) is Gaussian, once the frequencies are specified. Letting $\mathbb{E}$ denote expectation with respect to ξ_j and η_j, with the ω_j fixed, we see that

$$R_N(\tau) := \mathbb{E}\, Z_N(t+\tau)Z_N(t)$$

is given by

$$R_N(\tau) = \frac{1}{N^{2b}}\sum_{j=1}^{N} F^2(\omega_j)\cos(\omega_j \tau). \tag{7.2}$$

Notice that the oscillators $\xi_j \cos(\omega_j t)$ and $\eta_j \sin(\omega_j t)$ arise as solutions to an ODE with random initial data. The basic idea underlying the constructions of SDEs from ODEs in this section is to exploit this fact and to choose the function $F(\omega)$ and the sequence of frequencies ω_j so that $R_N(\tau)$ approximates $R(\tau)$ for large N, thus building an approximation $Z_N(t)$ of the stationary Gaussian process $Z(t)$ from solutions of ODEs. This idea is made more precise in the following subsections.

7.2. Skew-product systems

Here, we consider model problems with the form

$$\frac{\mathrm{d}x}{\mathrm{d}t} = f(x) + \sum_{j=1}^{N} k_j q_j, \tag{7.3}$$

$$m_j \frac{\mathrm{d}^2}{\mathrm{d}t^2} q_j + q_j = 0, \qquad j = 1, 2, \ldots, N, \tag{7.4}$$

where $m_j = \omega_j^{-2}$ and the k_j are constants to be determined.

To put this in the general framework of section 1 we set $y = (q, \mathrm{d}q/\mathrm{d}t)$, and $z = (x, y) = (x, q, \mathrm{d}q/\mathrm{d}t)$, where $q = (q_1, q_2, \ldots, q_N)^T$. The problem is in *skew-product* form: the y-dynamics evolves independently of the x-dynamics. Full coupling is considered in the next subsection.

We note that the q-equations derive from the Hamiltonian

$$H(p,q) = \frac{1}{2}\sum_{j=1}^{N} \frac{p_j^2}{m_j} + \frac{1}{2}\sum_{j=1}^{N} q_j^2.$$

Here $p_j = m_j\,(\mathrm{d}q/\mathrm{d}t)$. The functions $q_j(t)$ may be viewed as the trajectories of N independent harmonic oscillators with mass m_j, spring constant 1 and natural frequencies $\omega_j = m_j^{-1/2}$. Together, the N oscillators constitute a 'heat bath'. If we choose initial data for this heat bath from the Gibbs distribution at inverse temperature β, that is, we pick from density proportional to $\mathrm{e}^{-\beta H(p,q)}$ then

$$q(0) \sim \beta^{-1/2}\xi_j, \qquad \dot{q}(0) \sim \beta^{-1/2}\omega_j\,\eta_j,$$

where the random variables ξ_j and η_j are, as above, mutually independent sequences of i.i.d. $\mathcal{N}(0, 1)$ random variables.

To establish a connection with the previous subsection we choose the coupling constants k_j so that

$$k_j = \frac{F(\omega_j)}{N^b}.$$

Then,

$$\sum_{j=1}^{N} k_j q_j = \beta^{-1/2}\,Z_N(t),$$

where $Z_N(t)$ is given by (7.1). Thus the 'essential dynamics', $x(t)$, satisfy the randomly-driven ODE:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = f(x) + \beta^{-1/2}\,Z_N(t). \tag{7.5}$$

Example 7.1. We start with an example where $Z_N(t)$ approximates a coloured noise process. Choose $a \in (0, 1)$, $2b = 1 - a$ and $\omega_j = N^a\zeta_j$, where $\zeta := \{\zeta_j\}_{j=1}^{\infty}$ is an i.i.d. sequence with ζ_1 uniformly distributed in $[0, 1]$, $\zeta_1 \sim \mathcal{U}[0, 1]$. Defining $\Delta\omega = N^a/N$, which is the mean frequency spacing, (7.2) takes the form

$$R_N(t) = \sum_{j=1}^{N} F^2(\omega_j)\cos(\omega_j t)\,\Delta\omega,$$

which, as $N \to \infty$, is a Monte Carlo approximation to the Fourier-cosine transform of $F^2(\omega)$:

$$R(t) = \int_0^{\infty} F^2(\omega)\cos(\omega t)\,\mathrm{d}\omega.$$

If $F^2(\omega)$ is bounded and decays at least as fast as $1/\omega^{1+\delta}$, for some $\delta > 0$, then for almost every ζ, $R_N(t)$ converges to $R(t)$ point-wise and in $L^1[0, T]$, $T > 0$ arbitrary. The random forcing, (7.1), which takes the form

$$Z_N(t) = \sum_{j=1}^{N} F(\omega_j)[\xi_j\cos(\omega_j t) + \eta_j\sin(\omega_j t)](\Delta\omega)^{1/2},$$

then converges *weakly* (with respect to the probability space for $\{\xi_j\}$, $\{\eta_j\}$) in $C([0, T], \mathbb{R})$ to a zero mean Gaussian process with auto-covariance $R(t)$ as $N \to \infty$ (see [68] for details; see [69] for a general reference on weak convergence).

In particular, if

$$F^2(\omega) = \frac{2\alpha/\pi}{\alpha^2 + \omega^2},$$

where $\alpha > 0$ is a constant, then

$$R(\tau) = \exp(-\alpha|\tau|),$$

and $Z(t)$ is an OU process, like (6.18), defined by an Itô SDE. Finally, it can be shown that (7.5) defines a continuous mapping, $Z_N \mapsto x$, between $C([0,T],\mathbb{R})$ functions. Since weak convergence is preserved under continuous mappings (see [69]) it follows that $x(t)$ is approximated, for N large, by $X(t)$ solving the SDE

$$\begin{aligned} \frac{\mathrm{d}X}{\mathrm{d}t} &= f(X) + \beta^{-1/2} Z(t), \\ \frac{\mathrm{d}Z}{\mathrm{d}t} &= -\alpha Z + (2\alpha)^{1/2} \frac{\mathrm{d}B}{\mathrm{d}t}, \end{aligned} \tag{7.6}$$

where $B(t)$ is standard Brownian motion and $Z(t)$ is an OU process. This approximation holds for any $T > 0$. ■

Example 7.2. In this second example, $Z_N(t)$ approximates white noise, which may be viewed as a delta-correlated Gaussian (generalized) process. We set $b = 0$ and the ω_j are chosen deterministically:

$$\omega_j = 2(j-1), \qquad j = 1, 2, \ldots, N;$$

the $F(\omega_j)$ are given by

$$F(\omega_j) = \begin{cases} \left(\dfrac{1}{\pi}\right)^{1/2} & j = 1, \\ \left(\dfrac{2}{\pi}\right)^{1/2} & j = 2, \ldots, N. \end{cases}$$

This choice makes $R_N(t)$ a truncation of the formal Fourier series for a delta function. To exploit this fact rigorously it is necessary to work with the integral of $Z_N(t)$ which we will call $Y_N(t)$, normalizing by $Y_N(0) = 0$. The function $Y_N(t)$ converges almost surely, as $N \to \infty$, to a function in $C([-\pi/2, \pi/2], \mathbb{R})$ which may be identified with a realization of Brownian motion [70]. Thus, for large N, x is approximated by the SDE:

$$\frac{\mathrm{d}X}{\mathrm{d}t} = f(X) + \beta^{-1/2} \frac{\mathrm{d}U}{\mathrm{d}t}.$$

Here U is standard Brownian motion. The mapping $Y_N \to x$ is continuous and hence x converges strongly to X as $N \to \infty$ and error estimates can be found [71]. However, the convergence is only on a finite time interval $t \in [0, T]$, $T < \pi/2$, because of the periodicity inherent in the construction. ■

7.3. Hamiltonian systems

We now generalize the ideas developed in the last subsection to situations with back-coupling between the x and y variables so that the simplifying skew-product nature is lost. This leads to a class of model problems of clear physical significance.

We consider a mechanical system, which consists of a 'distinguished' particle which moves in a one-dimensional potential field, and experiences, in addition, interactions with a large collection of 'heat bath' particles. The goal is to derive a reduced equation for the distinguished particle, under the assumption that the initial data for the heat bath are random. Models of this type were first introduced in the 1960s by Kac and co-workers [72, 73] and by Zwanzig [9, 12]. The results reported here can be found, in full detail, in [68, 74].

The model problems we consider are defined by the following Hamiltonian,

$$H(P_N, Q_N, p, q) = \frac{1}{2} P_N^2 + V(Q_N) + \frac{1}{2} \sum_{j=1}^{N} \frac{p_j^2}{m_j} + \frac{1}{2} \sum_{j=1}^{N} k_j (q_j - Q_N)^2, \tag{7.7}$$

where Q_N, P_N are the coordinate and momentum of the distinguished particle, and q, p are, as before, vectors whose entries are the coordinates and momenta of the heat bath particles. The function $V(Q)$ is the potential field experienced by the distinguished particle; the jth heat bath particle has mass m_j and interacts with the distinguished particle by means of a linear spring with stiffness constant k_j; the jth heat bath particle has a characteristic frequency $\omega_j = (k_j/m_j)^{1/2}$. The subscript N in Q_N, P_N denotes the size of the heat bath as we will be considering systems of increasing size.

Hamilton's equations are

$$\ddot{Q}_N + V'(Q_N) = \sum_{j=1}^{N} k_j(q_j - Q_N),$$
$$\ddot{q}_j + \omega_j^2(q_j - Q_N) = 0$$

with initial conditions $Q_N(0) = Q_0$, $P_N(0) = P_0$, $q_j(0) = q_j^0$ and $p_j(0) = p_j^0$. The initial data Q_0 and P_0 are given, whereas the q_j^0 and p_j^0 are randomly drawn from a Gibbs distribution with inverse temperature β, i.e. from a distribution with density proportional to $\exp(-\beta H)$, conditional on knowledge of Q_0, P_0. It may be verified that this amounts to choosing

$$q_j^0 = Q_0 + \left(\frac{1}{\beta k_j}\right)^{1/2} \xi_j,$$
$$p_j^0 = \left(\frac{m_j}{\beta}\right)^{1/2} \eta_j,$$

where the sequences ξ_j, η_j are independent i.i.d. sequences as specified in the previous subsection.

The equations for q_j can be solved in terms of the past history of Q_N, and the q_j can then be substituted back into the equation for Q_N. This yields the following integro-differential equation for Q_N:

$$\ddot{Q}_N + V'(Q_N) + \int_0^t R_N(t-s)\dot{Q}_N(s)\,\mathrm{d}s = \beta^{-1/2} Z_N(t), \tag{7.8}$$

where

$$R_N(t) = \sum_{j=1}^{N} k_j \cos(\omega_j t)$$

and

$$Z_N(t) = \sum_{j=1}^{N} k_j^{1/2}[\xi_j \cos(\omega_j t) + \eta_j \sin(\omega_j t)].$$

Notice that, for given frequencies $\{\omega_j\}$, and taking expectations with respect to $\{\xi_j\}$ and $\{\eta_j\}$,

$$\mathbb{E} Z_N(t+\tau) Z_N(t) = R_N(\tau).$$

This is known as the *fluctuation–dissipation relation*—discussed in the general context of the Mori–Zwanzig reduction in section 3.

Equation (7.8) is an instance of a *generalized Langevin equation*, with memory kernel R_N and random forcing $\beta^{-1/2} Z_N$. By choosing the parameters k_j, ω_j to the different, limiting behaviours can be obtained as $N \to \infty$.

Example 7.3. We set $\omega_j = N^a \zeta_j$, with i.i.d. ζ_j and $\zeta_1 \sim \mathcal{U}(0, 1]$, $0 < a < 1$, $\Delta\omega = N^a/N$ and

$$k_j = F^2(\omega_j)\,\Delta\omega = \frac{2\alpha^2\gamma/\pi}{\alpha^2+\omega^2}\frac{N^a}{N}. \tag{7.9}$$

If $\gamma = \alpha^{-1}$ then the functions R_N, Z_N coincide with those in example 7.1. Thus R_N converges to $R(t) = \mathrm{e}^{-\alpha|t|}$, and Z_N weakly converges on any bounded interval to the OU process $Z(t)$ in (7.6). It can further be shown, using a continuity argument, that Q_N weakly converges to the stochastic process $Q(t)$ solving the stochastic IDE:

$$\ddot{Q} + V'(Q) + \int_0^t R(t-s)\dot{Q}(s)\,\mathrm{d}s = \beta^{-1/2}Z(t). \tag{7.10}$$

Moreover, Q solving (7.10) is equivalent to Q solving the SDE

$$\begin{aligned}
\frac{\mathrm{d}Q}{\mathrm{d}t} &= P,\\
\frac{\mathrm{d}P}{\mathrm{d}t} &= S - V'(Q),\\
\frac{\mathrm{d}S}{\mathrm{d}t} &= (-\alpha S - P) + \left(\frac{2\alpha}{\beta}\right)^{1/2}\frac{\mathrm{d}B}{\mathrm{d}t}
\end{aligned} \tag{7.11}$$

where $B(t)$ is standard Brownian motion. Thus, a Hamiltonian system with $2(N+1)$ variables has been reduced to an SDE for the distinguished particle with one auxiliary variable, $S(t)$, which embodies, for large N, the memory effects. Convergence of Q_N to Q can be proved in $C^2([0, T], \mathbb{R})$ for any $T > 0$ [68].

In figure 8, we show empirical distribution of $Q_N(t)$, $N = 5000$, for sample paths over a time interval of $T = 50\,000$ (open circles). The two graphs correspond to the cases of single-well, and double-well potential $V(Q)$. In each case, the solid line is the Boltzmann distribution, proportional to $\exp(-\beta V(Q))$, which is the empirical measure for the ergodic SDE (7.11) [75]. ■

Example 7.4. By a slight modification of the arguments in [74], a limit to a memoryless Langevin equation of the form

$$\begin{aligned}
\frac{\mathrm{d}Q}{\mathrm{d}t} &= P,\\
\frac{\mathrm{d}P}{\mathrm{d}t} &= -V'(Q) - \gamma P + \left(\frac{2\gamma}{\beta}\right)^{1/2}\frac{\mathrm{d}B}{\mathrm{d}t}
\end{aligned} \tag{7.12}$$

can be obtained. One way of doing this is to note that, if we do not take $\gamma = \alpha^{-1}$, then the limiting SDE becomes

$$\begin{aligned}
\frac{\mathrm{d}Q}{\mathrm{d}t} &= P,\\
\frac{\mathrm{d}P}{\mathrm{d}t} &= S - V'(Q),\\
\frac{\mathrm{d}S}{\mathrm{d}t} &= (-\alpha S - \alpha\gamma P) + \left(\frac{2\alpha^2\gamma}{\beta}\right)^{1/2}\frac{\mathrm{d}B}{\mathrm{d}t}.
\end{aligned} \tag{7.13}$$

Taking the limit $\alpha \to \infty$ then gives the desired memoryless Langevin equation.

Another way to obtain the same limit is by use of Fourier series, taking $\omega_j = (2j-1)$, and choosing the $k_j = F^2(\omega_j)$ appropriately so that R_N approximates a delta function, as in

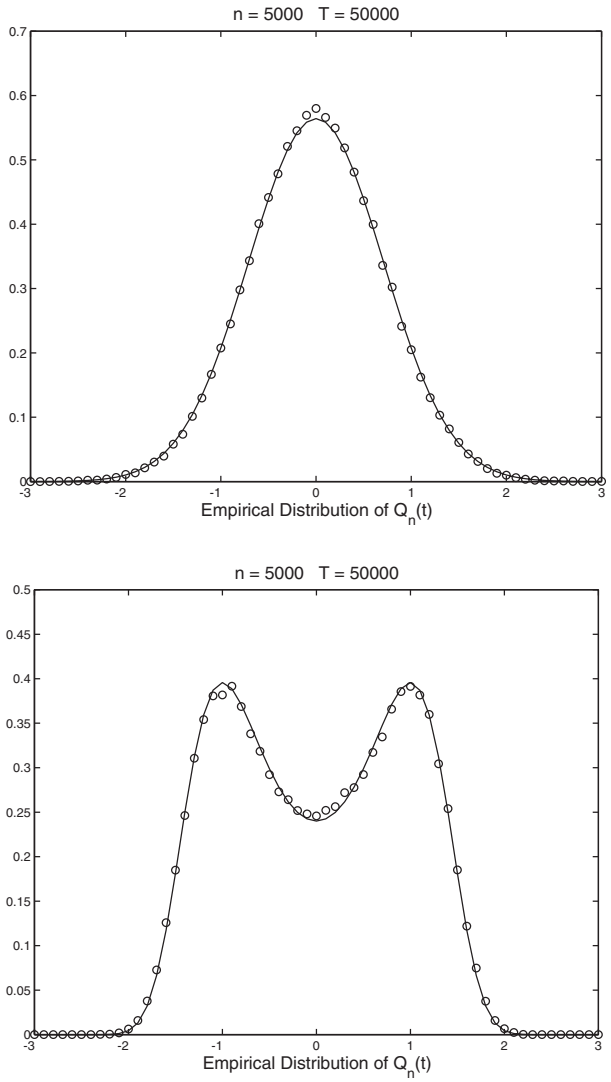

Figure 8. ○: empirical distribution of $Q_N(t)$ for a sample path of the Hamiltonian system in example 7.3, with parameters $n = 5000$, $\alpha = 1$, $\beta = 2$ and a sampling time of $T = 50\,000$. The solid line corresponds to the Boltzmann distribution. The graph on the top is for a single-well potential, $V(Q) = Q^2/2$; the graph on the bottom is for a double-well potential, $V(Q) = Q^4/4 - Q^2/2$.

example 7.2. As mentioned in that context, the use of the Fourier series means that long time behaviour cannot be studied without seeing the periodicity of R_N and Z_N; convergence to (7.12) only occurs on $t \in [0, T]$ with $T < \pi/2$. One way to circumvent this is to re-randomize the data in $\{p, q\}$ periodically in time which is done in [76]. ■

7.4. Discussion and bibliography

Ford *et al* [73] were the first to study mechanical models of masses and springs as models of a particle interacting with a heat bath; see also Ford and Kac [72] and Zwanzig [9, 12].

There exists a substantial amount of literature on the subject in both classical and quantum contexts (see [72,77–79]). The more recent work on heat bath models focuses on those aspects related to dimension reduction and coarse-grained time-stepping. In [74] the spectrum of the heat bath was chosen such that the frequencies are rationally related and the trajectories of the distinguished particle converge in a strong sense to Brownian motion. A drawback of this approach is its restriction to a fixed time interval since the approximate trajectories are periodic. Example 7.3 was studied in [68]. The main results concern the weak convergence of $Q_N(t)$ on bounded time intervals of arbitary size, overcoming the periodicity by use of random frequencies. Some convergence results, and in particular, some ergodic properties, were also proved for infinite time intervals. A generalization of these results for the case of heat bath interactions by means of nonlinear springs may be found in [80]. This paper contains, in addition, a systematic evaluation of the effective models by means of time-series analyses for the distinguished particle trajectories $Q_N(t)$, with N large. Aspects related to coarse-grained integration are studied in [74] and in [81]. A heat bath model that induces fractional Brownian motion is studied in [82].

8. Birth–death processes

Here we present a class of model problems of the form S–D, in the classification of section 1. Chemical reactions are often modelled through birth–death processes, counting molecular concentrations of different chemical species, with transition rates being proportional to algebraic expressions related to species concentrations [3]. In this section we start with a very simple model problem for a single variable species. We show how, in a certain limit, a closed ODE for the first moment describes the dynamics completely; formal expansions and numerical experiments are used to study this example. We follow this with a more involved model problem. In the space of probability measures we are showing that a master equation for a Markov chain in a very large or countable state space can be approximated by a Liouville equation (hyperbolic PDE (2.8) with $\Gamma = 0$) for a single moment; this is done by making the state space continuous and infinite (uncountable). One way to see that we are performing a form of dimension reduction is the following. We find a single closed equation for the first moment, namely the ODE whose characteristics satisfy the adjoint of the Liouville equation; this completely charcterizes the process. In the general case, a coupled infinite hierarchy for all the moments is required.

8.1. One variable species

Here we consider model problems for a single chemical species X, governed by the chemical reaction $X \rightleftharpoons_{k_1}^{k_2} A$, assuming the species A to be held at a fixed concentration a. Let $p_{ij}(t)$ denote the probability that at time t there are j particles of species X, given that there were i at time zero. The master equation (2.2) is then

$$\frac{\mathrm{d}p_{ij}}{\mathrm{d}t} = k_1 a p_{ij-1} + k_2(j+1)p_{ij+1} - (k_1 a + k_2 j)p_{ij}, \qquad p_{ij}(0) = \delta_{ij} \tag{8.1}$$

for $j = 1, 2, \ldots$, and

$$\frac{\mathrm{d}p_{i0}}{\mathrm{d}t} = k_2 p_{i1} - k_1 a p_{i0}, \qquad p_{i0}(0) = \delta_{i0} \tag{8.2}$$

for $j = 0$; see [3].

Recall that for SDEs we have a direct connection between the Chapman–Kolmogorov equation (master equation) and a pathwise description through SDEs. In this section the pathwise description is simply stated by defining an algorithm. Sample paths $z(t)$, $t \geqslant 0$, whose master equation (8.1) can be generated as follows. Assume that we are given $z(t_l)$, the number of particles at time $t_l \geqslant 0$. Let T, S be independent exponential random variables with rates $k_1 a$, $k_2 z(t_l)$, respectively. Define

$$t_{l+1} - t_l = \min(T, S).$$

We then set

$$z(t_{l+1}) = \begin{cases} z(t_l) + 1 & \text{if } T = \min(T, S), \\ z(t_l) - 1 & \text{if } S = \min(T, S). \end{cases} \tag{8.3}$$

This gives a process whose master equation is (8.1). Suitably modified, when $z(t_l) = 0$, it gives (8.2). This is an implementation of a birth–death process, a basic continuous-time countable state space Markov chain. Such a process can be written in the form of (1.1) with $\gamma(z)\,\mathrm{d}W/\mathrm{d}t$ replaced by a Poisson counting process $\mathrm{d}W(z, t)/\mathrm{d}t$.

In such systems, a full, 'detailed' description of the system includes the evolution of the infinite-dimensional vector of probabilities $p_{ij}(t)$, i fixed. Below, we show how in the limit of large numbers, a reduced description may be derived for the first moment of the distribution,

$$X(t) = \frac{1}{a} \sum_{j=0}^{\infty} j p_{ij}(t).$$

Example 8.1. Consider (8.1) with $k_1 = 2$, $k_2 = 1$ and $a = N$. Figure 9 shows three sample paths of this example, calculated with initial data $z(0) = N$. The different paths correspond to $N = 100$, 500 and 1000. In each case we plot $y(t) = N^{-1} z(t)$, together with a smooth curve which is the function $2 - \exp(-t)$, for reasons we now make apparent.

With $k_1 = 2$, $k_2 = 1$ and $a = N$ and $\rho_j = p_{Nj}$ equation (8.1) gives

$$\frac{\mathrm{d}\rho_j}{\mathrm{d}t} = 2N\rho_{j-1} + (j+1)\rho_{j+1} - (2N + j)\rho_j, \qquad \rho_j(0) = \delta_{Nj}.$$

This may be rewritten as, for $\Delta x = N^{-1}$,

$$\frac{\mathrm{d}\rho_j}{\mathrm{d}t} = \frac{(j+1)\Delta x \rho_{j+1} - j\Delta x \rho_j}{\Delta x} + 2\,\frac{\rho_{j-1} - \rho_j}{\Delta x}, \qquad \rho_j(0) = \delta_{Nj}. \tag{8.4}$$

Viewing $\rho_j(t)$ as a finite difference approximation of a continuous density $\rho(x, t)$, so that $\rho_j(t) \approx \rho(j\Delta x, t)$, we see that (8.4) formally approximates the PDE

$$\frac{\partial \rho}{\partial t} = \frac{\partial}{\partial x}(x\rho) - 2\frac{\partial \rho}{\partial x} \tag{8.5}$$

(such approximation is known as a Kramers–Moyal expansion [3]). But (8.5) is simply the Fokker–Planck equation for the ODE

$$\frac{\mathrm{d}X}{\mathrm{d}t} = 2 - X, \qquad X(0) = 1. \tag{8.6}$$

Equation (8.6) may be viewed as a closed equation approximating the first moment of the process. Since the limit dynamics is deterministic, because no diffusion is present in (8.5), all other moments are determined by the first one.

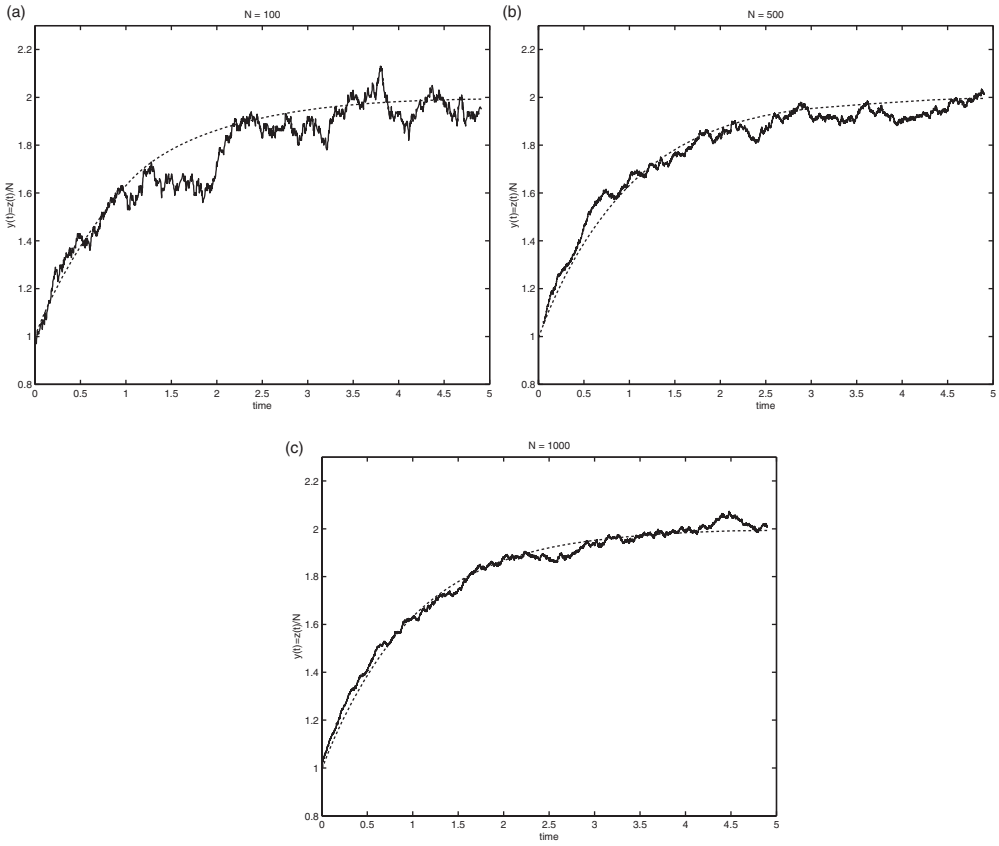

Figure 9. ——: sample paths of $z(t)/N$ for the birth–death process (8.3) with $z(0) = N$, $k_1 = 2$, $k_2 = 1$ and $a = N$ for (*a*) $N = 100$, (*b*) $N = 500$ and (*c*) $N = 1000$. - - - -: the curve $2 - \exp(-t)$.

This formal argument indicates that, as $N \to \infty$, the fluctuations in sample paths of (8.1) should diminish, following the deterministic dynamics given by (8.6), that is

$$X(t) = 2 - \exp(-t).$$

This is exactly what figure 9 shows. ■

8.2. Multiple variables species

A generalization of the previous model problems to multiple species and reactions is as follows (see Gillespie [83, 84]). As in the previous example it is simplest to write down a pathwise description through an algorithm. Let $x(t) = (x_1, \ldots, x_m) \in \mathbb{Z}^m$ denote the number of molecules of species x_j at time t for $j = 1, 2, \ldots, m$. Let $h_i(x)$, $i = 1, 2, \ldots, n$ denote a set of n reaction rates (which depend on the state x), and let $\nu_{\ell j} \in \mathbb{Z}$, $\ell = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$, denote the change in the number of molecules of species j after reaction ℓ. Assuming that the reactions occur at exponentially distributed times, independent of one another, gives rise to a birth–death process which can be computed as follows:

(i) Initialize $x_j(0)$, $j = 1, 2, \ldots, m$; set $k = 0$ and $t_k = 0$.
(ii) Compute the reaction rates $r_i = h_i(x(t_k))$, $i = 1, 2, \ldots, n$ and set $r = \sum_{i=1}^{n} r_i$.

(iii) Select a reaction: partition [0, 1] into n disjoint intervals I_i of length r_i/r and select a uniform random variable p_1 in [0, 1]. If p_1 falls in I_i then select reaction i.

(iv) Pick a second independent random variable p_2 uniformly in [0,1] and set

$$\tau = -\frac{\ln(p_2)}{r}, \qquad t_{k+1} = t_k + \tau.$$

(v) Set $x_j(t) = x_j(t_k)$, $t \in [t_k, t_{k+1})$ and $x_j(t_{k+1}) = x_j(t_k) + \nu_{ij}$, $j = 1, \ldots, m$.

(vi) Return to (ii) with $k \to k + 1$.

As in the previous section, such a process can be written in the form of (1.1) with $\gamma(z)\,\mathrm{d}W/\mathrm{d}t$ replaced by a Poisson counting process $\mathrm{d}W(z, t)/\mathrm{d}t$.

We assume that each $h_i(x)$ is a homogeneous polynomial of the form

$$h_i(x) = N\,\kappa_i \left(\frac{x_1}{N}\right)^{e_{i,1}} \left(\frac{x_2}{N}\right)^{e_{i,2}} \cdots \left(\frac{x_m}{N}\right)^{e_{i,m}} \equiv N\,\tilde{h}_i\left(\frac{x}{N}\right).$$

Then, if $X_i = x_i/N$ and $X = (X_1, \ldots, X_m)$ arguments similar to those in the previous section show that the master equation can be approximated by the Fokker–Planck equation

$$\frac{\partial \rho}{\partial t} + \sum_{i=1}^{m} \frac{\partial}{\partial X_i}[H_i(X)\rho] = 0,$$

where

$$H_i(X) = \sum_{j=1}^{n} \nu_{ji} \tilde{h}_j(X).$$

This indicates that the stochastic process for X will, for large N, be close to the deterministic system of ODEs

$$\frac{\mathrm{d}X_i}{\mathrm{d}t} = H_i(X), \qquad i = 1, \ldots, m.$$

Example 8.2. We illustrate this phenomenon with an example. Let $\ell = m = 3$ and $x(0) = (1, 1, N)$, $\tilde{h}_1(x) = x_1^2$, $\tilde{h}_2(x) = x_1 x_2$, $\tilde{h}_3(x) = x_3$. If

$$\begin{aligned} \nu_{11} &= -1, & \nu_{12} &= 1, \\ \nu_{22} &= -1, & \nu_{23} &= 1, \\ \nu_{33} &= -1, & \nu_{31} &= 1 \end{aligned}$$

with all other $\nu_{ij} = 0$ then the limiting ODE system is

$$\begin{aligned} \frac{\mathrm{d}X_1}{\mathrm{d}t} &= -X_1^2 + X_3, \\ \frac{\mathrm{d}X_2}{\mathrm{d}t} &= -X_1 X_2 + X_1^2, \\ \frac{\mathrm{d}X_3}{\mathrm{d}t} &= -X_3 + X_1 X_2. \end{aligned} \tag{8.7}$$

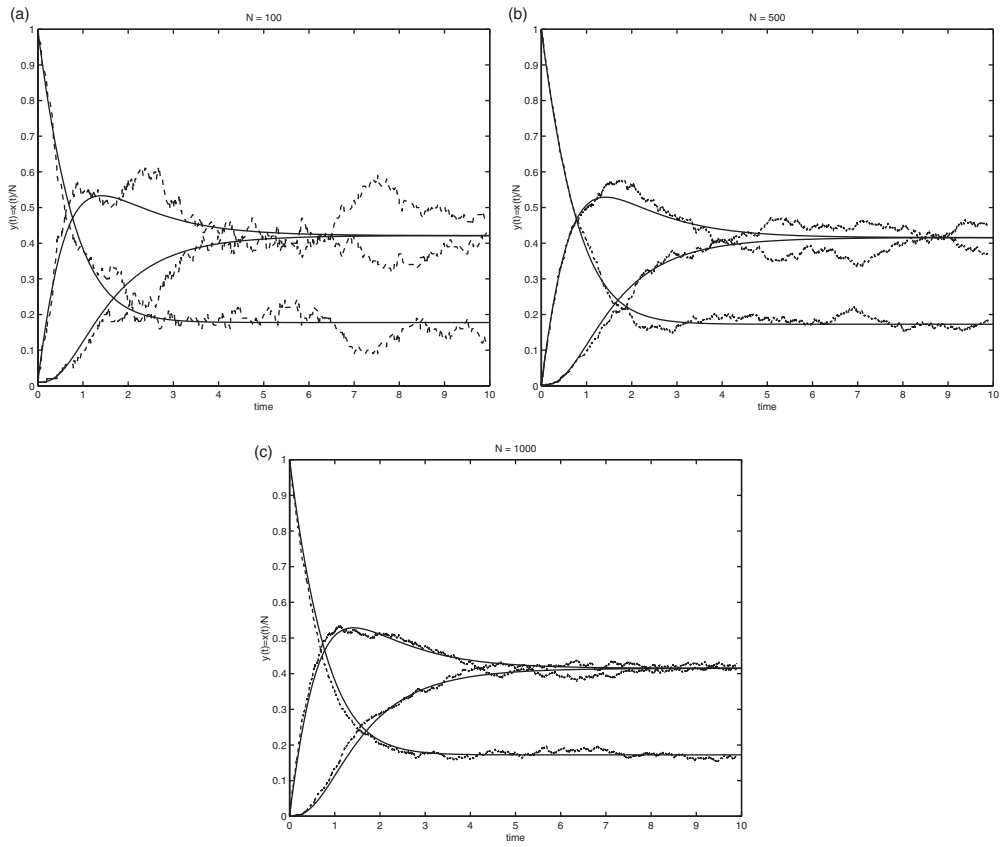

Figure 10. ——: sample paths of $x(t)/N$, where $x(t)$ is generated by the birth–death process with $\tilde{h}(x)$, $\nu_{\ell j}$ and $x(0)$ as described in the text, for (*a*) $N = 100$, (*b*) $N = 500$ and (*c*) $N = 1000$. - - - -: trajectories of $X_3(t)$ for $X(t)$ solving (8.7) with initial data $(1/N, 1/N, 1)$.

Figure 10 shows three sample paths of $x(t)/N$ for $x(t)$ given by the above birth–death process, for three values of N. As N increases, the paths of the stochastic process exhibit diminishing fluctuations about paths which solve the ODEs (8.7). ■

8.3. Discussion and bibliography

The fact that birth–death processes of the type studied here can be approximated by ODEs for large N, has been exploited by the physics and chemistry communities for some time [3]. Theorems making the formal asymptotic expansions given here rigorous may be found in [85, 86]; an overview of these theorems, from an applied perspective, may be found in the paper [87]. In general one can ask whether a closed system of equations exists for the evolution of a small number of moments. This section provides simple examples where the closed equations are for the first moment only.

When the birth–death process has spatial dependence through a lattice, then under appropriate scaling of the lattice variable with N it is possible to obtain PDEs, or stochastic PDEs when fluctuations remain in the limit; see [88,89], and the references therein, for example. The derivation of reaction–diffusion equations, and stochastic reaction–diffusion equations,

from birth–death processes combined with random walks, is an area of active interest in the physics community; see the lecture notes [90]. Related questions concerning derivation of the Boltzmann equation from a variety of stochastic hard sphere models are overviewed in [91,92].

9. Metastable systems

There is a final class of model problems which we introduce in this paper. The set-up differs substantially from that outlined in section 8. Specifically, this section is concerned with the extraction of small and finite state Markov chains from SDEs, or from large finite state Markov chains.

9.1. Finite state setting

In rough terms the ideas here apply to problems with backward equation (2.6) with the form

$$L = \begin{pmatrix} L^+ & \epsilon Q^\pm \\ \epsilon Q^\mp & L^- \end{pmatrix}. \tag{9.1}$$

The ideas may also be expressed in this way for the Chapman–Kolmogorov equation (2.9), but the technicalities are more complex for the uncountable state space setting. Hence, for SDEs, we describe the ideas pathwise.

For the moment we confine ourselves to the case of countable or finite state Markov chains, with backward equation (2.6), with generator of the form L given by (9.1). If L^+ and L^- both lead to ergodic Markov chains on the two subsets of variables χ^+, χ^- on which they act, then, for $\epsilon = 0$, L does not lead to an ergodic process: it has a two-dimensional null-space which can be parametrized by two eigenvectors supported wholly on χ^+ and χ^-, respectively. However, if the coupling matrices $Q^\pm$ and $Q^\mp$ are chosen so that L is ergodic for $\epsilon > 0$ then, on $\mathcal{O}(1)$ time-scales, the process will perform ergodic mixing in each of χ^+ or χ^- and then, on a time-scale of $\mathcal{O}(1/\epsilon)$, the communication between χ^+ and χ^- will be seen, leading to ergodic behaviour on the whole of $\chi = \chi^+ \oplus \chi^-$. By averaging separately over χ^+ and χ^-, with respect to the invariant measures of L^+ and L^-, it is possible to derive a two-state Markov chain, valid on time-scales of $\mathcal{O}(1/\epsilon)$, describing the dynamics between χ^+ and χ^-. This idea is developed in [93], and is closely related to earlier work in [94].

The idea generalizes to the derivation of m-state Markov chains when L has a block structure similar to (9.1), with m $\mathcal{O}(1)$ diagonal blocks and $\mathcal{O}(\epsilon)$ off-diagonal blocks. Such problems possess what are known as *metastable* states: the two invariant measures of the $\epsilon = 0$ dynamics are nearly invariant for the small ϵ but non-zero dynamics; they characterize the metastable states.

Note that the setting we are analysing corresponds to a situation where

$$L = L_1 + \epsilon L_2$$

and is hence, after time-rescaling, similar to situations discussed in sections 4–7. However, here the leading-order operator in L, L_1, is not ergodic, whereas in those previous sections our assumption about the fast dynamics is that it is ergodic.

9.2. The SDE setting

For SDEs (1.3), a natural class of model problems exhibiting this kind of behaviour are gradient systems with additive noise:

$$\frac{\mathrm{d}X}{\mathrm{d}t} = -\nabla V(X) + \sigma \frac{\mathrm{d}W}{\mathrm{d}t} \tag{9.2}$$

with $X \in \mathbb{R}^d$ and V a double-well potential. If σ is small then large deviation theory may be used to derive a two-state Markov chain describing transitions between the wells of V. The idea is that, if $\sigma = 0$, then Lebesgue a.e. initial data will converge, under the dynamics, to one of the two wells of V. This partitions the phase space into two sets, χ^+ and χ^-, centred on the two wells. If $\sigma \ll 1$ then this deterministic dynamics will govern the behaviour of the SDE. On a longer time-scale, exponentially large in small σ, the process will exhibit occasional transitions between the two sets χ^+ and χ^- (the wells), and a Markov chain may describe this transition process.

Similar ideas also apply to the second-order dynamics

$$\frac{\mathrm{d}^2 X}{\mathrm{d}t^2} = -\nabla V(X) - \frac{\mathrm{d}X}{\mathrm{d}t} + \sigma \frac{\mathrm{d}W}{\mathrm{d}t} \tag{9.3}$$

with $X \in \mathbb{R}^d$ and V again a double-well potential.

Example 9.1. A typical example of model problem (9.2) is given by (6.23). This corresponds to (9.2) with $d = 1$, $V(x) = \frac{1}{4}(1 - x^2)^2$. In figure 11 we plot three sample paths of (6.23) over a time interval of $T = 1000$ and different values of the noise coefficient σ. For $\sigma = 1$, there is no scale-separation and indeed, the paths does not exhibit metastable behaviour. For $\sigma = 0.5$ the time it takes to switch between wells is short compared with the time spent in each well, and clear metastable behaviour is observed. For $\sigma = 0.2$ the transition rate is low in comparison with the observation time, thus, no transition is observed; throughout the simulation the system remains in a single metastable state. ■

9.3. Discussion and bibliography

For finite state Markov chains the ideas described above are developed in [93]; they are closely related to earlier work in [94]. For both the SDEs described above, see the material on large deviations in [95].

10. Algorithms for stiff problems

In this section, and in all subsequent sections, we discuss algorithms. For problems with two time-scales, $\mathcal{O}(1)$ and $\mathcal{O}(\epsilon)$, it is important for reasons of efficiency to have time-stepping methods which only expend the minimum effort necessary in resolving the fast scales. For example it may be of interest to use time-step $\Delta t \approx \mathcal{O}(\epsilon)$, or even $\Delta t \gg \epsilon$. This is the problem of *stiffness*, which has been at the heart of the numerical analysis of ODEs since the work of Dahlquist in the 1950s. In this section we describe a variety of such situations that arise when the extraction of macroscopic dynamics is the desired outcome of the simulations, and relate the existing literature to the model problems and examples highlighted in the preceding sections. The discussion will show that this is a subject area where rigorous analysis underpins many of the algorithmic ideas. Much, but not all, of the analysis is restricted to situations where the separation into fast and slow variables is explicit.

10.1. Invariant manifolds

For problems with exponentially attracting invariant manifolds, on which the system may be approximated by a DAE, it is of importance to understand which numerical methods will reproduce the dynamics on the invariant manifold in a robust fashion, without using unnecessarily small time-steps. A relevant model problem used to evaluate this issue is

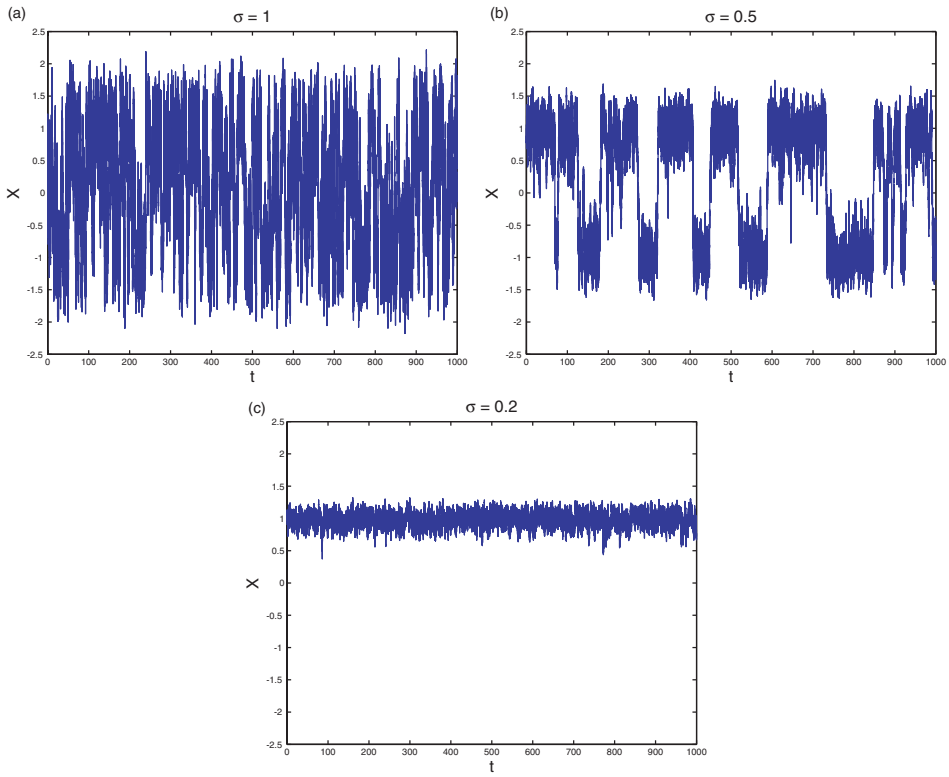

Figure 11. Sample paths of (6.23) over a time interval of $T = 1000$ for values of the noise coefficient: (a) $\sigma = 1$, (b) $\sigma = 0.5$ and (c) $\sigma = 0.2$.

equation (4.3) and example 4.1. In this context, the work of Nipp and Stoffer [96, 97] is of interest. They employ an invariant manifold theorem for maps, formulated in the unpublished report [98], and more recently detailed in the book [99]. For further discussion of the reduction of ODEs to DAEs, and the appropriate choice of numerical methods, see the texts [100, 101].

Roughly speaking, A-stable or $A(\alpha)$-stable methods, such as backward Euler, behave well for such problems; methods such as forward Euler behave poorly. The following illustration shows how implicitness achieves this good behaviour.

Illustration 10.1. We use example 4.1. Equations (4.6) can be written in the abstract form

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= Lx + ye, \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= -\frac{y}{\epsilon} + \frac{x_1 x_3}{\epsilon}, \end{aligned} \tag{10.1}$$

where $x = (x_1, x_2, x_3)^T$ and $e = (0, 0, 1)^T$. Recall that this problem has an attractive invariant manifold $y = x_1 x_3 + \mathcal{O}(\epsilon)$ for $\epsilon \ll 1$. Now consider the numerical method

$$\begin{aligned} x^{n+1} &= x^n + \Delta t L x^n + \Delta t y^n e, \\ y^{n+1} &= y^n - r y^{n+1} + r x_1^{n+1} x_3^{n+1}, \end{aligned} \tag{10.2}$$

where $r = \Delta t/\epsilon$ is the numerical time-step in units of the fast time-scale.

This linearly implicit method treats the stiff part of the problem by a backward Euler-like approximation, and the remainder by the forward Euler method. Figure 12(a) shows

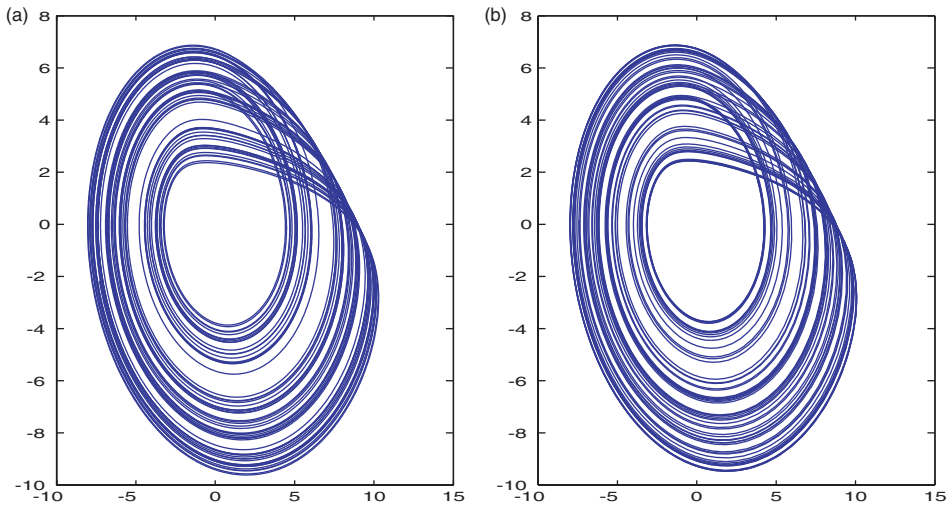

Figure 12. (*a*) Trajectory of the numerical approximation (10.3) for $\Delta t = 10^{-2}$ and $\epsilon = 10^{-4}$, projected onto the (x_1, x_2) plane. (*b*) Projected trajectory for the Rössler limit (4.8).

the numerically computed attractor for this approximation, with $\Delta t = 10^{-2}$ and $\epsilon = 10^{-4}$. Figure 12(*b*) shows the attractor for the Rössler limit (4.8). Comparison of the two figures shows that the method appears to compute the $\epsilon = 0$ limiting behaviour accurately, even though we are computing the fast variable in a highly under-resolved regime $r = \Delta t/\epsilon = 100$. The explanation for this is as follows. If we fix any $r > 0$ and consider $\Delta t \ll 1$ then the numerical method has an attractive invariant manifold

$$y = \eta^{\Delta t}(x), \qquad \eta^{\Delta t}(x) = x_1 x_3 + \mathcal{O}(\Delta t).$$

This should be compared with the invariant manifold for the equations themselves, given by (4.7). The fact that both manifolds agree to leading-order in Δt or ϵ explains the good behaviour of the numerical method. Substituting the form of the (numerical) invariant manifold into the discrete equation for x gives

$$\begin{aligned} x^{n+1} &= x^n + \Delta t L x^n + \Delta t \eta^{\Delta t}(x^n) e, \\ &\approx x^n + \Delta t L x^n + \Delta t x_1^n x_3^n e. \end{aligned}$$

The last expression is simply the forward Euler update for the Rössler limit (4.8), and it is known that this method has an attractor close (in the sense of upper semi-continuity [102]) to the attractor of the equation itself. Thus, we accurately capture the $\epsilon = 0$ limiting dynamics. Note further, that the approximation improves as r *increases* (greater under-resolution) in the sense that the rate of attraction to the invariant manifold increases. If, instead, the forward Euler method is applied directly to (10.1) then we obtain

$$\begin{aligned} x^{n+1} &= x^n + \Delta t L x^n + \Delta t y^n e, \\ y^{n+1} &= y^n - r y^n + r x_1^n x_3^n. \end{aligned} \tag{10.3}$$

This method has an invariant manifold of the desired form, only for $r = \Delta t/\epsilon < 2$. In summary this illustration shows two important facts: (i) that typical methods for model problems like (4.3) capture the correct $\epsilon = 0$ limiting dynamics for $\Delta t = \mathcal{O}(\epsilon)$ and smaller; (ii) that special methods, typically implicit, can capture the $\epsilon = 0$ limiting dynamics for arbitrary $r = \Delta t/\epsilon$, in particular for r large. ■

Gear and Kevrekidis [103] recently proposed a class of numerical methods for stiff problems with invariant manifolds. The slow variables are evolved in time with time-steps that are coarse compared to the fast variables. At each (coarse) time-step the y-equation is evolved for a short time, with x fixed, until it reaches the invariant manifold. The resulting value of y is then used to evaluate the time derivative of x for the coarse time integrator. Gear and Kevrekidis called this method 'projective integration'. E [104] generalized this method to stiff oscillatory systems as well (see below), and derived rigorous estimates of the difference between the projective integration algorithm and the asymptotic $\epsilon \to 0$ limit, for both stiff dissipative and stiff oscillatory problems. E *et al* [105] generalized this analysis to stochastic systems with separate treatment of intermediate- and long-time behaviour.

Illustration 10.2. Consider the following scale-separated system:

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= y, \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= \frac{1}{\epsilon}(-x + y - y^3). \end{aligned} \tag{10.4}$$

For $|x| < 1/\sqrt{3}$ the y-equation is bistable, with two attracting branches that we label $y = \eta_{\pm}(x)$. If $|x(0)| < 1/\sqrt{3}$ and $y(0)$ is, say, in the basin of attraction of the upper branch, then, for ϵ small, the trajectory remains close to the manifold $(x, \eta_+(x) + \mathcal{O}(\epsilon))$ with $x(t)$ increasing in time. When x crosses the critical point $1/\sqrt{3}$, y rapidly drops to the lower branch, and x starts to decrease along the manifold $(x, \eta_-(x) + \mathcal{O}(\epsilon))$, until it reaches the value $-1/\sqrt{3}$, where y rapidly climbs to the upper branch, giving rise to periodic behaviour.

The projective integration algorithm, in its simplest version, consists of the following steps.

(i) Given (x^n, y^n), keep x^n fixed and integrate the y-equation with a forward Euler scheme

$$y^{n,m+1} = y^{n,m} + \frac{\delta t}{\epsilon}(-x^n + y^{n,m} - (y^{n,m})^3), \qquad m = 0, 1, \ldots, M-1,$$

where $y^{n,0} = y^n$. The 'microscopic' time-step δt has to be small compared with ϵ, whereas the integration time $M\delta t$ has to be long enough so that $y^{n,m}$ reaches the vicinity of the invariant manifold. Then set $y^{n+1} = y^{n,M}$.

(ii) Evolve x in time with

$$x^{n+1} = x^n + \Delta t\, y^{n+1},$$

where Δt is the 'macroscopic' time-step, which is independent of ϵ.

This basic scheme can be generalized for higher-order solvers.

In figure 13 we compare $x(t)$ solving (10.4) with $\epsilon = 10^{-4}$ (solid line) with the output of the projective integration scheme with parameters $\Delta t = 5 \times 10^{-3}$, $\delta t = \epsilon/10$, and $M\delta t = 10\epsilon$ (dotted line). The agreement is good, and can be much improved by using higher-order solvers. Furthermore, notice that $M\delta t = 10^{-3} \ll \Delta t$. Thus the fast dynamics does not need to be integrated accurately over a complete macroscopic time-step to achieve accuracy. ■

The numerical calculation of slow dynamics, by dimension reduction in fast–slow systems and with application to problems in chemical kinetics, is described in [106] and in [107] with more recent developments in [108].

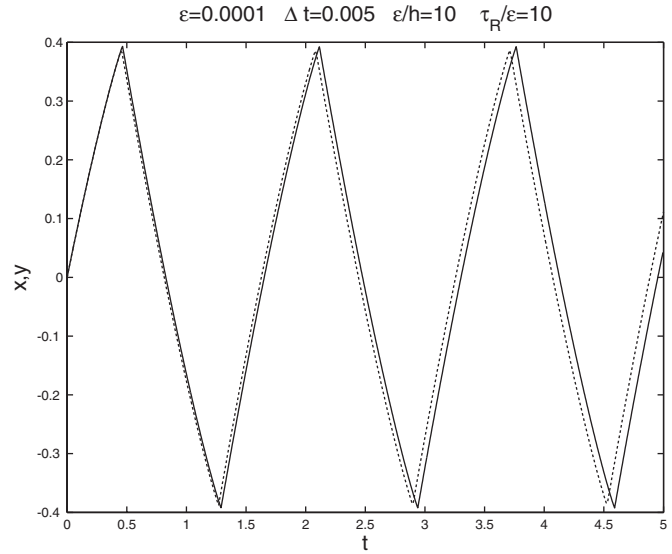

Figure 13. ——: $x(t)$ solving (10.4) with $\epsilon = 10^{-4}$. Dotted line: output of the projective integration scheme with parameters $\Delta t = 5 \times 10^{-3}$, $\delta t = \epsilon/10$ and $M\delta t = 10\epsilon$.

The idea of scale-separation and the resulting invariant manifolds, outlined in section 4, has been used as the basis for numerical algorithms applied to Galerkin truncations for PDEs. Here, x represents the low wave number modes, and y the remainder and the algorithms attempt to approximate numerically a function η: $\mathcal{Y} \to \mathcal{X}$ which is approximately invariant under the dynamics; this is the *nonlinear Galerkin method* for approximate inertial manifolds. A useful reference where this is studied in an applied context is [109], and a discussion of the rate of convergence of such algorithms may be found in [110]. A more recent perspective on these methods, and a cheap implementation through post-processing of the standard Galerkin method, is discussed in [111].

10.2. Averaging

Stiff oscillatory problems have received less attention in the literature than dissipative ones, but the literature is growing [112]. A major impetus is computational molecular dynamics. An important class of model problems is given by the Hamiltonian (5.5) and example 5.1. In particular, chapter XIII of the book [99] provides a good overview of this topic, and the references therein provide a comprehensive bibliography, including the known rigorous analysis. The paper [113] initiated a rigorous analysis of many of the issues raised by practitioners in this field.

For dissipative perturbations of Hamiltonian problems there is also interesting work on the design and analysis of numerical schemes which construct the correct slow dynamics (an attractive invariant torus for example) without resolving the fastest scales. See [114] and chapter XII of the book [99] for a description of the known rigorous analysis.

Illustration 10.3. We use example 5.1. Considering this as a model problem for molecular dynamics, it is natural to study the effect of approximation by explicit symplectic methods which are widely used in this application area. Applying the symplectic Euler

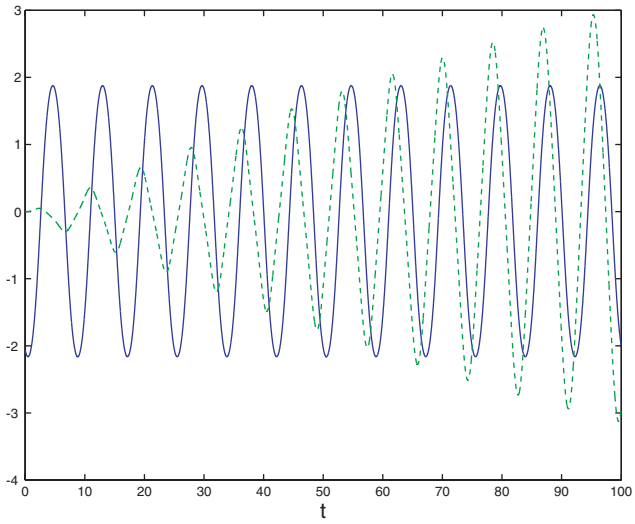

Figure 14. Time evolution of $x(t)$ solving (5.6) and computed using (10.5) with $\Delta t = 0.01$, $\epsilon = 0.01$ and $r = 1$ (——) and $X(t) - x(t)$, with $X(t)$ solving (5.8) (- - - -). We used $V(x) = -\cos x$ and $\omega(x) = 1 + 0.5 \sin x$.

method [115, 116] to equations (5.6) gives

$$\begin{aligned} x^{n+1} &= x^n + \Delta t p^n, \\ \eta^{n+1} &= \eta^n + r v^n, \\ p^{n+1} &= p^n - \Delta t V'(x^{n+1}) - \frac{\Delta t}{2}\omega'(x^{n+1})(\eta^{n+1})^2, \\ v^{n+1} &= v^n - r\omega(x^{n+1})\eta^{n+1}, \end{aligned} \tag{10.5}$$

where, as above, $r = \Delta t/\epsilon$.

Figure 14 shows a numerical simulation with $\Delta t = 0.01$ and $r = 1$ (under-resolved), using the same initial data as in figure 3. The time traces are of x, given by (10.5), and $x - X$, with X solving (5.8). It is clear that choosing the value $r = 1$ leads to an inaccurate numerical simulation of x as it is $\mathcal{O}(1)$ different from X solving (5.8). In contrast, the simulation in figure 3 uses $r = 0.1$ and correctly captures the limiting dynamics given by X.

It is natural to ask why the situation here is so different from the stiff dissipative illustration 10.1. In that context the linearly implicit method calculates the $\epsilon = 0$ limit correctly for $r = 100$, and even the explicit methods such as forward Euler do the same for $r = \mathcal{O}(1)$. In contrast, for this oscillatory example we require $r \ll 1$ to get the correct limit. The reason for the difference lies in the nature of the ergodic averages which must be represented. For the dissipative problems we need only approximate a delta measure, whereas for the oscillatory problems we are trying to capture a measure which has full support on the energy surface of the (η, v) dynamics, with (x, p) frozen. For example 5.1 the invariant measure for η is determined by η solving the following harmonic oscillator problem:

$$\frac{\mathrm{d}^2\eta}{\mathrm{d}t^2} + \frac{\omega(x)}{\epsilon^2}\eta = 0. \tag{10.6}$$

For fixed r the invariant measure for $\{\eta^n\}$ generated by the method (10.5) is determined by, for $x^n = x$ frozen,

$$\eta^{n+1} + [r^2\omega(x) - 2]\eta^n + \eta^{n-1} = 0. \tag{10.7}$$

This corresponds to a Leap-Frog discretization of (10.6) with time-step $\Delta t = r\epsilon$. For certain values of $r^2\omega(x)$ this equation has totally different invariant measure than does (10.6). For example, if $r^2\omega(x) = 1$, then the numerically generated invariant measure is supported on three points; in contrast the true invariant measure is absolutely continuous with respect to Lebesgue measure on an interval. As $x = x^n$ varies so does $r^2\omega(x)$ and the empirical measure generated by the numerical method will, for certain x, differ substantially from the true measure. These *resonances* are responsible for the $\mathcal{O}(1)$ error manifest in figure 14 where $r = 1$. Avoiding these resonances is only possible by correctly resolving the fast dynamics, as in figure 3, for which $r = 0.1$. ■

Illustration 10.4. The basic idea of projective integration can be extended to Hamiltonian problems such as example 5.1. The averaging algorithm of E [104] has to be adapted to the current situation in which it is of crucial importance to preserve the total energy of the system:

(i) Given (x^n, p^n, η^n, y^n), evolve (η, v) by solving (10.5), with $x = x^n$ and $p = p^n$ fixed, using a 'microscopic' time-step δt over an integration interval $M\delta t$. Denote the discrete solutions by $\eta^{n,m}$, $v^{n,m}$, $m = 0, 1, \ldots, M-1$, initializing with $\eta^{n,0} = \eta^n$, $v^{n,0} = v^n$.

(ii) Evolve the slow variables with a forward Euler method:

$$x^{n+1} = x^n + \Delta t\, F_1(x^n, p^n),$$
$$p^{n+1} = p^n + \Delta t\, F_2(x^n, p^n),$$

where Δt is the 'macroscopic' time-step. The functions F_1, F_2 are evaluated by averaging the right-hand sides of (10.5) over the solutions of the microsolver.

$$F_1(x^n, p^n) = x^n,$$
$$F_2(x^n, p^n) = -V'(x^n) - \frac{\omega'(x^n)}{2}\frac{1}{M}\sum_{m=0}^{M-1}(\eta^{n,m})^2.$$

(iii) (η^{n+1}, v^{n+1}) are randomly drawn from the energy shell with (x^{n+1}, v^{n+1}) given.

In figure 15 we compare $(x(t), p(t))$ solving (10.5) for $\epsilon = 10^{-3}$ and total energy $E = 1$ (solid line) with the output of the projective integration scheme with parameters $\Delta t = 5\times 10^{-2}$, $\delta t = \epsilon/3$ and $M\delta t = 3\epsilon$ (dotted line). As in the dissipative case, the agreement can be much improved by using higher-order solvers for both 'microscopic' and 'macroscopic' components. As with the previous illustration of projective integration the micro-solver does not need to be used for a complete macro-step: $M\delta t \ll \Delta t$. Thus the work required is substantially less than for resolved integration of the fastest scales. ■

It is worth noting here that for many deterministic Hamiltonian problems arising in practice, ergodicity is frequently difficult to prove, and often probably not true. A recent paper by Tupper [117] addresses this issue by considering a weakening of the definition of ergodicity (due to Khinchin) and shows that it is robust to perturbation, in particular to perturbations introduced numerically.

10.3. Stiff stochastic systems

Stiff stochastic systems have recently started to receive some attention in the literature. Representative model problems include (5.9), (6.3) and example 6.1. The papers [57, 105] describe how to implement the analytic program described in section 6.1 as a numerical algorithm. New dependent variables are used to achieve variance reduction in the averaging procedure. The techniques of section 6.1 facilitate the rigorous analysis and justification of this numerical method.

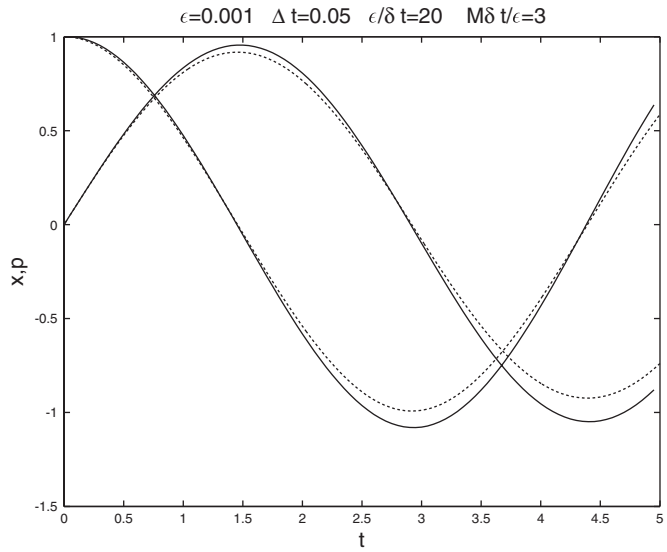

Figure 15. ——: $(x(t),\, p(t))$ solving (10.5) for $\epsilon = 10^{-3}$ and total energy $E = 1$. $\cdots\cdots$: output of the projective integration scheme with parameters $\Delta t = 5 \times 10^{-2}$, $\delta t = \epsilon/3$ and $M\delta t = 3\epsilon$.

Illustration 10.5. We use example 6.2. The calculation of the diffusion coefficient given in that example is very expensive requiring resolved simulation of (6.22) with time-step $\Delta t = \mathcal{O}(\epsilon)$ or smaller. In contrast, the method of Vanden Eijnden calculates the diffusion coefficient by means of a single simulation of the Lorenz equations, with an ϵ independent time-step. His method is based on the representations (6.16), (6.17) for the effective drift and diffusion. In example 6.2 we have $x \in \mathbb{R}$, $y = (y_1, y_2, y_3)^T \in \mathbb{R}^3$, and

$$f_0(x, y) = \tfrac{4}{90} y_2, \qquad f_1(x, y) = x - x^3.$$

Thus, the effective diffusion coefficient, σ, is given by

$$\frac{1}{2}\sigma^2 = \lim_{T\to\infty} \left(\frac{4}{90}\right)^2 \int_0^\infty \left\{ \frac{1}{T} \int_0^T \int_0^\infty y_2(s) y_2(s+t)\, \mathrm{d}s \right\} \mathrm{d}t.$$

Vanden Eijnden's method, applied to this problem consists of estimating the effective diffusion coefficient by a sampling over a discrete, properly resolved trajectory $y^n = y(n\Delta t)$ of (6.22), truncating the average at a finite time $T = M\Delta t$:

$$\sigma^2 \approx 2 \left(\frac{4}{90}\right)^2 \frac{\Delta t}{M} \sum_{m=0}^{M-1} \sum_{m'=0}^{M-1-m} (y_2)^m (y_2)^{m+m'}.$$

In the present case, the statistical errors associated with this method are large, because the numerically computed auto-covariance of $y_2(t)$ does not decay fast enough. By properly truncating the numerical auto-covariance we get $\sigma^2 = 0.13 \pm 0.01$, which is in agreement with the estimate obtained in example 6.2 by a different, but less efficient method. ■

10.4. The heterogeneous multiscale method

The paper [118] outlines a general framework for the numerical solution of problems with multiple scales—what the authors term the heterogeneous multiscale method. The basic idea is that, for problems exhibiting some form of ergodicity in the fast scale (which may be spatial

as well as temporal—PDEs are included) integration need only be performed on intervals sufficient to calculate the averaged effect of the fast scales on the slow scales; the length of such intervals is determined by decorrelation. This idea was used in illustrations 10.3, 10.4 and 10.5. Employing this idea can lead to substantial computational savings.

The approach allows for a unified treatment of the integration of both stiff dissipative and stiff oscillatory ODEs. Rigorous analysis is possible in some cases where the separation into fast and slow variables is *a priori* explicit. See [104] where two model problems, one dissipative and closely related to model problem (4.3), and the other oscillatory and closely related to the model problem (5.5), are studied. The key idea is that ergodicity of the fast variables, with the slow variable frozen, can be used to justify the use of an empirical measure under which the slow vector field is averaged; this empirical measure will be close to a Dirac mass for the stiff dissipative case, and will typically have the full Lebesgue measure on an appropriate energy surface for oscillatory problems. Similar ideas also apply to SDEs where, typically, the empirical measure will have support on the whole $\mathcal{Y}$-space; this is the basis of the algorithm described in [57] and in the previous subsection.

10.5. Heat baths

For many Hamiltonian problems arising in applications there is no clear separation of scales, but rather a broad spectrum. It is still of interest to integrate (1.1) by a time-stepping method which does not resolve all the time scales in $\mathcal{Y}$. In general this approach will fail because of numerical instabilities, or resonances between the time-step frequency and fast unresolved scales (see [119,120] for example). However, there are situations where unresolved simulations correctly reproduce macroscopic behaviour, and one example is for models similar to the Hamiltonian heat bath model given by (7.7) and examples 7.3, 7.4. These models are of interest because, whilst they contain a wide range of scales, they do not have explicit scale-separation. Thus they include a feature present in many realistic Hamiltonian models and which is not present in the model problem (5.5) and example 5.1. An open area for investigation is to develop useful model problems combining the characteristics of both (5.5) and (7.7). Such models would exhibit the features inherent in many molecular dynamics simulations.

Illustration 10.6. Figure 16 shows simulations of the Hamiltonian problem (7.7) with parameters similar to those detailed in example 7.4; the limiting SDE is of the form (7.12). It is possible to prove strong convergence to the SDE as $N \to \infty$ (see [74]). Figure 16(a) shows a path of the limiting SDE, together with simulation of (7.7) by the backward Euler method with $\omega_{\max}\Delta t = 10$; here $\omega_{\max}$ is the largest natural frequency of the bath variables $\{q_j\}_{j=1}^{N}$ with Q_N fixed. The backward Euler method in this under-resolved regime does not capture the irregularity of the SDE limit, but it calculates an excellent smoothed approximation. Figure 16 quantifies the errors in the backward Euler method, when compared with the SDE limit, under $\omega_{\max}\Delta t = 10$, as $N \to \infty$ ($\omega_{\max}$ is proportional to N); the results clearly show convergence. See [74] for details. ■

The paper [74] shows how the backward Euler method correctly predicts macroscopic behaviour of the distinguished particle, without resolving fast scales; various other time-stepping methods are also studied from this perspective. The paper [71] uses the model problem of example 7.2 to give rigorous analysis explaining some of the numerical experiments in [74]. The paper [121] shows a connection between these heat bath models and optimal prediction (described below). Generalizations of these ideas to other memory kernels are considered in [81]. A different, but related, study of the ability of a variety of different integrators, used in the under-resolved regime, to predict macroscopic behaviour, is due to Tupper [122]; he looks

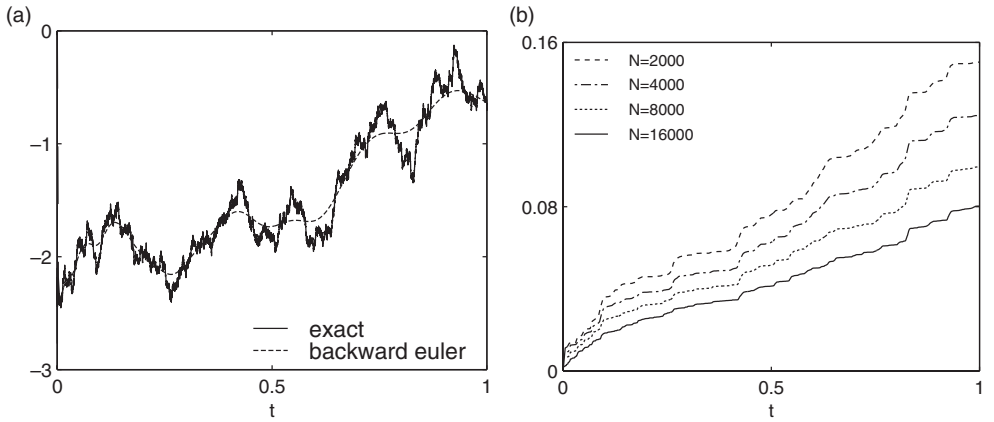

Figure 16. Numerical simulation of the Hamiltonian (7.7) by the backward Euler method, and comparison with the exact SDE limit: (*a*) shows sample paths; (*b*) shows the error between the backward Euler simulation and the SDE limit, for $N\Delta t$ fixed.

at interacting particle systems with Gaussian behaviour for a tagged particle, and compares numerical methods by their ability to compute the auto-correlation accurately.

11. System identification

Another approach to finding stochastic closures of deterministic dynamics is to use parameter estimation to fit SDE models to partially observed dynamics. Important applications include the atmospheric sciences [123] and molecular dynamics [124]. Typically these involve models which are, in some sense, close to problems which have been understood analytically. The subject of system identification is an enormous one, and here we have only referenced the literature that link directly with model problems outlined in this paper. For a general introduction on the subject see [125].

11.1. Atmospheric sciences

The work [61], motivated by problems in the atmospheric sciences, fits stochastic models to partially observed deterministic dynamical systems which, under heuristics concerning the OU-like behaviour (recall (6.18)) of the (deterministic but chaotic) Burgers' heat bath example of [56], are close to the scale-separated SDEs of section 6. The approach in [61] is to relate three models: (i) a large deterministic system coupling slow variables to a larger Burgers' bath, of the type introduced in [56]; (ii) a smaller stochastic system of the form of model problem (6.3) resulting when the Burgers' bath is replaced by an OU process; and (iii) a reduced stochastic model of the form (6.13). The results are compared by fitting data from simulations of (i) directly to models of the form (iii); or, instead, fitting the data to models of the form (ii), and then using the analysis of section 6.1 to derive a reduced stochastic model of the type (iii). Both methods produce comparable results, justifying the direct fit of low-dimensional stochastic models to high-dimensional chaotic deterministic ones. There is a connection here with the heat bath models problems of (7.7) in that, in both situations, large deterministic systems produce effective stochasticity in a small number of special variables.

11.2. Molecular dynamics

In the paper [80], motivated by the modelling of biomolecular conformational dynamics, SDEs are fit to heat bath models which generalize those in section 7. Specifically a non-Hookean coupling between the heat bath and the particle is introduced (it is the linearity of the Hookean coupling which facilitates the explicit construction of an SDE from an ODE). The Hamiltonian (7.7) is modified to replace the quadratic dependence in the last term by a quartic one, for example. More generally, Hamiltonians of the form

$$H(P_N, Q_N, p, q) = \frac{1}{2}P_N^2 + V(Q_N) + \sum_{j=1}^{N} \frac{p_j^2}{2m_j} + \sum_{j=1}^{N} k_j \Phi(q_j - Q_N) \tag{11.1}$$

are studied, the case $\Phi(x) = x^2/2$ being the analytically tractable case governed by the Hamiltonian (7.7). A formal linear response analysis, which is shown to apply at high enough temperatures, still suggests fitting data to models of the form (7.11) for the more general couplings in (11.1) and parameter estimation for the case $\Phi(x) = x^4/4$ is studied extensively. In [126] the parameter estimation techniques are also shown to work in the case $\Phi(x) = x^2/2$, where the approximation of ODEs by SDEs is rigorously justified as in examples 7.3 and 7.4. The paper [126] also introduces more sophisticated time-series methods for the estimation of parameters.

Illustration 11.1. We consider the Hamiltonian system (11.1) with quartic inter-particle potential $\Phi(x) = x^4/4$ as in [80]. The spring coefficients are chosen as follows: $k_j = g(\nu_j)\,\Delta\nu$, $j = 1, 2, \ldots, N$, where $g(x) = 0.4/(0.16 + x^2)$, $\nu_j = j\Delta\nu$ and $\Delta\nu = N^{-3/4}$. The masses of the heat bath particles are given by $m_j = k_j^{1/2}/\nu_j^2$. Given initial values for (Q_N, P_N) the heat bath variables (q, p) are randomly drawn from a Gibbs distribution with temperature 1. As shown in [80], the trajectories $Q_N(t)$ are well approximated, in a weak sense, by the solution $Q(t)$ of the stochastic system:

$$\begin{aligned} \frac{\mathrm{d}Q}{\mathrm{d}t} &= P, \\ \frac{\mathrm{d}P}{\mathrm{d}t} &= -V'(Q) + R, \\ \frac{\mathrm{d}R}{\mathrm{d}t} &= -(\alpha R + \beta P) + (2\alpha\beta)^{1/2}\frac{\mathrm{d}B}{\mathrm{d}t}, \end{aligned} \tag{11.2}$$

where $B(t)$ is standard Brownian motion, and $\alpha = 0.623$ and $\beta = 1/3.108$ are parameters fitted by the analysis of long time sequences of $Q_N(t)$.

In figure 17 we compare the empirical distribution and auto-covariance of sample paths $Q_N(t)$ solving the Hamiltonian system with the corresponding distribution and auto-covariance of the fitted stochastic process (11.2). ■

12. Evolving moments

The idea of deriving closed equations for a small number of moments is, of course, central to kinetic theory. The celebrated BBGKY hierarchy derives the equations of fluid mechanics from the Boltzmann equations by use of this approach [127]. The idea of propagating moments from microscopic simulations has already appeared in the literature; see [128]. Recently, there has been a concerted effort in the computational mathematics community to build algorithms to carry out this procedure. Rigorous analysis justifying these approaches is limited, but practical

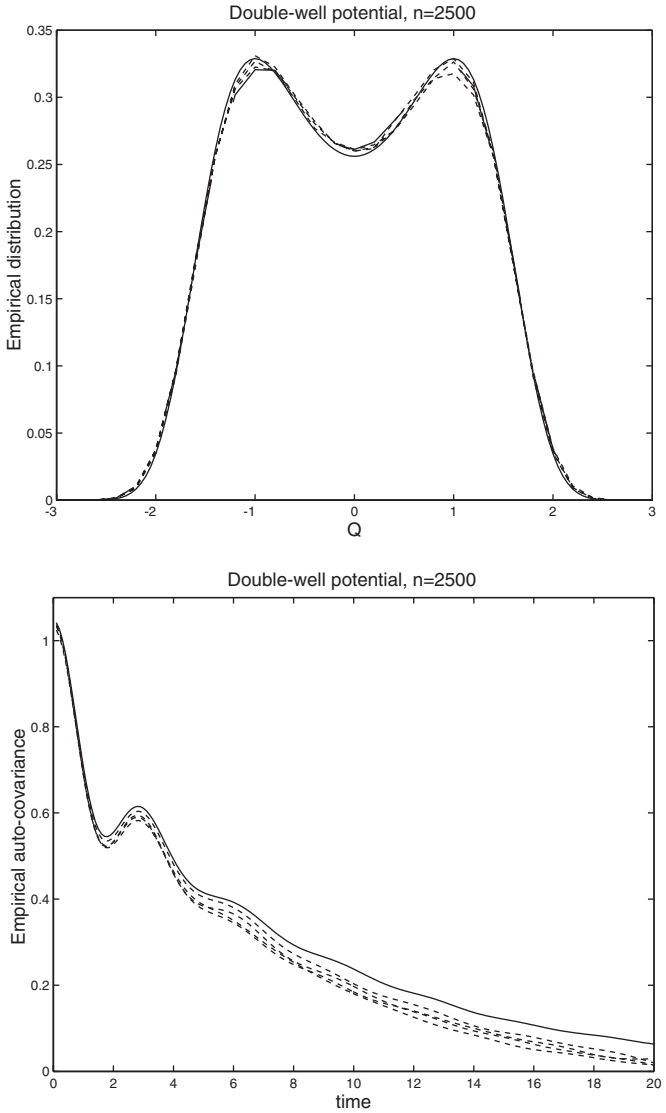

Figure 17. Top: empirical distribution of four sample paths solutions $Q_N(t)$ of the Hamiltonian system with Hamiltonian (11.1) with a quartic inter-particle potential $\Phi(x) = x^4/4$, $V(Q) = Q^4/4$ and $N = 2500$ heat bath particles (- - - -). The solid line is the equilibrium distribution of the approximating SDE (11.2). Bottom: the empirical auto-covariance for sample path of $Q_N(t)$ (- - - -) and the auto-covariance of the approximating SDE (——).

experience is building steadily. Here we outline two approaches used in the literature, and relate them.

12.1. Optimal prediction

In this work [19, 129–131] the underlying assumption is that the equation (1.1) carries a natural measure ν which is invariant under the flow induced by (1.1). For simplicity assume that $\gamma \equiv 0$

so that the problem is deterministic; thus $\alpha, \beta \equiv 0$, as well. From (1.2) we obtain

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= f(x, y), \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= g(x, y). \end{aligned} \tag{12.1}$$

The objective is to find an equation for the conditional expectation,

$$X(t) = \mathbb{E}[x(t) \mid x(0) = x_0], \tag{12.2}$$

where $\mathbb{E}$ is with respect to measure ν on $z(0) = (x(0), y(0))$. Thus $X(t)$ is the first moment of $x(t)$.

One basic approximation is to simply average $f(x, y)$ with respect to ν, conditional on x being held at its mean value, yielding

$$\frac{\mathrm{d}X}{\mathrm{d}t} = F(X), \qquad X(0) = x_0, \tag{12.3}$$

where

$$F(\zeta) = \int f(\zeta, y)\nu_\zeta(\mathrm{d}y) \tag{12.4}$$

and where ν_ζ is the appropriate conditional measure.

This method cannot work well in general, and errors between the solution of (12.3) and the desired solution of (12.2) can grow like t [132]. Model problem (7.7) and example 7.3 shows why this is so. For large N the variable $X(t) = (Q_N(t), \dot{Q}_N(t))$ satisfies an SDE and, in general, the moments of an SDE do not satisfy a closed ODE—they satisfy a PDE, namely the Chapman–Kolmogorov equation (2.9). Furthermore, the limit SDE has memory—it is Markovian only for $(Q, \dot{Q}, S)$ and the presence of the variable S is not accounted for in the method of optimal prediction. The discussion of Mori–Zwanzig projection operators in section 3 shows that both memory and noise are indeed typically present, and need to be accounted for in any elimination procedure. Suggestions for how to account for memory effects within the optimal prediction framework are presented and tested in [19]. The approach of Chorin and co-workers in [133] to overcome the absence of noise is to put fluctuations back into the model (12.3) to understand how typical paths $x(t)$ might behave, as well as to include a compensating damping term, accounting for memory. This is done on an *ad hoc* basis by fitting a diffusion coefficient γ in the model

$$\frac{\mathrm{d}X}{\mathrm{d}t} = F(X) - \gamma X + \sqrt{2\gamma T}\,\frac{\mathrm{d}U}{\mathrm{d}t},$$

where U is standard Brownian motion. In this approach the idea is to abandon the goal of finding closed equations for moments and, instead, to try and generate typical paths in $\mathcal{X}$ but without resolving the y variables. Another approach to overcoming the limitations of the method is to propagate further, higher-order, moments; this idea is discussed in the next subsection.

We should mention that, for certain model problems, the method of optimal prediction works well. One such set of problems are given by example 7.4 with V being quadratic. The limit dynamics is then a linear SDE (7.12) for which the expectation of $X(t) = (Q(t), \dot{Q}(t))$ does satisfy a closed memoryless ODE; this fact was observed empirically in [121]. Another set are problems where a separation of scales exists, and the optimal prediction approximation coincides with the method of averaging. For example, the effective equation (5.7) in example 5.1 can be derived from the full dynamics (5.6) by optimal prediction, in certain situations. Specifically, if ν is chosen to be the micro-canonical measure restricted to an

energy shell, then the 'unresolved' term η^2 is replaced by its ensemble average over the micro-canonical ensemble with fixed energy, E, conditioned by the given values of the 'resolved' variables, (x, p).

Illustration 12.1. The following example is taken from [19]. Consider the system of equations

$$\begin{aligned} \frac{\mathrm{d}q_1}{\mathrm{d}t} &= p_1, \\ \frac{\mathrm{d}p_1}{\mathrm{d}t} &= -q_1(1+q_2^2), \\ \frac{\mathrm{d}q_2}{\mathrm{d}t} &= p_2, \\ \frac{\mathrm{d}p_2}{\mathrm{d}t} &= -q_2(1+q_1^2), \end{aligned} \tag{12.5}$$

which is a Hamiltonian system with the Hamiltonian

$$H(p_1, q_1, p_2, q_2) = \tfrac{1}{2}(p_1^2 + p_2^2 + q_1^2 + q_2^2 + q_1^2 q_2^2).$$

The initial conditions are assumed to be randomly drawn from a canonical distribution with unit temperature, i.e. from a distribution with density proportional to e^{-H}. At time $t = 0$ only $(q_1(0), q_2(0))$ are given, and the goal is to calculate their average over all sets of initial conditions that are compatible with the initial data, $\mathbb{E}[q_1(t) \mid q_1(0), p_1(0)]$ and $\mathbb{E}[p_1(t) \mid q_1(0), p_1(0)]$. The optimal prediction equations for this system can be derived explicitly: $Q_1(t) \approx \mathbb{E}[q_1(t) \mid q_1(0), p_1(0)]$ and $P_1(t) \approx \mathbb{E}[p_1(t) \mid q_1(0), p_1(0)]$ are governed by

$$\begin{aligned} \frac{\mathrm{d}Q_1}{\mathrm{d}t} &= P_1, \\ \frac{\mathrm{d}P_1}{\mathrm{d}t} &= -Q_1\left(1 + \frac{1}{1+Q_1^2}\right). \end{aligned}$$

In figure 18, we compare $\mathbb{E}[q_1(t) \mid q_1(0), p_1(0)]$ and the optimal prediction approximation $Q_1(t)$. The former was generated by averaging over a large ensemble of solutions with initial conditions drawn from the canonical distribution, conditioned by the partial initial data at hand $q_1(0) = 1$, $p_1(0) = 0$. The graph shows that optimal prediction is accurate, in this case, for short times only. Note the decay of the true average, due to the 'dephasing' of the solutions in the ensemble, in contrast with the periodic behaviour of the optimal prediction solution. This periodic behaviour reflects the fact that, for Hamiltonian systems and canonical measure ν, the optimal prediction equations are themselves Hamiltonian [19]. Various methods of incorporating memory effects into the optimal prediction framework, to overcome this difficulty, were considered in [19]. ■

12.2. The moment map

In its most basic form, the aim of this approach is, as for optimal prediction, to produce algorithms for X defined by (12.2), given (12.1). The methods generalize to random driving, as for optimal prediction. In this basic form, the moment map is

$$X_{n+1} = \Gamma(X_n), \tag{12.6}$$

where

$$\Gamma(X) = \int \rho(x, y, t; X)x \,\mathrm{d}x \,\mathrm{d}y \tag{12.7}$$

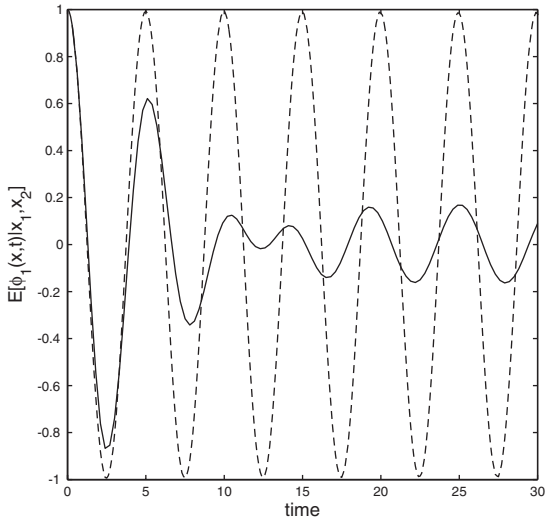

Figure 18. Comparison between $\mathbb{E}[q_1(t) \mid q_1(0), p_1(0)]$ (——) and $Q_1(t)$ (- - - -) for initial data $q_1(0) = 1$, $p_1(0) = 0$.

and where ρ satisfies the Liouville equation

$$\frac{\partial \rho}{\partial t} + \nabla_x \cdot (f\rho) + \nabla_y \cdot (g\rho) = 0 \tag{12.8}$$

with initial data given by an invariant measure ν, conditional on X. Notice that the map contains a parameter t which needs to be selected.

The motivation for the approach is that, if the time t map induced by the flow of the vector field F given by (12.4) can be approximated numerically, a variety of algorithms from computational bifurcation theory can be used to compute families of steady solutions, their stability, periodic solutions and so forth. However, in this basic form the method suffers from the limitations of optimal prediction outlined in the previous subsection. But it is possible, under certain circumstances, that stochastic equations for z may have a choice of coarse-grained variables x, for which X is effectively deterministic. The model problems of section 8 give examples of this set-up, and the algorithm has been used in this context. Also for more complex systems such as lattice Boltzmann models for fluids, the idea of using moments to represent the system is physically natural, and algorithmically successful. Note that for these problems in fluid mechanics, several moments are propagated and coupled together.

In situations where the approximation (12.3), (12.4) or (12.6), (12.7) fails, the approach in [134] proposes a rational closure scheme, in contrast to the somewhat *ad hoc* closure proposed in [133]. The idea in [134] is to propagate a number of moments of $x(t)$ rather than just the mean. Let

$$X^{(j)} = \mathbb{E}[x(t) \otimes x(t) \otimes \cdots \otimes x(t) \mid x(0) = x_0], \qquad j = 1, \cdots, k,$$

where the tensor product is over j terms. We then let $\rho(x, y, t; X^{(1)}, \ldots, X^{(k)})$ denote the probability density function for (1.2), started from measure ν, conditional on knowing the first k moments of $x(0)$. The natural generalization of (12.6), (12.7) is the map

$$X_{n+1}^{(j)} = \Gamma^{(j)}(X_n^{(1)}, \ldots, X_n^{(k)}),$$

where

$$\Gamma^{(j)}(X^{(1)}, \ldots, X^{(k)}) = \int \rho(x, y, t; X^{(1)}, \ldots, X^{(k)}) x \otimes x \otimes \cdots \otimes x \, \mathrm{d}x \, \mathrm{d}y.$$

(Again the tensor product involves j contributions.) By increasing k this gives a rational way of improving the approximation underlying equations (12.3), (12.4) because, under certain regularity assumptions, the moments of x do form a basis for the master equation. There is, however, no general analysis to determine the appropriate number of moments to propagate in order to get good approximations to the Liouville equation with a small number of moments. In principle, it is possible to add moments until the numerical results are unaffected by further addition. But in practice it is prohibitively expensive to propagate more than a few moments. The use of this method in more complex problems, where closure in the dynamics of a small number of moments may not occur, is far from understood, but some numerical evidence is encouraging. For example, the experiments in [135] show that the method works well in example 7.3 where the fixed points of the moment map may be identified with the metastable states which arise when transitions between the wells of V are rare. A simpler example, also studied in [135], is example 9.1. This problem does have a scale-separation, as detailed in section 9, but not of the explicit form exploited in the algorithms of section 10. Rather the scale-separation leads to a partition of the phase space, based on the potential wells, with a slow Markov chain between elements of the partition. The moment map, which nonlinearizes the linear flow of the Fokker–Planck equation (2.8), has fixed points in each potential well. For example, if the mean and variance matrix are propagated in the moment map, then fixed points are introduced corresponding to Gaussian approximations of the invariant measure, restricted to a particular well of the potential. There is related work on the stabilization of metastable states in [136].

The approach outlined here may yield considerable savings when embedded in bifurcation or continuation software. One approach to coarse time-stepping is introduced in [137], with applications to problems of the type described in section 8 covered, for example, in [134, 138]. Application of related algorithmic ideas to the dynamics of a biomolecule may be found in [139]. The approach is studied in the context of the heat bath examples of section 7.3, in [135].

Another possible saving arises if the computational approximation to $F(X)$ found by integrating the full system over step t, is used to propagate the system through time $T > t$, using

$$X_{n+1} = X_n + T F(X_n).$$

This idea is closely related to the projective integration method; see illustrations 10.1 and 10.4. A complete understanding of the stability and consistency issues associated with this method has not yet been developed; for some initial investigations in this direction see [103]. As mentioned earlier, there are similarities with the heterogeneous multiscale method, and it is likely that rigorous justification of the approach would require analysis similar to that developed in [118]. The papers [137, 140, 141] show applications of ideas similar to those outlined here, but for infinite-dimensional problems.

Illustration 12.2. To illustrate the method we consider the SDE

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -x(x^2 - \mu) + \nu + \frac{\mathrm{d}U}{\mathrm{d}t}, \tag{12.9}$$

where U is standard Brownian motion and where μ and ν are real parameters; the case ν small is of particular interest. The potential has two local minima for $|\nu| < 2(\mu/3)^{3/2}$ and one global minimum otherwise. When the diffusion coefficient is small relative to the escape times from the wells, it is natural to try and find a reduction of the stochastic dynamics which captures the behaviour within wells, and between wells. We employ the simplest version of the moment map, which propagates only the mean. Thus the infinite-dimensional linear flow

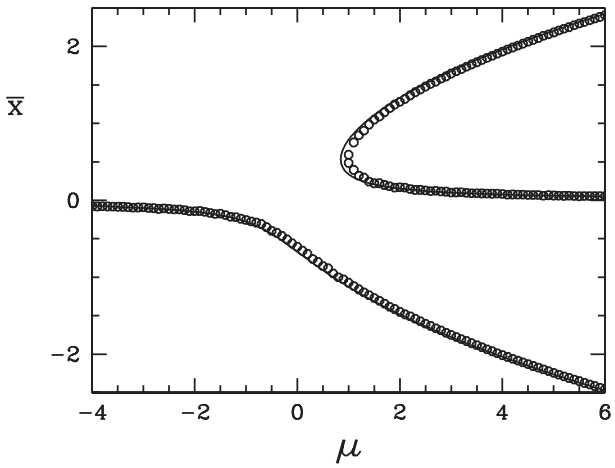

Figure 19. Fixed points of the moment map applied to (12.9) (○) and zeros of the drift (——) as μ varies.

of the Fokker–Planck equation for the SDE is approximated by a one-dimensional nonlinear map. Figure 19, taken from [135], shows the bifurcation diagram for fixed points of this map, as μ varies (with $\nu = 0.3$). The fixed points of the map are, in this case, very close to the zeros of the drift. (Similar features arise if a two-dimensional map, for the mean and variance, is used.) It is apparent that the moment map captures the coarse features of the SDE in that the metastable states of the system, which are centred on the wells of V (shown with solid lines), become stable fixed points of the moment map. Thus the moment map is effective in the identification of metastable states.

Illustration 13.1 shows how the transfer operator approach may be used as an alternative method for identifying metastable states, and that it also extracts a Markov chain describing the dynamics between these states. The question of how to use the moment map to extract effective dynamics between metastable states, for example in finding transition pathways, is still under investigation. ■

12.3. Optimal prediction and the moment map are related

Here we show how the two methods of this section are related to one another. Let $\rho(x, y, t; X)$ denote the probability density function for (12.1), started from measure ν, conditional on $x(0) = X$. In its simplest form the moment map is given by (12.6) and (12.7). We assume smoothness of ρ in t so that

$$\rho(x, y, t; X) \approx \rho(x, y, 0; X) + t\frac{\partial \rho}{\partial t}(x, y, 0; X).$$

Now ρ satisfies the Liouville equation (12.8) and so, for $\rho = \rho(x, y, 0; X)$,

$$\begin{aligned}\Gamma(X) &\approx \int \rho x \,\mathrm{d}x \,\mathrm{d}y - t\int [\nabla_x \cdot (f\rho) + \nabla_y \cdot (g\rho)]x \,\mathrm{d}x \,\mathrm{d}y \\ &= X + t\int f\rho \,\mathrm{d}x \,\mathrm{d}y.\end{aligned}$$

Here we have used the fact that $\rho(x, y, 0; X)$ acts as a delta function $\delta(x - X)$ when integrated against functions of x alone, and the divergence theorem in x and y on the second and third

terms respectively. We also assume that the density ρ decays to zero at infinity. Thus,

$$\frac{\Gamma(X) - X}{t} \approx F(X)$$

with $F(X)$ given by (12.4). Thus (12.6) is seen to be of the form

$$X_{n+1} \approx X_n + tF(X_n),$$

an approximation of (12.3), (12.4) which recovers the solution of these equations in the limit $t \to 0$.

13. Identifying variables

All the algorithms described in sections 10–12 proceed on the assumption that $\mathcal{X}$ is known *a priori*. In many situations this is a reasonable assumption, since a mathematical structure such as scale-separation, or physical reasoning, both help identify $\mathcal{X}$. However, many important applications arise where scale-separation may not occur, or may be hidden; furthermore not all physical problems clearly dictate a choice for $\mathcal{X}$. Thus it is of importance to develop algorithms which both identify $\mathcal{X}$ and find the dynamics within it. This is a much harder problem than simply identifying dynamics within $\mathcal{X}$, with $\mathcal{X}$ known *a priori*; the literature reflects this fact with only limited successes in the area of the identification of $\mathcal{X}$. However, it is to be anticipated that substantial investigation of this topic will ensue, driven by problems of great practical significance, such as the identification of transition pathways in complex chemical systems. Here we describe some of the existing literature.

13.1. Transfer operator approach

This is a method which attempts to find a small and finite state Markov chain within $\mathcal{X}$. In its most basic form, this method is aimed at problems such as those outlined in section 9. The aim is to identify a small finite set of variables, and a Markov chain describing the dynamics between elements of the set. The identification of variables on which the Markov chain lives proceeds through the study of spatial dependence of the eigenfunctions of the transition kernel P of a large Markov chain. For continuous time Markov chains with generator L, $P = \exp(\mathcal{L}t)$. For SDEs such as (1.1), $P = \exp(\mathcal{L}t)$, with $\mathcal{L}$ given by (2.7). It is worth mentioning that the spatial structure of the eigenfunctions of $P = \exp(\mathcal{L}t)$ is also very revealing for the scale-separated model problems of section 6 and for the heat bath problems of section 7 (see [76]). Hence, at least in principle, the transfer operator approach also applies to the reduction of (1.1) to (1.3) in some generality.

In practice it is prohibitively expensive to use the idea as a numerical method, without significant algorithmic sophistication, exploiting, for example, multiscale structures in the problem. The method has its origins in work that has been designed to compute attractors for dissipative dynamical systems [142, 143]. There was then an attempt to lift this work to applications in Hamiltonian mechanics, in particular, to molecular dynamics in [144]. The use of this method in applications to molecular dynamics was fully realized when applied to a constant temperature formulation of molecular dynamics, achieved through randomized momenta, described in [145]. To simplify exposition of the method, we describe an idealization studied in [76]. In practice the algorithm is used in a more complex fashion and details may be found in [93].

Imagine that we are given a sampled time-series $z_n = z(n\Delta) \in \mathcal{Z}$ for some $\Delta > 0$, $z(t)$ solving (1.1). By projecting into $\mathcal{X}$ (assumed known—we are going to identify reduced

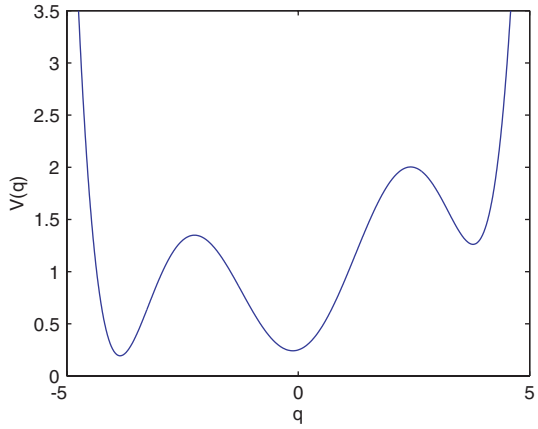

Figure 20. The potential $V(x)$ in illustration 13.1.

variables in $\mathcal{X}$) we find the sequence $x_n \in \mathcal{X}$, $n \in \mathbb{Z}$. From this it is possible to find an empirical Markov chain on some finite partition of $\mathcal{X}$, say $\mathcal{X}^\delta$. The maximum likelihood estimator of this Markov chain simply counts the number of transitions x_n makes from state $i \in \mathcal{X}^\delta$ to $j \in \mathcal{X}^\delta$ as a proportion of all transitions from i. This gives a Markov transition matrix P. The idea of the transfer operator approach is to try and extract from this matrix P, a simpler Markov chain on a state space of low dimension. This idea works well, and can be rigorously justified, when the matrix P has a single eigenvalue on the unit circle, necessarily at 1, with a small cluster of eigenvalues, say $m-1$, next to the unit circle, and the rest of the eigenvalues separated by an order one amount (measured in terms of the nearness of the $m-1$ dominant eigenvalues) from the unit circle; such systems are termed metastable. It is then possible to find a Markov chain on an m-dimensional state space which accurately approximates the coarse dynamics of the problem—see section 9.1.

A good model problem for the transfer operator approach is a countable state Markov chain with generator L given by (9.1). Then $P = \exp(Lt)$, $t = \mathcal{O}(1)$, will have exactly the desired property for $m = 2$. A useful example for SDEs is example 9.1. The stochastic dynamics are approximated by a two-state Markov chain, describing transitions between the neighbourhoods of ± 1. In [76] various models similar to examples 7.1, 7.2 are used to evaluate the transfer operator approach.

Illustration 13.1. We study an illustrative example from [76]. Consider equation (1.3) with $A(X) = \sqrt{2/\beta}$, $F(X) = -V'(X)$ and $V(X)$ as given in figure 20; thus V has three minima and the invariant density of the SDE is as shown in figure 21 for $\beta = 2.0$. Thus $m = 3$ and this is reflected in the structure of the spectrum for $P = \exp(Lt)$ with $t = 1$; the eigenvalues are found (numerically) to be

λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	...
1.000	0.950	0.915	0.387	0.227	0.125	...

The eigenvectors associated with the three eigenvalues near 1 help identify a coarsening of the state space and facilitate the construction of a three-state Markov chain on this coarsening.

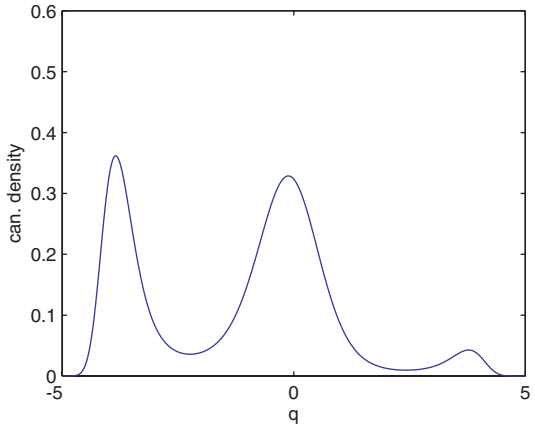

Figure 21. The invariant density for the illustration 13.1.

Note that an alternative method for identifying metastable states is by use of the moment map; see illustration 12.2. ■

13.2. SVD-based techniques

This is a method which attempts to identify $\mathcal{X}$, as well as to find approximate dynamics in $\mathcal{X}$. Consider the deterministic version of (1.1), with $\gamma \equiv 0$. The basic idea is to observe a single path of (1.1), use a singular value decomposition (SVD or proper orthogonal decomposition—POD) to extract dominant modes, and then project the equation (1.1) onto them. We assume that $\mathcal{Z}$ has finite dimension d, and try to identify a space $\mathcal{X}$ of dimension $k < d$.

Let

$$Z = [z(t_1), z(t_2), \ldots, z(t_N)] \in \mathbb{R}^{d\times N}$$

be a matrix formed from a large number of samples of a single path of (1.1). The (reduced) SVD factorizes Z as

$$Z = U\Sigma V^*,$$

where U, $\Sigma \in \mathbb{R}^{d\times d}$ and $V \in \mathbb{R}^{N\times d}$. By retaining only k columns of U (U_k) and k rows of Σ (Σ_k) we make the approximation

$$Z \approx U_k \Sigma_k V_k^*,$$

where U_k^* projects from $\mathcal{Z}$ into a low-dimensional subspace of dimension k which we identify with $\mathcal{X}$. Thus $U_k \in \mathbb{R}^{d\times k}$, $\Sigma_k \in \mathbb{R}^{k\times k}$ and $V_k \in \mathbb{R}^{N\times k}$. Here Σ is a diagonal matrix and we assume that the singular values are ordered in decreasing fashion, from left to right. Then Σ_k contains the k leading singular values of Σ. Hence, the dynamics in (1.1) is approximated by

$$\dot{\xi}(t) = U_k^*(U_k\xi(t)). \tag{13.1}$$

To understand how the method works, note that

$$ZZ^T = U\Sigma\Sigma U^*$$

so that the columns of U comprise the eigenvectors of ZZ^T. The diagonal entries of Σ (which is a diagonal matrix) are the square roots of the eigenvalues. The entries of ZZ^T comprise correlation information between *different* components of z, averaged over time.

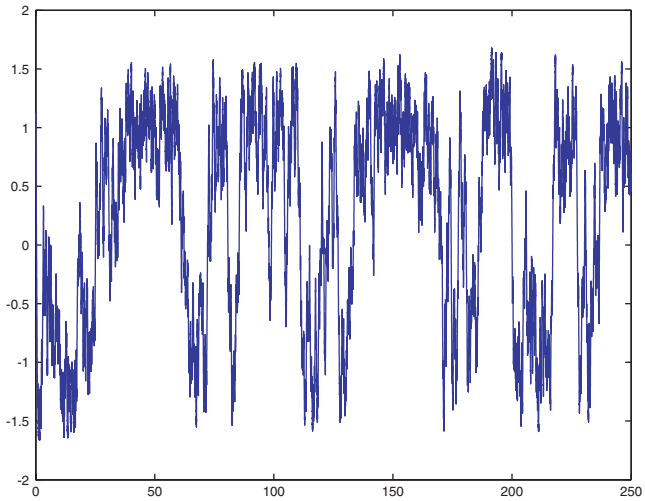

Figure 22. Sample path of x from (6.22) with $\epsilon = 0.25$ showing effective stochastic behaviour.

The dominant eigenvalues, and associated eigenvectors, isolate a coordinate basis which highlights the directions which maximize this correlation. However, the seed of the failure of the algorithm can be seen from this description: the data is found by time-averaging and correlation information *in time* is completely lost. More sophisticated data analysis, exploiting spectral representation theorems for stationary Gaussian processes, would employ correlation information in time and perform approximation in the Kahunen–Loeve basis.

Illustration 13.2. We consider example 6.2. Recall that, if the variable x is taken to characterize $\mathcal{X}$ then we can fit the SDE (6.23) to time-series data from (6.22); an efficient method for doing this is outlined in illustration 10.5. We apply the SVD-based method to see what happens if it tries to identify a one-dimensional subspace $\mathcal{X}$.

Note immediately that the method cannot capture the stochastic dynamics of (6.23) because if $k = 1$ then (13.1) is a one-dimensional ODE with trajectories which, provided bounded, converge to the set of equilibria. The SVD method has a further drawback: it does not even identify the coordinate x as significant. Figure 22 shows x solving (6.22) with $\epsilon = 0.25$ on the time interval $t \in [0, 250]$. The stochastic behaviour of the limit equation (6.23) is evident in the component x. However the SVD method applied to the complete time-series (here $d = 4$, $N = 1 + 10^5$ and $k = 1$) identifies $\mathcal{X}$ as span$\{26y_1 + 27y_2 + 10\,000y_3 + 38x\}$ which is clearly a variable dominated by y_3 and hardly affected by x at all. (The numbers obtained are very sensitive to the precise data used, but the conclusion that y_3 dominates is robust.) ■

The previous illustration touched on an issue of some importance, namely that dimension reduction by the SVD method may constrain the dynamics in a manner which makes complex behaviour in the original system impossible in the reduced one. This idea is nicely illustrated for the Lorenz equations in [146]. For applications of this SVD-based approach in fluid mechanics see [147] and [148]; for applications in molecular dynamics see [149]. The usefulness of this method is limited by the fact that the low-dimensional basis, onto which the solution is projected, is calculated from information which is global in time. Nonetheless, information from PODs is still used in a variety of situations to identify an appropriate choice for the subspace $\mathcal{X}$ in situations where it is not identifiable *a priori*.

13.3. Model reduction

This is a method which attempts to identify $\mathcal{X}$, as well as to find approximate dynamics in $\mathcal{X}$. Here the usual application domain is control theory, and most work to date concerns linear systems. With this in mind we set

$$h(z) = Az, \qquad \gamma(z) = B, \qquad \frac{\mathrm{d}W}{\mathrm{d}t}(t) = u(t)$$

in equation (1.1) and obtain

$$\frac{\mathrm{d}z}{\mathrm{d}t} = Az + Bu.$$

We assume that the object of interest is a linear function of z:

$$\eta = Cz.$$

The objective of model reduction is to find $\hat{A}$, $\hat{B}$, $\hat{C}$ so that the reduced dynamics

$$\frac{\mathrm{d}X}{\mathrm{d}t} = \hat{A}X + \hat{B}u,$$
$$\eta' = \hat{C}X$$

provides a good approximation η' to η, for a range of controllers u. In the (x, y) picture of equations (1.2), this corresponds to finding coordinates in which the y variable can be effectively eliminated without introducing memory. On the assumption that A is negative-definite, by use of the Laplace transform, the question reduces to finding $\hat{A}$, $\hat{B}$, $\hat{C}$ so that $\eta'_L(s)$ is a good approximation to $\eta_L(s)$, where

$$\eta_L(s) = C(sI - A)^{-1}B,$$
$$\eta'_L(s) = \hat{C}(sI - \hat{A})^{-1}\hat{B}.$$

This problem in approximation theory can be tackled in a number of different ways, depending on the range of s over which good approximation is required. Two basic approaches are Krylov subspace (moment matching) and SVD-based (see [150–152]).

It would be of interest to extend these ideas to systems, such as those in section 7.3, where stochastic effects arise when eliminating variables. This would occur, for example, if A is skew, contrasting with most of the existing work on model reduction where A is negative definite.

14. Miscellaneous

In this paper we confine ourselves primarily to ODEs. However related issues arise, of course, for infinite-dimensional problems, including PDEs, and are both of pressing importance from the viewpoint of making optimal use of computational resources, and from the point of view of applications. It would be very hard to do justice to the breadth of literature in this context, and we limit ourselves to a few papers which themselves lead to the wider literature. The very general approach outlined in [118] includes PDEs, with applications across a range of problems. For PDEs with multiple-scales there is interesting recent work, using finite elements built on micro-structure, which addresses spatial issues analogous to the temporal issues considered in this paper; see for example [153, 154]. See [155] for a recent review on multiscale methods in general, especially in the context of PDEs. An interesting link between variable reduction and the renormalization group is pointed out in [156].

Application areas which draw on the ideas outlined in this review are numerous: molecular dynamics, materials science and fluid mechanics, to name a few.

Molecular models based on Newtonian dynamics typically involve a variety of force fields which induce a wide range of time-scales. The bond stretch and bond angle potentials induce fast vibrations, whereas their interaction with slower forces, such as electrostatic Coulomb interaction, and van der Waals repulsion lead, for example, to complex biomolecular conformations, such as protein folding, on very long time-scales. Thus, the elimination of fast scales, and the prediction of macroscopic information about molecular conformations, and transitions between them, is a major goal; the topics of this review impinge on molecular dynamics in many different ways. The use of heat baths through random frequencies has been attempted in molecular dynamics simulations—see [157–159]. The idea of fitting SDEs to partially observed dynamics in the context of biomolecular dynamics was undertaken in [124]. A natural idea is to try and eliminate the fast scales by use of averaging techniques, like those in section 5, to produce constrained mechanical systems (see [32, 36, 160]). However, for molecules in thermal contact with a heat bath, there is a range of physics literature suggesting other effective smoothed potentials, consistent with the macrocanonical ensemble; see [161–163]. A prevalent approach in this field is to use different time-steps for the fast and slow contributions to the force; see [113] and chapter XIII of [99].

In materials science an active and open area of research concerns the interfacing of molecular and continuum models. A concrete example where this is of interest is the study of crack propagation where detailed molecular information is required near the crack tip, but where a continuum description suffices in the far field. Since continuum models are essentially macroscopic, and molecular models microscopic, the issue of extraction of macroscopic dynamics is the key issue. The topic is a large one and there is not sufficient space to provide a full bibliography. An important early reference is [164] and recent papers with a more mathematical slant include [165, 166].

The atmospheric sciences provide a wealth of applications in fluid mechanics where the extraction of macroscopic information is key [52, 58, 59, 123]. Another important application area is non-Newtonian fluid mechanics where, increasingly, stochastic models are used to describe the microscopic information required to represent macroscopic quantities such as the stress tensor [167–170].

Acknowledgments

This paper is based on the 2002 Ron Diperna Memorial Lecture, given by AMS at the Mathematics Department, University of California, Berkeley, 7 February 2002. The authors are grateful to Xinyu He for helping with some preliminary numerical calculations and to Zvi Artstein, Alexandre Chorin, Robert Krasny, Christian Lubich, Weinan E, Greg Pavliotis, Sebastian Reich, Christof Schütte, Paul Tupper and Petter Wiberg for helpful comments on a preliminary draft. The authors are particularly grateful to Eric Vanden Eijnden who made a number of significant suggestions about how to structure the paper. DG and RK are supported in part by the Israel Science Foundation founded by the Israel Academy of Sciences and Humanities, and by the Applied Mathematical Sciences subprogram of the Office of Energy Research of the US Department of Energy under Contract DE-AC03-76-SF00098. AMS is supported by the EPSRC (UK).

References

[1] Norris J 1997 *Markov Chains* (Cambridge: Cambridge University Press)
[2] Risken H 1984 *The Fokker-Planck Equation* (New York: Springer)
[3] Gardiner C W 1985 *Handbook of Stochastic Methods* 2nd edn (New York: Springer)

[4] Øksendal B 1998 *Stochastic Differential Equations* 5th edn (Berlin: Springer)
[5] Rogers L C G and Williams D 2000 Diffusions, Markov processes and martingales *Itô Calculus* 2nd edn, vol 2 (Cambridge: Cambridge University Press)
[6] van Kampen N G 1985 Elimination of fast variables *Phys. Rep.* **124** 69–160
[7] Papanicolaou G C 1974 Introduction to the asymptotic analysis of stochastic equations *Modern Modeling of Continuum Phenomena* ed R C DiPrima (Providence, RI: AMS)
[8] Papanicolaou G C 1978 Asymptotic analysis of stochastic equations *Studies in Probability Theory* vol 18 *Studies in Mathematics* ed M Rosenblatt (Math. Assoc. Am.)
[9] Zwanzig R 1980 Problems in nonlinear transport theory *Systems Far from Equilibrium* ed L Garrido (New York: Springer) pp 198–225
[10] Evans D and Morriss G 1990 *Statistical Mechanics of Nonequilibrium Liquids* (London: Academic)
[11] Mori H 1965 Transport, collective motion, and Brownian motion *Prog. Theor. Phys.* **33** 423–50
[12] Zwanzig R 1973 Nonlinear generalized Langevin equations *J. Stat. Phys.* **9** 215–20
[13] Mori H, Fujisaka H and Shigematsu H 1974 A new expansion of the master equation *Prog. Theor. Phys.* **51** 109–22
[14] Givon D, Hald O H and Kupferman R 2004 Existence of orthogonal dynamics *Israel J. Math.* in press
[15] Vanden-Eijnden E and Greco A 1998 Stochastic modelling of turbulence and anomalous transport in plasmas *J. Plasma Phys.* **59** 683–94
[16] Kramer P R, Majda A J and Vanden-Eijnden E 2003 Testing approximate closures for turbulent diffusion on some model flows *J. Stat. Phys.* **111** 565–679
[17] Just W, Kantz H, Rödenbeck C and Helm M 2001 Stochastic modelling: replacing fast degrees of freedom by noise *J. Phys.* A: *Math. Gen.* **34** 3199–213
[18] Just W, Gelfert K, Baba N, Riegert A and Kantz H 2003 Elimination of fast chaotic degrees of freedom: on the accuracy of the Born approximation *J. Stat. Phys.* **112** 277–92
[19] Chorin A J, Hald O H and Kupferman R 2002 Optimal prediction with memory *Physica* D **166** 239–57
[20] Rossler O E 1976 An equation for continuous chaos *Phys. Lett.* **35a** 397–8
[21] O'Malley R E 1991 *Singular Perturbation Methods for Ordinary Differential Equations* (New York: Springer)
[22] Tikhonov A N, Vasiléva A B and Sveshnikov A G 1985 *Differential Equations* (Berlin: Springer)
[23] Carr J 1980 *Applications of Centre Manifold Theory* (New York: Springer)
[24] Wiggins S 1990 *Introduction to Applied Nonlinear Dynamical Systems and Chaos* (New York: Springer)
[25] Kreiss H-O 1992 Problems with different time-scale s *Acta Numer.* (Cambridge: Cambridge University Press)
[26] Constantin P, Foias C, Nicolaenko B and Temam R 1994 *Integral Manifolds and Inertial Manifolds* (New York: Springer)
[27] Hale J K 1988 Asymptotic behavior of dissipative systems *Mathematical Surveys and Monographs* vol 25 (Providence, RI: AMS)
[28] Teman R 1999 *Infinite Dimensional Dynamical Systems* 2nd edn (New York: Springer)
[29] Wells J C 1976 Invariant manifolds of nonlinear operators *Pac. J. Math.* **62** 285–93
[30] Fenichel N 1971 Persistence and smoothness of invariant manifolds for flows *Indiana Univ. Math. J.* 193–226
[31] Fenichel N 1979 Geometric singular perturbation theory for ordinary differential equations *J. Diff. Eqns* **31** 53–98
[32] Rubin H and Ungar P 1957 Motion under a strong constraining force *Commun. Pure Appl. Math.* **10** 65–87
[33] Neistadt A I 1984 The separation of motions in systems with rapidly rotating phase *J. Appl. Math. Mech.* **48** 511–17
[34] Bennetin G, Galgani L and Giorgilli A 1987 Realization of holonomic constraints and freezing of high frequency degrees of freedom in the light of classical perturbation theory: Part I *Commun. Math. Phys.* **113** 87–103
[35] Bornemann F 1998 *Homogenization in Time of Singularly Perturbed Mechanical Systems* (New York: Springer)
[36] Bornemann F A and Schütte Ch 1997 Homogenization of Hamiltonian systems with a strong constraining potential *Physica* D **102** 57–77
[37] Bender C M and Orszag S A 1999 *Advanced Mathematical Methods for Scientists and Engineers* (New York: Springer)
[38] Takens F 1980 Motion under the influence of a strong constraining force *Global Theory of Dynamical Systems (Evanston, 1979)* ed Z Nitecki and C Robinson (Berlin: Springer)
[39] Sanders J A and Verhulst F 1985 *Averaging Methods in Nonlinear Dynamical Systems* (New York: Springer)
[40] Arnold V I 1989 *Mathematical Methods of Classical Mechanics* 2nd edn (New York: Springer)
[41] Lochak P and Meunier C 1988 *Multiple Phase Averaging for Classical Systems* (New York: Springer)
[42] MacKay R S 2003 Slow manifolds *Energy Localisation and Transfer* ed T Dauxois *et al* (Singapore: World Scientific)

[43] Papanicolaou G C 1976 Some probabilistic problems and methods in singular perturbations *Rocky Mtn. J. Math.* **6** 653–73
[44] Kifer Y 1992 Averaging in dynamical systems and large deviations *Invent. Math.* **110** 337–70
[45] Kifer Y 1995 Limit theorems for averaging in dynamical systems *Ergod. Theory Dynam. Sys.* **15** 1143–72
[46] Kifer Y 2001 Averaging and climate models *Prog. Prob.* **49** 171–88
[47] Kifer Y 2001 Stochastic versions of Anosov's and Neistadt's theorems on averaging *Stochast. Dynam.* **1** 1–21
[48] Artstein Z 2002 On singularly perturbed ordinary differential equations with measure-valued limits *Math. Bohem.* **127** 139–52
[49] Artstein Z and Slemrod M 2001 On singularly perturbed retarded functional differential equations *J. Diff. Eqns* **171** 88–109
[50] Artstein Z and Vigodner A 1996 Singularly perturbed ordinary differential equations with dynamic limits *Proc. R. Soc. Edinb.* **126A** 541–69
[51] Kurtz T G 1973 A limit theorem for perturbed operator semigroups with applications to random evolutions *J. Funct. Anal.* **12** 55–67
[52] Majda A J, Timofeyev I and Vanden-Eijnden E 2001 A mathematical framework for stochastic climate models *Commun. Pure Appl. Math.* **LIV** 891–947
[53] Mao X 1997 *Stochastic Differential Equations and Applications* (Chichester: Horwood)
[54] Papanicolaou G C and Kohler W 1974 Asymptotic theory of mixing stochastic ordinary differential equations *Commun. Pure Appl. Math* **XXVII** 641–68
[55] Katok A and Hasselblatt B 1995 *Introduction to the Modern Theory of Dynamical Systems* (Cambridge: Cambridge University Press)
[56] Majda A and Timofeyev I 2000 Remarkable statistical behavior for truncated Burgers–Hopf dynamics *Proc. Natl Acad. Sci. USA* **97** 12413–17
[57] Vanden-Eijnden E 2003 Numerical techniques for multi-scale dynamical systems with stochastic effects *Commun. Math. Sci.* **1** 377–84
[58] Majda A J, Timofeyev I and Vanden-Eijnden E 1999 Models for stochastic climate prediction *Proc. Natl Acad. Sci. USA* **96** 14687–91
[59] Majda A J, Timofeyev I and Vanden-Eijnden E 2003 Systematic strategies for stochastic mode reduction in climate *J. Atmos. Sci.* **60** 1705–22
[60] Beck C 1990 Brownian motion from deterministic dynamics *Physica* A **169** 324–36
[61] Majda A J, Timofeyev I and Vanden-Eijnden E 2002 *A priori* tests of a stochastic mode reduction strategy *Physica* D **170** 206–52
[62] Givon D and Kupferman R 2003 White noise limits for discrete dynamical systems driven by fast deterministic dynamics *Physica* A submitted
[63] Papanicolaou G C and Varadhan S R S 1973 A limit theorem with strong mixing in Banach space and two applications to stochastic differential equations *Commun. Pure Appl. Math.* **XXVI** 497–524
[64] Khasminskii R Z 1963 Principle of averaging for parabolic and elliptic differential equations and for Markov processes with small diffusion *Theory. Probab. Appl.* **8** 1–21
[65] Khasminskii R Z 1966 A limit theorem for the solutions of differential equations with random right-handed sides *Theory. Probab. Appl.* **11** 390–406
[66] Karlin S and Taylor H M 1975 *A First Course in Stochastic Processes* (New York: Academic)
[67] Kahane J P 1985 *Some Random Series of Functions* (Cambridge: Cambridge University Press)
[68] Kupferman R, Stuart A M, Terry J R and Tupper P F 2002 Long term behaviour of large mechanical systems with random initial data *Stochast. Dynam.* **2** 533–62
[69] Billingsley P 1968 *Convergence of probability measures* (New York: Wiley)
[70] Krylov N V 1995 Introduction to the theory of diffusion processes *AMS Translation of Monographs* vol 142 (Providence, RI: AMS)
[71] Cano B, Stuart A M, Süli E and Warren J O 2001 Stiff oscillatory systems, delta jumps and white noise *Found. Comput. Math.* **1** 69–100
[72] Ford G W and Kac M 1987 On the quantum Langevin equation *J. Stat. Phys.* **46** 803–10
[73] Ford G W, Kac M and Mazur P 1965 Statistical mechanics of assemblies of coupled oscillators *J. Math. Phys.* **6** 504–15
[74] Stuart A M and Warren J O 1999 Analysis and experiments for a computational model of a heat bath *J. Stat. Phys.* **97** 687–723
[75] Tropper M M 1977 Ergodic properties and quasideterministic properties of finite-dimensional stochastic systems *J. Stat. Phys.* **17** 491–509
[76] Huisinga W, Schütte C and Stuart A M 2003 Extracting macroscopic stochastic dynamics: model problems *Commun. Pure Appl. Math.* **56** 234–69

[77] Jakšić V and Pillet C-A 1997 Ergodic properties of the non-Markovian Langevin equation *Lett. Math. Phys.* **41** 49–57
[78] Bianucci M and Mannella R 1996 Linear response of Hamiltonian chaotic systems as a function of the number of degrees of freedom *Phys. Rev. Lett.* **77** 1258–61
[79] Lindenberg K and Seshadri V 1981 Dissipative contributions of internal multiplicative noise: I. Mechanical oscillator *Physica* A **109** 483–99
[80] Kupferman R and Stuart A M 2004 Fitting SDE models to nonlinear Kac–Zwanzig heat bath models *Physica* D in press
[81] Hald O H and Kupferman R 2002 Asymptotic and numerical analyses for mechanical models of heat baths *J. Stat. Phys.* **106** 1121–84
[82] Kupferman R 2004 Fractional kinetics in Kac–Zwanzig heat bath models *J. Stat. Phys.* **114** 291–326
[83] Gillespie D T 1976 A general method for numerically simulating the stochastic time evolution of coupled chemical reactions *J. Comput. Phys.* **22** 403–34
[84] Gillespie D T 1977 Exact stochastic simulation of coupled chemical reactions *J. Phys. Chem.* **81** 2340–61
[85] Kurtz T G 1976 Limit theorems and diffusion approximations for density dependent Markov chains *Math. Prog. Stud.* **5** 67
[86] Kurtz T G 1978 Strong approximation theorems for density dependent Markov chains *Stochast. Proc. Appl.* **6** 223
[87] Fox R F and Keizer J 1991 Amplification of intrinsic fluctuations by chaotic dynamics in physical systems *Phys. Rev.* A **43** 1709–20
[88] Haselwandter C and Vvedensky D 2002 Fluctuations in the lattice gas for Burgers' equation *J. Phys. A: Math. Gen.* **35** 579–84
[89] Vvedensky D 2003 Edwards–Wilkinson equation from lattice transition rules *Preprint* Physics Department, Imperial College
[90] Cardy J Field theory and nonequilibrium statistical physics *Notes* http://www.thphys.physics.ox.ac.uk/users/JohnCardy/home.html
[91] Rezakhanlou F 1996 Kinetic limits for a class of interacting particle systems *Probab. Theory Rel. Fields* **104** 97–146
[92] Rezakhanlou F 2004 Kinetic limits for interacting particle systems *Springer Lecture Notes in Mathematics* at press
[93] Deuflhard P, Huisinga W, Fischer W and Schütte C 2000 Identification of almost invariant aggregates in reversible nearly uncoupled Markov chains *Linear Algebra Appl.* **315** 39–59
[94] Courtois P J 1977 *Decomposability: Queuing and Computer System Applications* (New York: Academic)
[95] Freidlin M I and Wentzell A D 1984 *Random Perturbations of Dynamical Systems* (New York: Springer)
[96] Nipp K and Stoffer D 1995 Invariant manifolds and global error estimates of numerical integration schemes applied to stiff systems of singular perturbation type: I. RK methods *Numer. Math.* **70** 245–57
[97] Nipp K and Stoffer D 1996 Invariant manifolds and global error estimates of numerical integration schemes applied to stiff systems of singular perturbation type: II. Linear multistep methods *Numer. Math.* **74** 305–23
[98] Nipp K and Stoffer D 1992 Attractive invariant manifolds for maps: existence, smoothness and continuous dependence on the map *Research Report* 92-11, SAM, ETH, Zurich
[99] Hairer E, Lubich C and Wanner G 2002 *Geometric Numerical Integration* (New York: Springer)
[100] Hairer E, Lubich C and Roche M 1989 *The Numerical Solution of Differential-Algebraic Systems by Runge–Kutta Methods* (New York: Springer)
[101] Hairer E and Wanner G 1996 *Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems* (New York: Springer)
[102] Stuart A M and Humphries A R 1996 *Dynamical Systems and Numerical Analysis* (Cambridge: Cambridge University Press)
[103] Gear C W and Kevrekidis I G 2003 Projective methods for stiff differential equations: problems with gaps in their eigenvalue spectrum *SIAM J. Sci. Comput.* **24** 1091–106
[104] E W 2003 Analysis of heterogeneous multiscale method for ordinary differential equations *Commun. Math. Sci.* at press
[105] E W, Liu D and Vanden-Eijnden E 2004 Analysis of numerical techniques for multiscale stochastic dynamical systems *Commun. Pure Appl. Math.* at press
[106] Lam S and Goussis D 1994 The csp method for simplifying kinetics *Int. J. Chem. Kin.* **26** 461–86
[107] Maas U and Pope S B 1992 Simplifying chemical kinetics: intrinsic low-dimensional manifolds in composition space *Combust. Flame* **88** 239–64
[108] Deuflhard P and Heroth J 1996 Dynamic dimension reduction in ODE models *Scientific Computing in Chemical Engineering* ed F Keil *et al* (Berlin: Springer)

[109] Kevrekidis I G, Jolly M S and Titi E S 1990 Approximate inertial manifolds for the Kuramoto–Sivashinsky equation: analysis and computations *Physica* D **44** 38–60
[110] Devulder C, Marion M and Titi E S 1993 On the rate of convergence of the nonlinear Galerkin methods *Math. Comput.* **60** 495–514
[111] Garcia-Archilla B, Novo J and Titi E S 2002 Postprocessing Fourier spectral methods: the case of smooth solutions *Appl. Numer. Math.* **43** 191–209
[112] Petzold L, Jay L O and Yen J 1997 Numerical solution of highly oscillatory ordinary differential equations *Acta Numer.* **6** 437–83
[113] Archilla B Garcia, Serna J M Sanz and Skeel R D 1998 Long-time-step methods for oscillatory differential equations *SIAM J. Sci. Comput.* **20** 930–63
[114] Kirchgraber U, Lasagni F, Nipp K and Stoffer D 1991 On the application of invariant manifold theory, in particular to numerical analysis *Int. Ser. Numer. Math.* vol 97 (Basel: Birkhauser) pp 189–97
[115] Sanz-Serna J M and Calvo M P 1994 *Numerical Hamiltonian Problems* (London: Chapman and Hall)
[116] Hairer E, Norsett S and Wanner G 1993 *Solving Ordinary Differential Equations: Nonstiff Problems* (Berlin: Springer)
[117] Tupper P F 2004 Ergodicity and numerical simulation *SIAM J. Numer. Anal.* in press
[118] E W and Engquist B 2003 The heterogenous multiscale method *Commun. Math. Sci.* **1** 87–132
[119] Ascher U and Reich S 1999 The midpoint scheme and variants for Hamiltonian systems: advantages and pitfalls *SIAM J. Sci. Comput.* **21** 1045–65
[120] Schlick T, Mandziuk M, Skeel R D and Srinivas K 1998 Nonlinear resonance artifacts in molecular dynamics simulations *J. Comput. Phys.* **140** 1–29
[121] Kast A P 2000 Optimal prediction of stiff oscillatory mechanics *Proc. Natl Acad. Sci. USA* **97** 6253–7
[122] Tupper P F 2004 A test problem for molecular dynamics integrators *IMA J. Numer. Anal.* in press
[123] Miller R N, Carter E F and Blue S T 1999 Data assimilation into nonlinear stochastic models *Tellus* **51A** 167–94
[124] Grubmüller H and Tavan P 1994 Molecular dynamics of conformational substates for a simplified protein model *J. Chem. Phys.* **101** 5047–57
[125] Rao B L S P 1999 *Statistical Inference for Diffusion Type Processes* (London: Arnold)
[126] Pokern Y, Stuart A M and Wiberg P 2003 Parameter estimation for partially observed hypo-elliptic diffusions *J. R. Stat. Soc. Ser.* B submitted
[127] Cercignani C 1988 *The Boltzmann Equation and its Applications* (New York: Springer)
[128] Xu K and Prendergast K H 1994 Numerical Navier–Stokes solutions from gas kinetic theory *J. Comput. Phys.* **114** 9–17
[129] Chorin A J, Kast A and Kupferman R 1998 Optimal prediction of underresolved dynamics *Proc. Natl Acad. Sci. USA* **95** 4094–8
[130] Chorin A J, Kupferman R and Levy D 2000 Optimal prediction for Hamiltonian partial differential equations *J. Comput. Phys.* **162** 267–97
[131] Chorin A J, Hald O H and Kupferman R 2000 Optimal prediction and the Mori–Zwanzig representation of irreversible processes *Proc. Natl Acad. Sci. USA* **97** 2968–73
[132] Hald O H 1999 Optimal prediction of the Klein–Gordon equation *Proc. Natl Acad. Sci. USA* **96** 4774–9
[133] Bell J, Chorin A J and Crutchfield W 2000 Stochastic optimal prediction with application to averaged Euler equations ed C A Lin *Proc. 7th Natl Conf. on Computer Fluid Mech.* (Taiwan: Pingtung) pp 1–13
[134] Makeev A, Maroudas D and Kevrekidis Y 2002 'Coarse' stability and bifurcation analysis using stochastic simulators: kinetic Monte Carlo examples *J. Chem. Phys.* **116** 10083–91
[135] Barkley D, Kevrekidis I G and Stuart A M 2003 Coarse integration for large scale dynamical systems: a numerical study, in preparation
[136] Hillermeier C, Kunstmann N, Rabus B and Tavan P 1994 Topological feature maps with self-organized lateral connections: a population coded, one-layer model of associative memory *Biol. Cyber.* **72** 103–17
[137] Theodoropoulos K, Qian Y-H and Kevrekidis I G 2000 Coarse stability and bifurcation analysis using timesteppers: a reaction diffusion example *Proc. Natl Acad. Sci.* **97** 9840–3
[138] Makeev A G, Maroudas D, Panagiotopoulos A Z and Kevrekidis I G 2002 Coarse bifurcation analysis of kinetic Monte Carlo simulations: a lattice gas model with lateral interactions *J. Chem. Phys.* **117** 8229–40
[139] Hummer G and Kevrekidis I G 2002 Coarse molecular dynamics of a peptide fragment: free energy, kinetics and long time dynamics computations *J. Chem. Phys.* submitted
[140] Kevrekidis I G, Gear C W, Hyman J M, Kevrekidis P G, Runborg O and Theodoropoulos K 2003 Equation-free coarse-grained multiscale computation: enabling microscopic simulators to perform system-level tasks *Commun. Math. Sci.* **1** 715–62

[141] Gear C W, Kevrekidis I G and Theodoropoulos C 2002 'Coarse' integration/bifurcation analysis via microscopic simulators: micro-Galerkin methods *Comput. Chem. Engng.* **26** 941–63
[142] Dellnitz M and Junge O 1998 An adaptive subdivision technique for the approximation of attractors and invariant measures *Comput. Vis. Sci.* **1** 63–8
[143] Dellnitz M and Junge O 1999 On the approximation of complicated dynamical behaviour *SIAM J. Numer. Anal.* **36** 491–515
[144] Deuflhard P, Dellnitz M, Junge O and Schutte C 1999 Computation of essential molecular dynamics by subdivision techniques *Computational Molecular Dynamics: Challenges, Methods, Ideas* (*Lecture Notes in Computational Science and Engineering*) ed P Deuflhard *et al* (Berlin: Springer)
[145] Schütte C, Fischer A, Huisinga W and Deuflhard P 1999 A direct approach to conformational dynamics based on hybrid Monte Carlo *J. Comput. Phys.* **151** 146–68
[146] Palmer T N 2001 A nonlinear dynamical perspective on model error: a proposal for non-local stochastic-dynamics parameterization in weather and climate prediction models *Quart. J. Met. Soc.* **572** 279–304
[147] Berkooz G, Holmes P and Lumley J 1996 Coherent structures, dynamical systems and symmetry *Cambridge Monograph on Mechanics* (Cambridge: Cambridge University Press)
[148] Sirovich L and Sirovich C H 1989 Low dimensional description of complicated phenomena, in connection between finite and infinity dimensional flows *Cont. Math.* **99** 277–305
[149] Amadei A, Linssen B M and Berendsen H J C 1993 Essential dynamics of proteins *Proteins: Structure, Function, and Genetics* **17** 412–25
[150] Bai Z and Golub G 2002 Computation of large scale quadratic forms and transfer functions using the theory of moments, quadrature and Pade approximation *Modern Methods in Scientific Computing and Applications* (*NATO Science Series*) ed A Bourlioux and M J Gander (Dordrecht: Kluwer)
[151] Freund R W 1999 Reduced-order modelling techniques based on Krylov subspaces and their use in circuit simulation *Appl. Comput. Control Signals Circuits* **1** 435–98
[152] Antoulas A C, Sorensen D C and Gugerrin S 2001 A survey of model reduction methods for large scale systems *Contemp. Math.* (Providence, RI: AMS)
[153] Hughes T J R 1995 Multiscale phenomena: Green's functions, the Dirichlet-to-Neumann formulation, subgrid-scale models, bubbles and the origins of stabilized methods *Comput. Methods Appl. Mech. Eng.* **127** 387–401
[154] Efendiev Y R, Hou T Y and Wu X H 2000 Convergence of a nonconformal multiscale finite element method *SIAM J. Numer. Anal.* **37** 888–910
[155] Brandt A 2001 Multiscale scientific computing: review 2001 *Multiscale and Multiresolution Methods: Theory and Applications* ed T J Barth *et al* (Heidelberg: Springer)
[156] Chorin A J 2003 Averaging and renormalization for the KdV–Burgers equation *Proc. Natl Acad. Sci. USA* **100** 9674–9
[157] Xu D and Schulten K 1994 Coupling of protein motion to electron transfer in a photosynthetic reaction center: investigating the low temperature behaviour in the framework of the spin-boson model *Chem. Phys.* **182** 91–117
[158] Schulten K 1995 Curve crossing in a protein: coupling of the elementary quantum process to motions of the protein *Proc. Ecole de Physique des Houches* ed D Bicout and M J Field (Berlin: Springer) pp 85–118
[159] Damjanovic A, Kosztin I, Kleinekathoefer U and Schulten K 2002 Excitons in a photosynthetic light-harvesting system: a combined molecular dynamics, quantum chemistry and polaron model study *Phys. Rev.* E **65** 031919
[160] van Kampen N G and Lodder J J 1984 Constraints *Am. J. Phys.* **52** 419–24
[161] Fixman M 1974 Classical statistical mechanics of constraints: a theorem and application to polymers *Proc. Natl Acad. Sci. USA* **71** 3050–3
[162] Reich S 1995 Smoothed dynamics of highly oscillatory Hamiltonian systems *Physica* D **89** 28–42
[163] Reich S 2000 Smoothed Langevin dynamics of highly oscillatory systems *Physica* D **138** 210–24
[164] Rudd R E and Broughton J Q 1998 Coarse-grained molecular dynamics and the atomic limit of finite elements *Phys. Rev.* B **58** R5893–6
[165] E W and Huang Z 2001 Matching conditions in atomistic-continuum modeling of materials *Phys. Rev. Lett.* **87**
[166] E W and Huang Z 2003 A dynamic atomistic-continuum method for the simulation of crystalline materials
[167] Ottinger H C 1996 *Stochastic Processes in Polymeric Fluids* (Berlin: Springer)
[168] Muratov C and E W 2002 Theory of phase separation kinetics in polymer-liquid crystal systems *J. Chem. Phys.* **116** 4723
[169] Li T, Vanden-Eijnden E, Zhang P and E W 2003 Stochastic models of complex fluids at small Deborah number *J. Non-Newt. Fluid Mech.* submitted
[170] Le Bris C, Jourdain B and Lelievre T 2004 Existence of solution for a micro-macro model of polymeric fluid: the FENE model *J. Funct. Anal.* **209** 162–93