Extracting Macroscopic Dynamics: Model Problems & Algorithms

Dror Givon [1], Raz Kupferman [1] [2] and Andrew Stuart [3]

Abstract

In many applications, the primary objective of numerical simulation of time-evolving systems is the prediction of macroscopic, or coarse-grained, quantities. A representative example is the prediction of biomolecular conformations from molecular dynamics. In recent years a number of new algorithmic approaches have been introduced to extract effective, lower-dimensional, models for the macroscopic dynamics; the starting point is the full, detailed, evolution equations. In many cases the effective low-dimensional dynamics may be stochastic, even when the original starting point is deterministic.

This review surveys a number of these new approaches to the problem of extracting effective dynamics, highlighting similarities and differences between them. The importance of model problems for the evaluation of these new approaches is stressed, and a number of model problems are described. When the macroscopic dynamics is stochastic, these model problems are either obtained through a clear separation of time-scales, leading to a stochastic effect of the fast dynamics on the slow dynamics, or by considering high dimensional ordinary differential equations which, when projected onto a low dimensional subspace, exhibit stochastic behaviour through the presence of a broad frequency spectrum. Models whose stochastic microscopic behaviour leads to deterministic macroscopic dynamics are also introduced.

The algorithms we overview include SVD-based methods for non-linear problems, model reduction for linear control systems, optimal prediction techniques, asymptotics-based mode elimination, coarse time-stepping methods and transfer-operator based methodologies.

---

[1] Institute of Mathematics, The Hebrew University, Jerusalem, 91904 Israel. `{givon,raz}@math.huji.ac.il`

[2] Department of Mathematics, Lawrence Berkeley Laboratory, Berkeley CA 94720 USA.

[3] Mathematics Institute, Warwick University, Coventry, CV4 7AL, England. `stuart@maths.warwick.ac.uk`

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 $z \in \mathcal{Z}$:

$$\frac{dz}{dt} = h(z) + \gamma(z)\frac{dW}{dt}, \tag{1.1}$$

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 multivariate Brownian motion and (1.1) is a standard Itô stochastic differential equation (SDE). In addition, we will also touch on the case where $\gamma(z)dW(t)/dt$ is replaced by a Poisson counting process $dW(z,t)/dt$, 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 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 co-ordinates x in $\mathcal{X}$ and y in $\mathcal{Y}$ we obtain from (1.1) the coupled SDEs

$$\begin{aligned} \frac{dx}{dt} &= f(x,y) + \alpha(x,y)\frac{dU}{dt} \\ \frac{dy}{dt} &= g(x,y) + \beta(x,y)\frac{dV}{dt} \end{aligned} \tag{1.2}$$

where U,V are again noise processes.

We will study situations where the y variable can be eliminated, and an effective, approximate equation for x alone is obtained. In many cases we will be looking for a stochastic differential equation for $X \in \mathcal{X}$:

$$\frac{dX}{dt} = F(X) + A(X)\frac{dU'}{dt}, \tag{1.3}$$

where $X(t)$ approximates $x(t)$ in a sense to be determined for each class of problems and U' is a noise process. In other cases, where memory must be captured to adequately represent the dynamics in x, 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 discrete time model in $\mathcal{X}$, 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 paths of (1.1); this requires consideration of the Master Equation for probability densities.

The 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 the approaches currently emerging. In order to do this we will spend a substantial fraction of the paper explaining situations in which it is possible to find closed equations for X which adequately approximate the dynamics of $x \in \mathcal{X}$. Thus most 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 primary importance in order to make clear statements about the situations in which we expect 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 and give proofs— we present the essential ideas and reference the literature for details 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 $W(t)$ 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 at the level of the Master Equation. Sections 4–8 describe a variety of situations where an effective equation for the dynamics in $\mathcal{X}$ can be derived. Section 9 is 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 $\mathcal{Z}$.

The following provides an overview of a number of important themes

running throughout this paper.

i) *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 the model (1.1) and/or from physical considerations; in others its determination may be the most taxing part of the algorithm.

ii) *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 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 also describes situations where a memory effect remains, but can be modelled by adding a small number of extra degrees of freedom.

iii) *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, 5, 6, 7, 8 are of the form D-D, D-D, S-S, D-S and S-D respectively where D denotes *Deterministic*, and S denotes *Stochastic*, and the first (resp. second) letter defines the type of dynamics in $\mathcal{Z}$ (resp. $\mathcal{X}$).

iv) *Scale Separation.* Sections 4 – 6 all rely on explicit time-scale separation to achieve the memoryless effect. 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.

2 The Master Equation

In this section we consider dynamical systems of the form (1.1) within a probabilistic setting, by considering the evolution of probability measures induced by the dynamics of paths of (1.1). There are two essential reasons for considering a probabilistic description rather than a pathwise one:

i) Variable reduction is often related to uncertainty in initial data, hence with 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 due 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 in (some components of) a molecular system, since it is a measurable macroscopic quantity, without specifying the exact positions and velocities; this corresponds to specifying a probability measure on the positions and velocities, with variance 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 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 Subsection 2.1 we describe the derivation of the equation governing probability measures for countable state space Markov chains; in Subsection

2.2 we generalize this to the case of Itô SDEs, which give rise to Markov processes on uncountable state spaces.

2.1   Countable State Space

Consider a continuous time Markov chain $z(t)$, $t \geq 0$, taking values in the state space $\mathcal{I} \subseteq \{0, 1, 2, \dots\}$. 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 $t, \Delta t \geq 0$,

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

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,

$$\frac{dp_{ij}}{dt} = \sum_k p_{ik} \ell_{kj}. \tag{2.1}$$

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

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

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

$$\frac{dP}{dt} = PL, \quad P(0) = I. \tag{2.2}$$

Since $P(t) = \exp(Lt)$ solves this problem we see that P and L commute so that $P(t)$ also solves

$$\frac{dP}{dt} = LP, \quad P(0) = I. \tag{2.3}$$

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 $\mu(t) = (\mu_0(t), \mu_1(t), \dots)^T$ 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),

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

where e_i is the ith unit vector, zero in all entries except the ith, 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 $v(t) = (v_0(t), v_1(t), \dots)^T$ denote the vector with i^{th} entry

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

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 $w = (w_0, w_1, \dots)^T$ 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:

$$\frac{dv}{dt} = Lv, \quad v(0) = w. \tag{2.6}$$

Equation (2.6) is the discrete version of the Chapman-Kolmogorov equation described below.

If we set $w_j = j^k$ for $k = \{0, 1, \dots\}$, and denote the solution of (2.6) by $v^{(k)}(t)$ (the k-th 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. 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 generalizes to continuous state space processes, for example those in the next subsection.

This suggests a more a general question: for a given Markov chain on $\mathcal{I}$ starting in state i, do there exist a small number of linear functions of $p(t)$ (i.e., expectation values of functions on $\mathcal{I}$) which, at least approximately, characterize the behaviour of (some) components of the process? This question is at the heart of the model problems and algorithms that we study here.

Some of the algorithms we highlight work in discrete time. Then the analogue of (2.4) is the iteration

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

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 aspects of its behaviour.

2.2 Fokker-Planck and Chapman-Kolmogorov Equations

The concepts introduced in Subsection 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 multidimensional 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, via

$$\nabla\phi = \frac{\partial\phi}{\partial z_i}e_i, \qquad \nabla v = \frac{\partial v_i}{\partial z_j}e_i e_j^T$$

for orthonormal basis $\{e_i\}$ in $\mathbb{R}^d$. The divergence acts on vector valued functions v, or matrix valued functions A via

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

In the preceding we are assuming the Einstein summation convention, whereby repeated indices imply a summation. Below we will use ∇_x (resp. ∇_y) to denote gradient or divergence with respect to x (resp. y) co-ordinates alone.

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

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

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

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

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—$A : B = \text{trace}(A^T B)$. We will also be interested in the operator $\mathcal{L}^*$ defined by

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

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

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

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-Plank equation

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

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 $\mathbb{E}$ expectation with respect to this measure. Whether or not such a smooth density exists depends on the (hypo-) ellipticity properties of $\mathcal{L}$.

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 $v(z_0, t) = \mathbb{E}[w(z(t))|z(0) = z_0]$, 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

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

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

$$e^{\mathcal{L}t}w(y) = \int_{\mathcal{Z}} (e^{\mathcal{L}^*t}\rho_0)(z)w(z) \, dz, \qquad (2.10)$$

where $\rho_0(z) = \delta(z - y)$. This is the analogue of (2.5) in the countable state space case. Indeed, 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 Discussion and Bibliography

A good background reference on Markov chains is Norris [Nor97]. For a discussion of SDEs from the Fokker-Planck viewpoint, see Risken [Ris84] or Gikhman and Skorokhod [GS96] . For a discussion of the generator $\mathcal{L}$, and the Chapman Kolmogorov equation, see Oksendal [Øks98]. For a discussion concerning ellipticity, hypo-ellipticity and smoothness of solutions to these equations see Rogers and Williams [RW00].

3 Mori-Zwanzig Projection Operators

The Mori-Zwanzig formalism is a technique developed for irreversible statistical mechanics to reduce, at least formally, the dimensionality of a system of ODEs. 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

$$\frac{dx(t)}{dt} = \bar{f}(x(t)) + \int_0^t K(x(t-s), s)\, ds + n(x(0), y(0), t). \qquad (3.1)$$

The first term on the right hand side is only a function of of the instantaneous value of x at time t, and therefore represents a "Markovian" term. The second terms 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 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[1], 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 $w(x, y)$ has then 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 $\mathbb{E}[w|x]$. The conditional expectation defines a mapping $w \mapsto \mathscr{P}w = \mathbb{E}[w|x]$ from functions of (x, y) to functions of x. Specifically,

$$(\mathscr{P}w)(x) = \frac{\int_{\mathcal{Y}} \rho(x, y) w(x, y)\, dy}{\int_{\mathcal{Y}} \rho(x, y)\, dy}.$$

With the initial value $y(0)$ 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

---

[1]Not to be confused with the projection P defined in the Introduction

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 $K(x, t)$ satisfy what is known as the *fluctuation-dissipation relation.* Equation (3.1) is often called a *generalized Langevin equation.* Analogously 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.

The derivation of (3.1) is quite abstract, 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 function $x(t)$ is more accurately written as $P\varphi^t(x, y)$, where here (x, y) are the initial data. The x-equation in (1.2) is

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

and, 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 as follows:

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

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

The next stage is to re-arrange the terms in the square brackets in (3.3). Defining the operator $\mathcal{L} = f(x, y) \cdot \nabla_x + g(x, y) \cdot \nabla_y$, the noise function $n(x, y, t)$ is defined as the solution of the orthogonal dynamics equation:

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

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

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

hence (3.3) takes the form (3.1).

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

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

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

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

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

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

known in the Physics literature as Dyson's formula [EM90].

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 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 and 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.

3.1 Discussion and Bibliography

The original derivation of Mori-Zwanzig formalism can be found in Mori [Mor65] and Zwanzig [Zwa73]. 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. Zwanzig in [Zwa80] 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 [MFS74]. The existence

of solutions to the orthogonal dynamics equation (3.4) turns out to be very subtle; see Givon et al. [GHK03]. 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 [VEG98] and a more recent applications in [KMVE03]. Recent uses of the Mori-Zwanzig formalism in the context of variable reduction can be found in Just et al. [JKRH01, JGB$^+$03] and Chorin et al. [CHK02].

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. For such problems the initial conditions in $\mathcal{Y}$ are irrelevant, after, perhaps, an initial transient.

4.1 Theory

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:

$$
\begin{aligned}
f(x, y) &= L_1 x + \hat{f}(x, y), \\
g(x, y) &= L_2 y + \hat{g}(x, y).
\end{aligned}
\tag{4.1}
$$

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 large relative to the size of $\hat{f}$ and $\hat{g}$ (and the argument can be localized by use of cut-off functions), then the existence of an exponentially attractive and 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

$$\frac{dX}{dt} = L_1 X + \hat{f}(X, \eta(X))$$
$$X(0) = x(0).$$

(4.2)

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 assumptions on the initial value of the discarded variables $y(0)$.

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 $\epsilon \ll 1$,

$$\frac{dx}{dt} = f(x, y)$$
$$\frac{dy}{dt} = -\frac{y}{\epsilon} + \frac{\tilde{g}(x)}{\epsilon}.$$

(4.3)

In reference to the spectral properties of L_1, L_2 in (4.1), 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 $f(x, y)$.

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

$$y = \tilde{g}(x) + \epsilon g_1(x) + O(\epsilon^2)$$

gives, substituting into the y-equation in (4.3), expanding to $O(1)$ and using the x-equation in (4.3),

$$\nabla \tilde{g}(x) \cdot f(x, \tilde{g}(x)) = -g_1(x).$$

Thus, up to errors of $O(\epsilon^2)$, the reduced dynamics are

$$\frac{dX}{dt} = f(X, G(X)),$$

(4.4)

$$G(X) = \tilde{g}(X) - \epsilon \nabla \tilde{g}(X) \cdot f(X, \tilde{g}(X))),$$

(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

$$
\begin{aligned}
\frac{dx_1}{dt} &= -x_2 - x_3 \\
\frac{dx_2}{dt} &= x_1 + \frac{1}{5}x_2 \\
\frac{dx_3}{dt} &= \frac{1}{5} + y - 5x_3 \\
\frac{dy}{dt} &= -\frac{y}{\epsilon} + \frac{x_1 x_3}{\epsilon},
\end{aligned}
\tag{4.6}
$$

so that $\mathcal{X} = \mathbb{R}^3$ and $\mathcal{Y} = \mathbb{R}$. The expression (4.5) with $\epsilon = 0$ indicates that, for small ϵ, $y \approx x_1 x_3$, 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 Rossler system [Ros76]

$$
\begin{aligned}
\frac{dX_1}{dt} &= -X_2 - X_3 \\
\frac{dX_2}{dt} &= X_1 + \frac{1}{5}X_2 \\
\frac{dX_3}{dt} &= \frac{1}{5} + X_3(X_1 - 5)
\end{aligned}
\tag{4.7}
$$

Figure 4.1 shows the attractor for (x, y), projected onto (x_1, x_2), compared with the attractor for X projected onto (X_1, X_2) at the value $\epsilon = 0.01$. Notice the clear similarities between the two. Figure 4.2 compares the histograms for x_1 and X_1 over 10^5 time units; the two histograms are clearly closely related.

4.3   Discussion and Bibliography

The reduction of differential systems with attracting slow manifold into differential-algebraic systems is due originally to independent work of Levinson and Tikhonov (O'Malley [O'M91], Tikhonov et al. [TVS85]). The use of a spectral gap sufficiently large relative to the size of the nonlinear terms is used in the construction of local stable, unstable and center manifolds (e.g., Wiggins [Wig90]), slow manifolds (Kreiss [Kre92]) and inertial manifolds
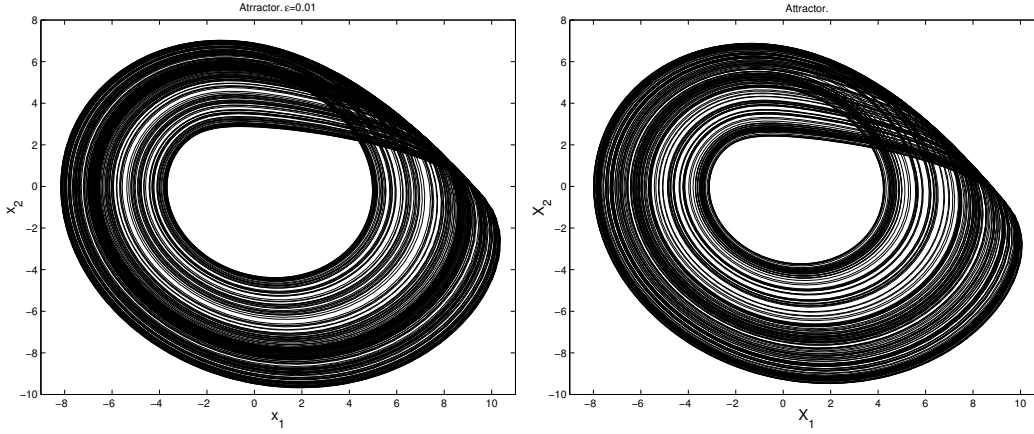
Figure 4.1: Comparison between the attracting sets for (4.6) with $\epsilon = 0.01$ (left) and (4.7) (right), projected on the (x_1, x_2) and (X_1, X_2) planes, respectively.
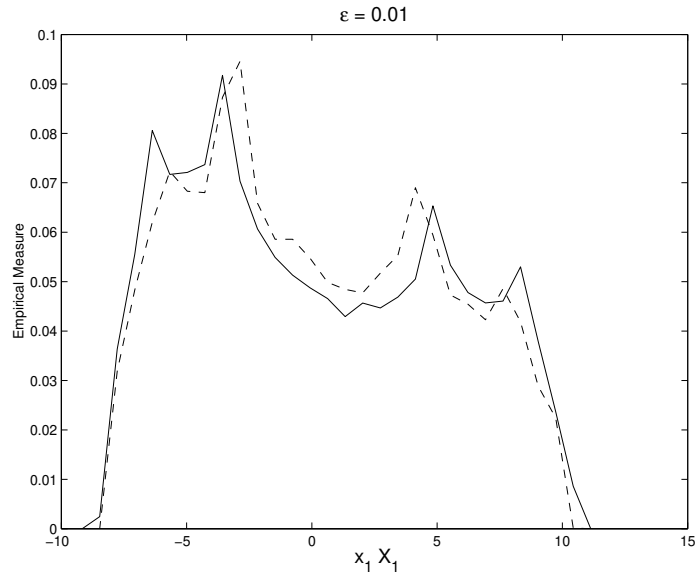

Figure 4.2: Comparison between the empirical measures of x_1 solving (4.6) with $\epsilon = 0.01$ (solid line), and X_1 solving (4.7) (dashed line). The empirical measure was computed from a trajectory over a time interval of 10^5 units.

(Constantin et al. [CFNT94]). A variety of methods of proof exist, predominantly the Lyapunov-Perron approach (Hale [Hal88], Temam [Tem99]) and the Hadamard graph transform (Wells [Wel76]). The numerical calculation of slow dynamics, via dimension reduction in fast-slow systems and with application to problems in chemical kinetics, is described in [DH96].

5 Scale Separation and Averaging

There exists 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. 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.

5.1 The averaging method

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

$$\begin{aligned}
\frac{dx}{dt} &= f(x, y) \\
\frac{dy}{dt} &= \frac{1}{\epsilon} g(x, y),
\end{aligned} \tag{5.1}$$

where $\epsilon \ll 1$.

   The starting point is to analyze 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

$$\frac{d}{dt}\varphi_x^t(y) = g(x, \varphi_x^t(y)), \qquad \varphi_x^0(y) = y \tag{5.2}$$

(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 functions $\psi : \mathcal{Y} \to \mathbb{R}$ the limit of the time average

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

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.

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,

$$\frac{dX}{dt} = F(X) \tag{5.3}$$

$$F(\zeta) = \int_{\mathcal{Y}} f(\zeta, y) \, \mu_\zeta(dy). \tag{5.4}$$

Similar ideas prevail if the invariant measure generated by the y dynamics depends upon the initial data in $\mathcal{Y}$, but are complex to state in general. We outline some of these situations in the remainder of this section.

5.2   Model Problem

An application area where averaging techniques are of current interest is molecular dynamics. Here one encounters Hamiltonian systems with strong potential forces, responsible for fast, small amplitude, oscillations around a constraining sub-manifold. The goal is to describe the evolution of the slowly evolving degrees of freedom by averaging over the rapidly oscillating variables. Bornemann and Schütte [BS97] developed a systematic treatment of such systems that builds upon earlier work by Rubin and Ungar [RU57], using the techniques of time-homogenization [Bor98].

The general setting is a Hamiltonian of the form

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

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 [BS97], shows how problems with this form of Hamiltonian can be cast in the general set-up of equations (5.1). 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: $\omega(x) \geq \bar{\omega} > 0$ for all x. The corresponding equations of motion are

$$\frac{dx}{dt} = p \qquad\qquad \frac{dy}{dt} = v$$
$$\frac{dp}{dt} = -V'(x) - \frac{\omega'(x)}{2\epsilon^2}y^2 \qquad \frac{dv}{dt} = -\frac{\omega(x)}{\epsilon^2}y.$$

The assumption that the energy does not depend on ϵ implies that $y^2 \leq 2\epsilon^2 E/\bar{\omega}$ and the solution approaches the submanifold $y = 0$ 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 variables $\eta = y/\epsilon$. The equations then read

$$\frac{dx}{dt} = p \qquad\qquad \frac{d\eta}{dt} = \frac{1}{\epsilon}v$$
$$\frac{dp}{dt} = -V'(x) - \frac{\omega'(x)}{2}\eta^2 \qquad \frac{dv}{dt} = -\frac{\omega(x)}{\epsilon}\eta.$$

In these variables we recover a system of the form (5.1) with "slow" variables, (x, p), and "fast" variables, (η, 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 [BO99], but it is more 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

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

Harmonic oscillators satisfy an equi-partition 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 $(X, P) \approx (x, p)$,

$$\begin{aligned}
\frac{dX}{dt} &= P \\
\frac{dP}{dt} &= -V'(X) - \frac{\omega'(X)}{2\omega(X)}\left[E - \frac{1}{2}P^2 - V(X) \right],
\end{aligned} \tag{5.5}$$

with initial data E, $X_0 = x_0$ and $P_0 = p_0$. It is easily verified that (X, P) satisfying (5.5) conserve the following invariant,

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

Thus, (5.5) reduces to the simpler form

$$\begin{aligned}
\frac{dX}{dt} &= P \\
\frac{dP}{dt} &= -V'(X) - c\,\omega'(X),
\end{aligned}$$

where

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

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

$$V\mathrm{eff}(x) = V(x) + c\,\omega(x).$$

Note that the limiting equation contains memory of the initial conditions for
the fast variables, through the initial energy E. Thus the situation differs
slightly from the Anosov theorem described previously.

This heuristic derivation is made rigorous in [BS97], using time-homogenization
techniques, and it is also generalized to higher dimension. The condition
that $\omega(x)$ be bounded away from zero may be viewed as a non-resonance
condition. Resonances become increasingly important as the co-dimension,
m, increases, limiting the applicability of the time-homogenization approach
(Takens [Tak80]).

5.3 Discussion and Bibliography

A detailed account on the averaging method, as well as numerous examples
can be found in Sanders and Verhulst [SV85]. Much of the literature on av-
eraging was published in Russian and has remained untranslated; an English
review of this literature is found in Lochak and Meunier [LM88].

Anosov's theorem requires the fast dynamics to be ergodic. Often ergod-
icity fails due to the presence of "resonant zones"—regions in $\mathcal{X}$ for which the
fast dynamics is not ergodic. Arnold and Neistadt [LM88] extended Anosov's
result for situations in which the ergodicity assumption fails on a sufficiently
small set of $x \in \mathcal{X}$.

The situations in which the fast dynamics tend to fixed points, periodic
solutions, or chaotic solutions can be treated in a unified manner through
the introduction of Young measures. Artstein and co-workers considered a
class of singularly perturbed system of type (5.1), with attention given to the
limiting behaviour of both slow and fast variables. In all of the above cases
the pair (x, y) can be shown to converge to (X, μ_X), where X is the solution
of

$$\frac{dX}{dt} = \int_{\mathcal{Y}} f(X, y)\,\mu_X(dy),$$

and μ_X is the ergodic measure on $\mathcal{Y}$; the convergence of y to μ_X is in the
sense of Young measure. (In the case of a fixed point the Young measure is
concentrated at a point.) A general theorem along these lines is proved in
[Art02].

There are many generalization to this idea. The case of non-autonomous fast dynamics, as well as a case with infinite dimension are covered in [AS01]. 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 [AV96].

In the context of stochastic systems, an interesting set-up for averaging is to consider systems of the form

$$
\begin{aligned}
\frac{dx}{dt} &= f(x, y) \\
\frac{dy}{dt} &= \frac{1}{\epsilon} g(x, y) + \frac{1}{\sqrt{\epsilon}} \frac{dV}{dt}.
\end{aligned}
\tag{5.6}
$$

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 $S - D$. This idea is generalized in Subsection 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 Section 4 a separation of time-scales led to deterministic dynamics in $\mathcal{X}$ and the "slaving" of the y variable to the x variable. In the Fokker-Planck picture this corresponds to a situation where, after a short transient, the probability density has the approximate form

$$
\rho(x, y, t) \approx \delta(y - \eta(x)) \bar{\rho}(x, t)
$$

(the measure is concentrated near the approximation to the invariant manifold, namely $y = \eta(x)$).

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, but 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 theory will be developed 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 literature when full back-coupling between the variables is present.

In the skew-product case 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. This ansatz 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 analog of the slaving and averaging techniques of the previous two sections. When full back-coupling is present, the approximate solution will take the form

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

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 re-study the problem using the Fokker-Planck approach. The two approaches are both illustrated by a model problem, 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

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

that is

$$\begin{aligned}
\frac{dx}{dt} &= \frac{1}{\epsilon}f_0(x, y) + f_1(x, y) \\
\frac{dy}{dt} &= \frac{1}{\epsilon^2}g_0(y) + \frac{1}{\epsilon}\beta_0(y)\frac{dV}{dt}.
\end{aligned}$$

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 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. 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 f_1 the effective drift. One way to see this is by using the Chapman-Kolmogorov equation and we now develop this idea. Although the underlying theory in this area is developed by Kurtz [Kur73], we follow the recent presentation given by Majda et al. [MTVE01] in which the mathematical structure is made very clear through perturbation expansions.

Recall that $v(x, y, t)$ satisfying 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,

$$\frac{\partial v}{\partial t} = \frac{1}{\epsilon^2}\mathcal{L}_1 v + \frac{1}{\epsilon}\mathcal{L}_2 v + \mathcal{L}_3 v$$
$$v(x, y, 0) = w(x), \tag{6.3}$$

where

$$\mathcal{L}_1 v = g_0 \cdot \nabla_y v + \frac{1}{2}(\beta_0\beta_0^T) : \nabla_y(\nabla_y v) \tag{6.4}$$

$$\mathcal{L}_2 v = f_0 \cdot \nabla_x v \tag{6.5}$$

$$\mathcal{L}_3 v = f_1 \cdot \nabla_x v. \tag{6.6}$$

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 (6.3) and equating powers of ϵ gives a hierarchy of equations, the first three of which are

$$\mathcal{L}_1 v_0 = 0 \tag{6.7}$$

$$\mathcal{L}_1 v_1 = -\mathcal{L}_2 v_0 \tag{6.8}$$

$$\mathcal{L}_1 v_2 = \frac{\partial v_0}{\partial t} - \mathcal{L}_2 v_1 - \mathcal{L}_3 v_0. \tag{6.9}$$

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

Note that $\mathcal{L}_1$, given by (6.4), 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)

$$\lim_{t \to \infty} e^{\mathcal{L}_1 t} \phi(y_0) = \lim_{t \to \infty} \int_\mathcal{Y} \phi(y) \left(e^{\mathcal{L}_1^* t} \rho_0 \right)(y) \, dz = \int_\mathcal{Y} \phi(y) \rho_\infty(y) \, dy = \langle \phi \rangle, \tag{6.10}$$

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.

We now argue that functions ϕ satisfying $\mathcal{L}_1 \phi = 0$ are independent of y. Indeed,

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

so that $\phi(y) = e^{\mathcal{L}_1 s} \phi(y)$ for all s. Letting $s \to \infty$ and using (6.10) we get

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

and the latter is independent of y. Thus, the first equation (6.7) in the hierarchy implies $v_0 = v_0(x, t)$.

Consider next the v_1 equation (6.8). 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 (6.5):

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

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 (6.9) the solvability condition for v_2 becomes

$$\frac{\partial v_0}{\partial t} = -\left\langle \mathcal{L}_2 \mathcal{L}_1^{-1} \mathcal{L}_2 v_0 \right\rangle + \left\langle \mathcal{L}_3 v_0 \right\rangle. \tag{6.11}$$

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 SDE in $\mathcal{X}$:

$$\frac{dX}{dt} = F(X) + A(X) \frac{dU}{dt},$$

U being standard Brownian motion. That is, (6.11) 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 $\mathcal{L}_1 r(x, y) = f_0(x, y)$. Hence

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

and (6.11) takes the explicit form

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

Thus $A(x)$ satisfies

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

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, $-2\left\langle r f_0^T \right\rangle$, is positive definite. Notice that, for all constant vectors a,

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

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

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

Finally,
$$F(x) = \langle f_1 - f_0 \cdot \nabla_x r \rangle .$$

Note that the explicit extraction of $A(X), F(X)$ may not be possible in general since it requires the inversion of $\mathcal{L}_1$.

6.2   Model Problem

Consider the equations in $\mathcal{X} = \mathcal{Y} = \mathbb{R}$:

$$\frac{dx}{dt} = -\lambda x + \frac{1}{\epsilon}yx,$$
$$\frac{dy}{dt} = -\frac{1}{\epsilon^2}y + \frac{1}{\epsilon}\frac{dV}{dt},$$

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

$$\mathcal{L}_1 v = -y\frac{\partial v}{\partial y} + \frac{1}{2}\frac{\partial^2 v}{\partial y^2}$$
$$\mathcal{L}_2 v = xy\frac{\partial v}{\partial x}$$
$$\mathcal{L}_3 v = -\lambda x\frac{\partial v}{\partial x}.$$

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 (6.11) 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) SDE

$$\frac{dX}{dt} = \left(\frac{1}{2} - \lambda\right)X + X\frac{dU}{dt}. \tag{6.12}$$

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

$$\frac{dY}{dt} = \left(\frac{\partial g}{\partial t} + \frac{1}{2}\frac{\partial^2 g}{\partial U^2}\right) + \frac{\partial g}{\partial U}\frac{dU}{dt}.$$

Using this, it is immediately verified that equation (6.12) has the exact solution

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

In order to test the theory we compare the behaviour of x against known theoretical properties of X solving the limiting SDE (6.12). From the exact solution $X(t)$ and the properties of Brownian motion (see Mao [Mao97], p. 105) it follows that:

$$\begin{cases} \lambda > 0 \Leftrightarrow & \lim_{t\to\infty} X(t) = 0 \text{ a.s.} \\ \lambda = 0 \Leftrightarrow & \limsup_{t\to\infty} |X(t)| = \infty \text{ and } \liminf_{t\to\infty} |X(t)| = 0 \text{ a.s.} \\ \lambda < 0 \Leftrightarrow & \lim_{t\to\infty} |X(t)| = \infty \text{ a.s.} \end{cases}$$
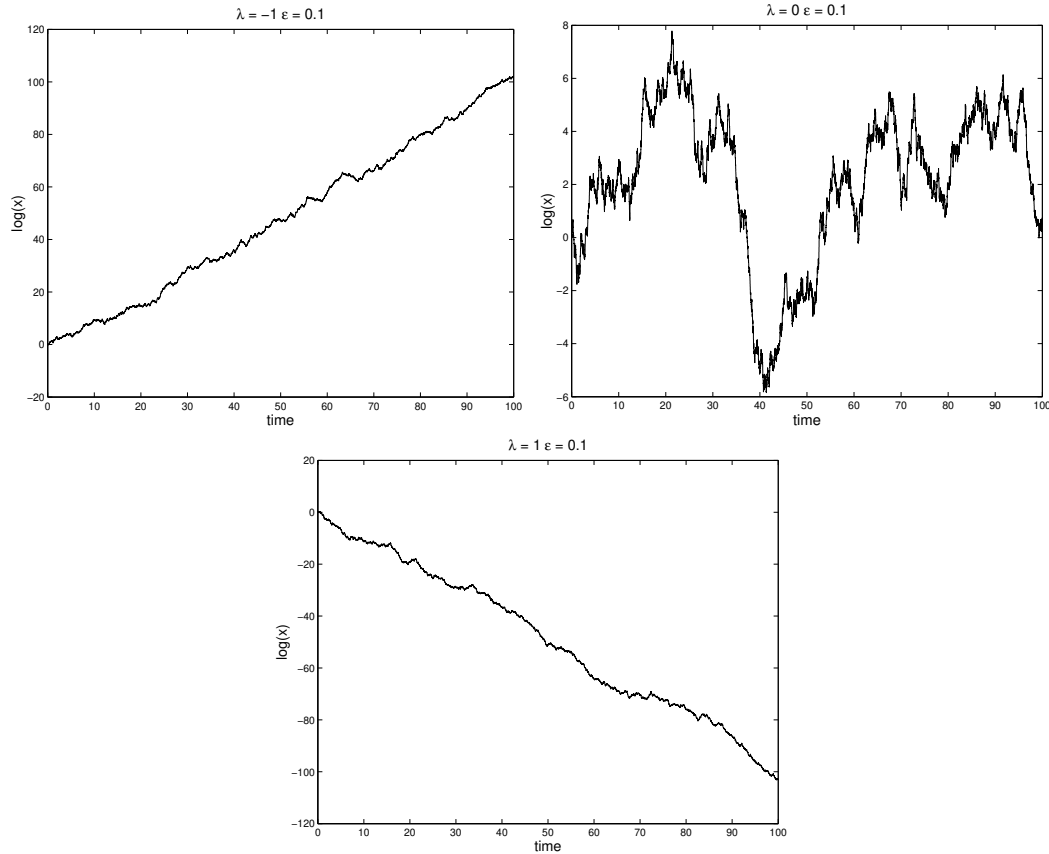
Figure 6.1: Time evolution of $\log x(t)$ for $\epsilon = 0.1$ and (a) $\lambda = -1$, (b) $\lambda = 0$, and (c) $\lambda = 1$.

In Figure 6.1 we show three trajectories of $\log x(t)$ for $\lambda = -1, 0$ and 1 respectively. The value of ϵ is 0.1 . In Figure 6.2 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 $-\infty$).
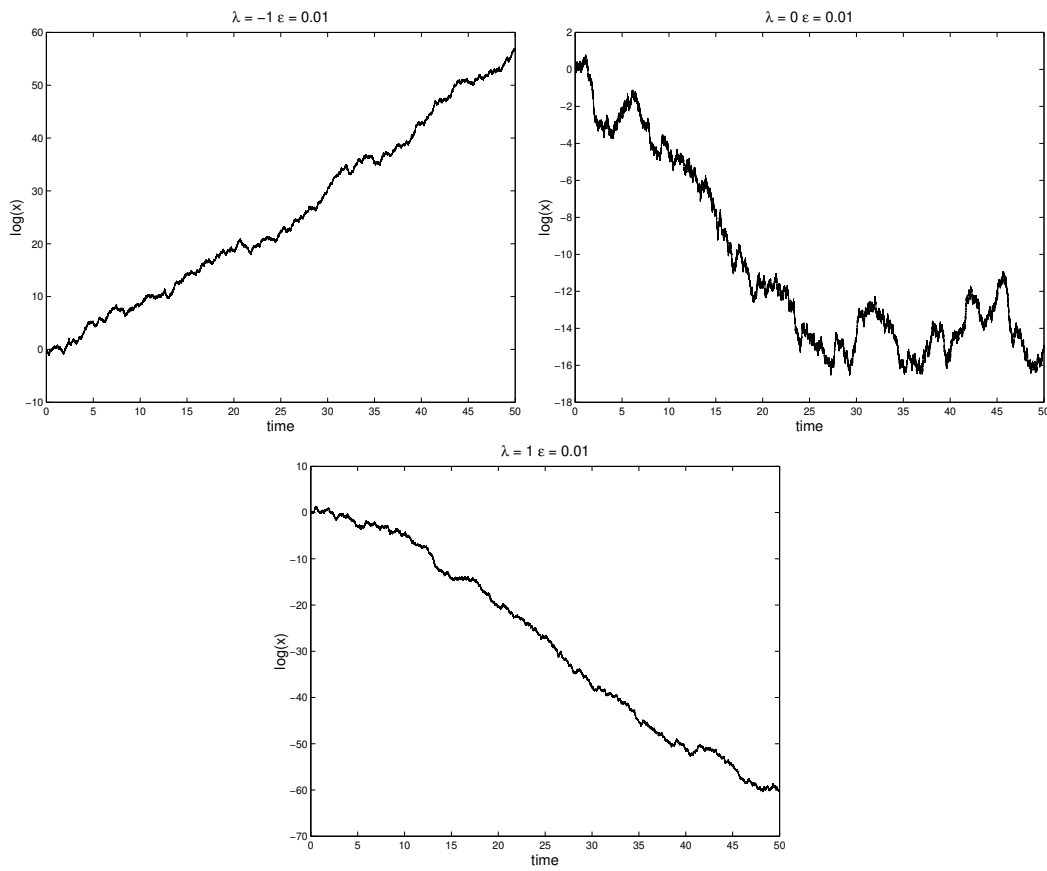
Figure 6.2: Time evolution of $\log x(t)$ for $\epsilon = 0.01$ and (a) $\lambda = -1$, (b) $\lambda = 0$, and (c) $\lambda = 1$.

6.3   The Fokker-Planck Picture

If $\beta_0 \equiv 0$ then the basic ideas outlined in Subsection 6.1 still apply formally, provided the dynamics in z is mixing in a sufficiently strong sense[2]. The approach based on the Chapman-Kolmogorov equations is no longer appropriate in the deterministic setting as it does not allow averaging with respect to initial data; hence we develop a treatment similar to that in Subsection 6.1, but using the Fokker-Planck equation. We will retain the noise term β_0, but allow it to be zero when interpreting the final results. In the Fokker-Planck picture this corresponds to parabolic regularization.

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

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

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 submanifold, 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

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

---

[2]Examples of deterministic dynamics with provably strong mixing properties include geodesic flow on manifolds with negative curvature [KH95]; empirically there are many interesting systems which appear to obey this condition, including the Lorenz equations, for example, and the recent interesting "Burger's bath" of Majda and Timofeyev [MT00].

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 $\bar\rho(x,t)$:

$$\frac{\partial\bar\rho}{\partial t} = -\int_\mathcal{Y} \mathcal{L}_2^*(\mathcal{L}_1^*)^{-1}\mathcal{L}_2^*\rho_\infty\bar\rho\, dy + \int_\mathcal{Y} \mathcal{L}_3^*\rho_\infty\bar\rho\, dy. \tag{6.13}$$

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 $\mathcal{X}$:

$$\frac{dX}{dt} = F(X) + A(X)\frac{dU}{dt},$$

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

6.4 Model Problem

Consider the equations

$$\begin{aligned}
\frac{dx}{dt} &= x - x^3 + \frac{4}{90\epsilon}y_2 \\
\frac{dy_1}{dt} &= \frac{10}{\epsilon^2}(y_2 - y_1) \\
\frac{dy_2}{dt} &= \frac{1}{\epsilon^2}(28y_1 - y_2 - y_1y_3) \\
\frac{dy_3}{dt} &= \frac{1}{\epsilon^2}(y_1y_2 - \frac{8}{3}y_3)
\end{aligned} \tag{6.14}$$

Note that the vector $y = (y_1, y_2, y_3)^T$ solves the Lorenz equations, at parameter values where the solution is chaotic [Wig90]. 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

$$\frac{dX}{dt} = X - X^3 + \sigma\frac{dW}{dt}, \qquad (6.15)$$

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 (6.13) 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

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

where

$$\sigma^2 = \frac{8}{90}\int_{\mathcal{Y}} y_2 r(y)\,dy.$$

Also

$$\int_{\mathcal{Y}}\mathcal{L}_3^*\rho_\infty\bar{\rho}\,dy = -\frac{\partial}{\partial x}\left[(x - x^3)\bar{\rho}\right].$$

Thus the limiting equation for probability densities is

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

which is the Fokker-Planck equation for the SDE (6.15).

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 time step Δt. We study the time-series γ_n defined by

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

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 (6.15) 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^3.

Figure 6.3 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 6.4 (which is over the long time 10^4), generated from $x(t)$, with the exact invariant measure for the SDE (6.15), at the estimated value of σ. The agreement is very good.

6.5   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 [Kur73]. Applications to climate models, where the atmosphere evolves slowly relative to the fast oceanic variations, are surveyed in Majda et al. [MTVE01]; as mentioned above, it is the presentation in [MTVE01] which we have followed here.

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 [PK74]. The idea outlined here is carried out in discrete time by Beck [Bec90] who also uses a skew-product structure to enable the analysis; the ideas can then be rigorously justified in some cases. Related ideas in continuous time are

---

[3]This method of parameter estimation for the stochastic model is quite general and does not exploit the scale-separation in an optimal fashion; other methods could be used which do exploit scale-separation, such as the method outlined in [VE03].
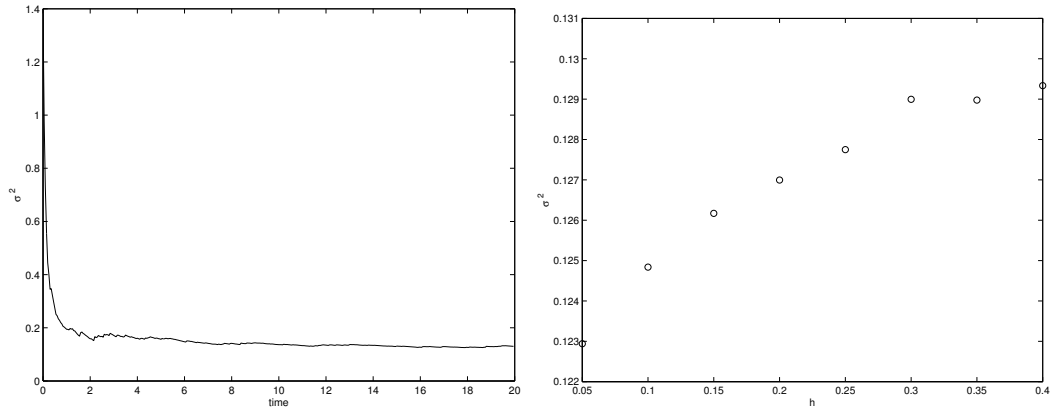
Figure 6.3: Left: estimated value of σ as function of the size of the time interval for $\epsilon^2 = 0.001$ and $h = 0.05$. Right: estimated value of σ as function of the sampling interval h.
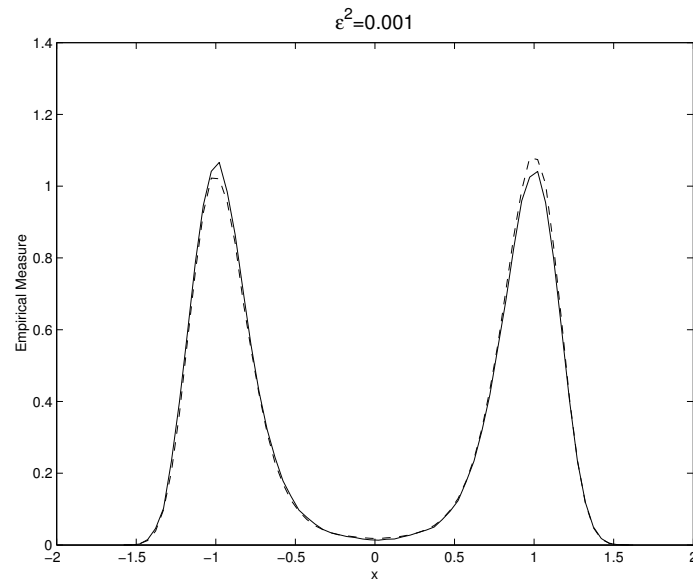

Figure 6.4: Empirical measure of $x(t)$ solving (6.14) for $\epsilon^2 = 0.001$ (solid line) compared with the empirical measure of $X(t)$ solving (6.15).

addressed in [JKRH01, JGB$^+$03] 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) [CHK02], 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 Subsection 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 Subsection 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

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

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

$$\frac{dX}{dt} = \frac{1}{\epsilon}F_0(X) + F_1(X) + A(X)\frac{dU}{dt},$$

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, is covered in a series of papers by Papanicolaou and co-workers—see [PV73, Pap74, PK74, Pap76], building on original work of Khasminkii [Kha63, Kha66]. See also [JKRH01, JGB$^+$03, Bec90, MTVE01].

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 $\alpha = 0$

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

The equation for y is thus

$$\frac{dy}{dt} = \frac{1}{\epsilon^2}g_0(x,y) + \frac{1}{\epsilon}\beta_0(x,y)\frac{dV}{dt}.$$

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

$$\frac{dY}{dt} = g_0(\zeta,y) + \beta_0(\zeta,y)\frac{dV}{dt}, \tag{6.16}$$

where ζ is viewed as a parameter. If this equation is ergodic, for each fixed ζ, with invariant measure π_ζ, then it is natural to try and generalize the studies of the previous sections, replacing $\langle \cdot \rangle$ by by averaging with respect to π_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 $\rho_\infty(\zeta,y)$ is the invariant density for (6.16). Such ideas can be developed systematically; see [Kha63, Kha66, MTVE01, PV73, PK74, Pap76, JKRH01, JGB$^+$03] 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 last section we showed how effective low-dimensional stochastic models can arise from either higher dimensional SDEs or ODEs, when a separation of time-scales occurs. We worked with the Chapman-Kolmogorov or Fokker-Planck equations, rather than 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. Here we will study the system from a pathwise perspective, rather than using the Chapman-Kolmogorov or Fokker-Planck picture. We are studying problems of the form D-S in the classification of Section 1.

7.1 Trigonometric Approximation of Gaussian Processes

Recall that 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 [KT75]) 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 trigonometric series for Gaussian processes (see Kahane [Kah85]). We consider the approximation of Gaussian processes by finite series of the form

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

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 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

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

where

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

The idea is 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 $Z(t)$. This basic idea, as well as its applications, are made more precise in the following subsections.

7.2 Skew-Product Systems

Consider now the system of ODEs:

$$\frac{dx}{dt} = f(x) + \sum_{j=1}^N k_j q_j$$

$$m_j \frac{d^2}{dt^2} q_j + q_j = 0, \qquad j = 1, 2, \ldots, N,$$

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 $y = (q, \frac{dq}{dt})$, and $z = (x, y) = (x, q, \frac{dq}{dt})$, 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 $p_j = m_j \, (dq/dt)$. 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 the density

$$\frac{1}{Z} e^{-\beta H(p,q)},$$

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. $\mathcal{N}(0, 1)$.

To establish a connection with the previous subsection we choose the spring 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), and thus the "essential dynamics", $x(t)$, satisfy the randomly-driven ODE:

$$\frac{dx}{dt} = f(x) + \beta^{-1/2} \, Z_N(t). \tag{7.3}$$

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 $F^2(\omega)$:

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

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 $R(t)$ (see [KSTT02] for details; see [Bil68] 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 Ornstein-Uhlenbeck (OU) process defined by an Itô SDE. Finally, it can be shown that (7.3) defines a continuous mapping, $Z_N \mapsto x$, between $C([0,T], \mathbb{R})$ functions. Since weak convergence is preserved under continuous mappings it follows that $x(t)$ is approximated, for N large, by $X(t)$ solving the SDE

$$\frac{dX}{dt} = f(X) + \beta^{-1/2} Z(t), \tag{7.4}$$

$$\frac{dZ}{dt} = -\alpha Z + (2\alpha)^{1/2} \frac{dB}{dt}, \tag{7.5}$$

where $B(t)$ is standard Brownian motion and $Z(t)$ is an OU process.

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

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

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 to a function in $C([-\pi/2, \pi/2], \mathbb{R})$ which may be identified with Brownian motion ([Kry95]). Thus, for large N, x is approximated by the SDE:

$$\frac{dX}{dt} = f(X) + \beta^{-1/2}\frac{dU}{dt}.$$

The mapping $Y_N \to x$ is continuous and hence Here U is standard Brownian motion. x converges strongly to X as $N \to \infty$ and error estimates can be found [CSSW01]. However the convergence is only on a finite time-interval, 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.

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 Ford, Kac and Mazur

[FK87, FKM65] and by Zwanzig and co-workers [Zwa73, Zwa80]. The results reported here can be found, in full detail, in [SW99, KSTT02].

The mechanical system is defined by the following Hamiltonian,

$$H(P_N, Q_N, p, q) = \frac{1}{2}P_N^2 + V(Q_N) + \sum_{j=1}^{N} \frac{p_j^2}{2m_j} + \sum_{j=1}^{N} \frac{k_j}{2}(q_j - Q_N)^2, \quad (7.6)$$

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 j-th heat bath particle has mass m_j and interacts with the distinguished particle via a linear spring with stiffness constant k_j; the j-th 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

$$\ddot{Q}_N + V'(Q_N) = \sum_{j=1}^{N} k_j(q_j - Q_N)$$
$$\ddot{q}_j + \omega_j^2(q_j - Q_N) = 0, \quad (7.7)$$

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 system is set up so that 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 $\exp(-\beta H)$. It is easily verified that this amounts to choosing

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

where the sequence ξ_j, η_j are defined as above.

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:

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

where

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

and

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

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 different limiting behaviours can be obtained as $N \to \infty$.

EXAMPLE 7.3 By choosing $\omega_j = N^a \zeta_j$, with i.i.d ζ_j and $\zeta_1 \sim \mathcal{U}(0, 1]$, $0 < a < 1$, and

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

the functions R_N, Z_N coincide with those in Example 7.1. Thus R_N converges to $R(t) = e^{-\alpha|t|}$, and Z_N weakly converges on any bounded interval to the OU process $Z(t)$ in (7.5). It can further be shown, using a continuity argument, that Q_N weakly converges to the stochastic process $Q(t)$ solving the stochastic IDE:

$$\ddot{Q} + V'(Q) + \int_0^t R(t - s)\dot{Q}(s) \, ds = \beta^{-1/2} Z(t). \tag{7.9}$$

Moreover, Q solving (7.9) is equivalent to Q solving the SDE

$$\begin{aligned}
\frac{dQ}{dt} &= P \\
\frac{dP}{dt} &= Z - V'(Q) \\
\frac{dZ}{dt} &= (-\alpha Z - P) + (2\alpha/\beta)^{1/2} \frac{dB}{dt}
\end{aligned} \tag{7.10}$$

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, $Z(t)$, which embodies, for large N, the memory effects.
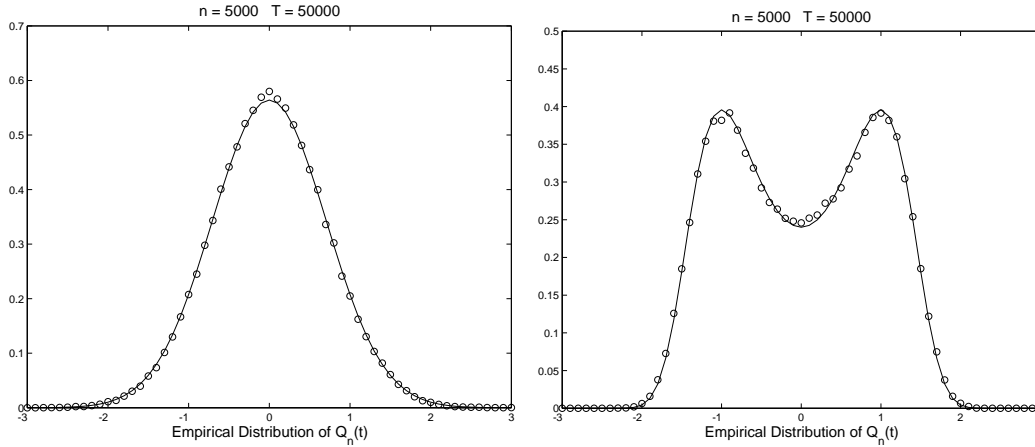
Figure 7.1: Circles: empirical distribution of $Q_N(t)$ for a sample path of the Hamiltonian system in Example 7.3, with parameters $n = 5000$, $\alpha = 1$, $\beta = 2$, and a sampling time of $T = 50000$. The solid line corresponds to the Boltzmann distribution. The graph on the left is for a single-well potential, $V(Q) = Q^2/2$; the graph on the right is for a double-well potential, $V(Q) = Q^4/4 - Q^2/2$.

In Figure 7.1 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 (7.10).

EXAMPLE 7.4 By a slight modification of the arguments in [SW99], a limit to a memoryless Langevin equation can be obtained through 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 Example 7.2. As mentioned in that context, the use of Fourier series means that long time behaviour cannot be studied without seeing periodicity of K_N and Z_N. One way to circumvent this is to re-randomize the data in $\{p, q\}$ periodically in time which is done in [HSS03].

7.4   Discussion and Bibliography

Ford, Kac and Mazur [FKM65] were the first to study mechanical models
of masses and springs as a model of a particle interacting with a heat bath;
see also Ford and Kac [FK87] and Zwanzig [Zwa73, Zwa80]. There exists an
substantial amount of literature on the subject in both classical and quan-
tum contexts (see [FK87, JP97, BM96, LS81]). The more recent work on
heat bath models focuses on those aspects related to dimensional reduction
and coarse grained time stepping. In [SW99] the spectrum of the heat bath
was chosen such that the frequencies are rationally related and the trajec-
tories 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
[KSTT02]. While the main results concern the weak convergence of $Q_N(t)$
on bounded time intervals, certain convergence results, and notable ergodic
properties were also proven for infinite time intervals. A generalization of
those results for the case of heat bath interactions via nonlinear springs may
be found in [KS03]. This work 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 [SW99] and in [HK02a].

8   Chemical Kinetics

Here we present a class of examples 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 proportional to algebraic expressions related to species
concentrations [Gar85]. In this section we start with a very simple example
of this kind of model showing 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 example where numerical experiments show that a similar
closed system of ODEs governs the dynamics.

8.1 Introductory Example

Consider 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 thus

$$\frac{dp_{ij}}{dt} = k_1 a \ p_{ij-1} + k_2(j+1) \ p_{ij+1} - (k_1 a + k_2 j)p_{ij}, \quad p_{ij}(0) = \delta_{ij}, \quad (8.1)$$

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

$$\frac{dp_{i0}}{dt} = k_2 \ p_{i1} - k_1 a p_{i0}, \quad p_{i0}(0) = \delta_{i0} \quad (8.2)$$

for $j = 0$; see [Gar85].

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 through defining an algorithm. Sample paths $z(t)$, $t \geq 0$, with the master equation (8.1) can be generated as follows. Assume that we are given $z(t_l)$, the number of particles at time $t_l \geq 0$. 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

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

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.

Figure 8.1 shows three sample paths of this model, calculated with initial data $z(0) = N$, with $k_1 = 2$, $k_2 = 1$, and $a = N$; 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.

Consider (8.1) with $k_1 = 2$, $k_2 = 1$, and $a = N$. Setting $\rho_j = p_{Nj}$ so that

$$\frac{d\rho_j}{dt} = 2N \rho_{j-1} + (j+1) \rho_{j+1} - (2N + j) \rho_j, \quad \rho_j(0) = \delta_{Nj}.$$
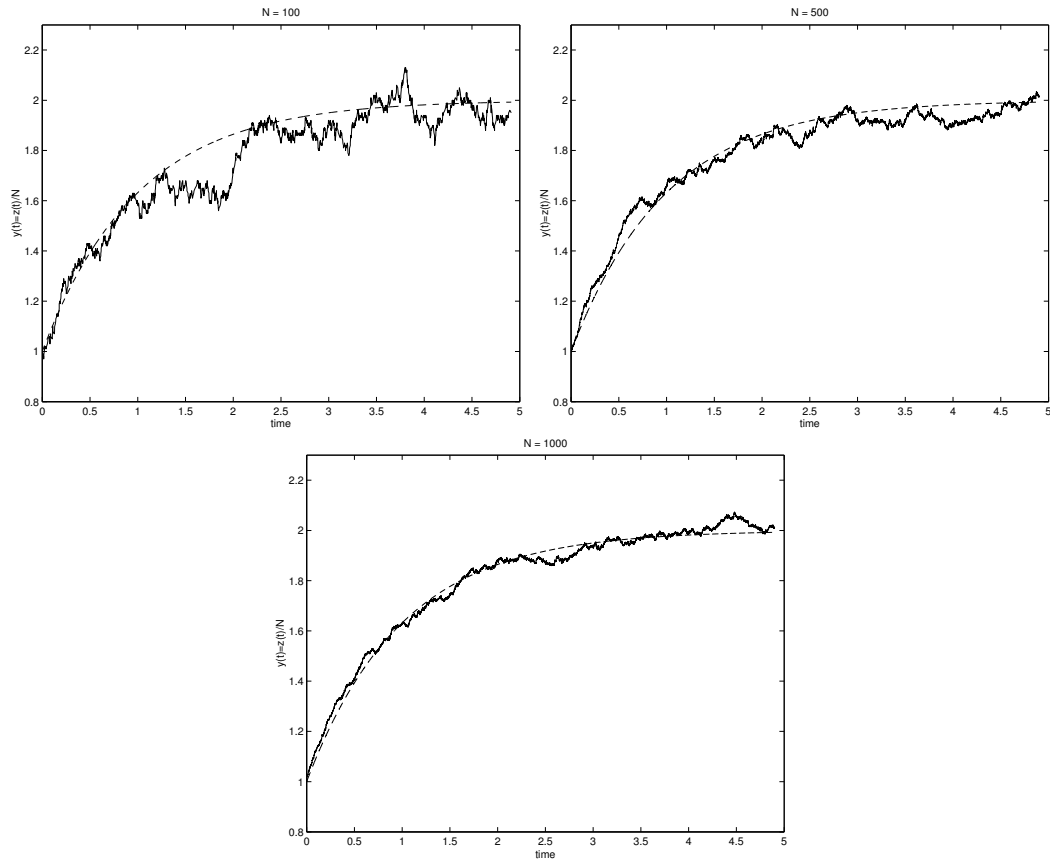
Figure 8.1: Solid lines: sample paths of $z(t)/N$ for the birth death process (8.3) with $z(0) = N$, $k_1 = 2$, $k_2 = 1$, and $a = N$ for (a) $N = 100$, (b) $N = 500$, and (c) $N = 1000$. Dashed lines: the curve $2 - \exp(-t)$.

This may be re-written as, with $\Delta x = N^{-1}$,

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

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

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

(such approximation is known as a Kramers-Moyal expansion [Gar85]). But this is simply the Fokker-Planck equation for the ODE

$$\frac{dX}{dt} = 2 - X, \quad X(0) = 1. \qquad (8.6)$$

This may be viewed as a closed equation for 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.

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 8.1 shows.

8.2 More Involved Example

A generalization of the previous example to more general chemical reactions is as follows (see Gilespie [Gil76, Gil77]). 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

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

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

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

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 subsection 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

$$\frac{dX_i}{dt} = H_i(X), \qquad i = 1, \ldots, m.$$

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

$$
\begin{aligned}
\frac{dX_1}{dt} &= -X_1^2 + X_3 \\
\frac{dX_2}{dt} &= -X_1 X_2 + X_1^2 \\
\frac{dX_3}{dt} &= -X_3 + X_1 X_2
\end{aligned}
\tag{8.7}
$$

Figure 8.2 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, 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 for large N by ODEs, has been exploited in the physics and chemistry communities for some time [Gar85]. Theorems making the formal asymptotic expansions given here rigorous may be found in [Kur76, Kur78]; an overview of these theorems, from an applied perspective, may be found in the paper [FK91]. 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 and stochastic PDEs when fluctuations remain in the limit; see [HV02, Vve03], and the references therein, for example. The derivation of reaction-diffusion equations, and stochastic reaction-diffusion equations, from a birth-death process combined with random walks, is an area of active interest in the physics community; see the lecture notes [Car].

9 Algorithms

There has been an explosion of interest in developing algorithms to compute the evolution of macroscopic quantities from detailed microscopic systems, at minimum cost. Here we briefly highlight some of these approaches, giving pointers to the literature for details.

Recall from the discussion of *Reduction Principles* in the Introduction that any algorithm for dimension reduction comprises two components: determination of the space $\mathcal{X}$ and then determination of the dynamics in $\mathcal{X}$.
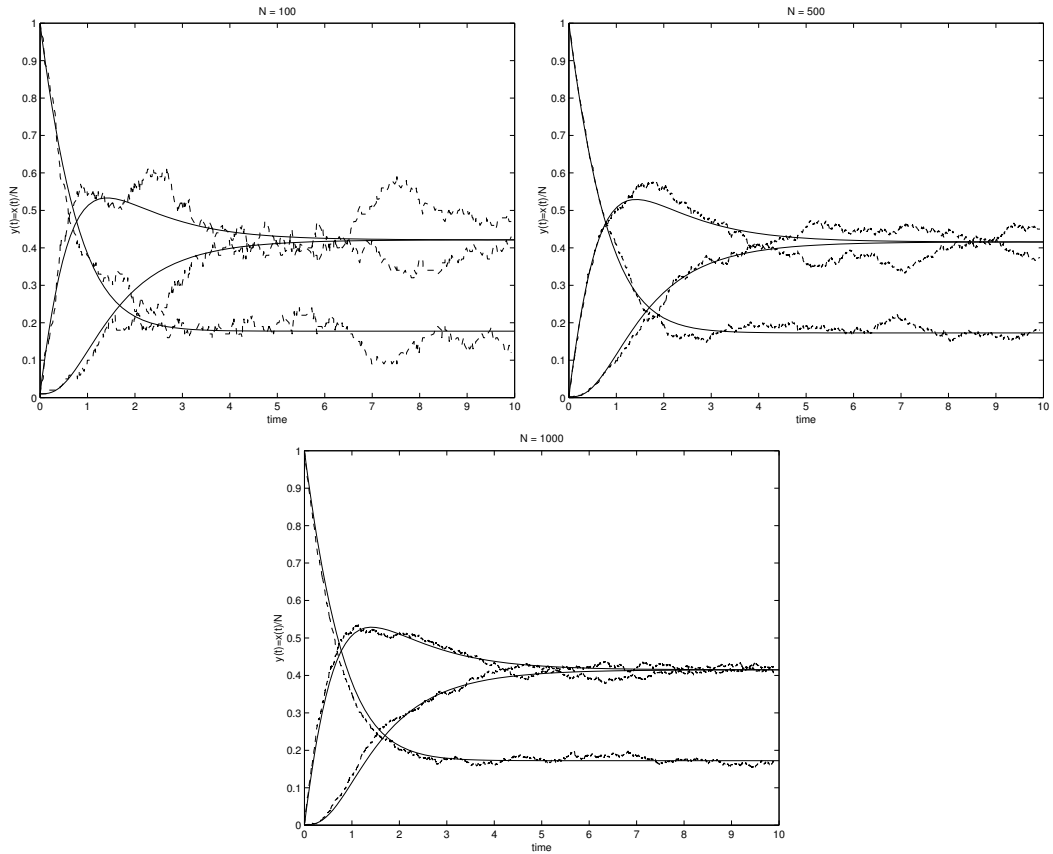
Figure 8.2: Solid lines: sample paths of $x(t)/N$, where $x(t)$ is generated by the birth death process with $\tilde{h}(x)$, $\nu_{\ell j}$, and $x(0)$ as described in the text, for (a) $N = 100$, (b) $N = 500$, and (c) $N = 1000$. Dashed lines: trajectories of $X_3(t)$ for $X(t)$ solving (8.7) with initial data $(1/N, 1/N, 1)$.

The model problems in the previous sections have not addressed the first issue at all: the space $\mathcal{X}$ is explicit from the form of the model problems in all cases. In what follows, the first two subsections, on POD-Based Techniques and Model Reduction, together with the last subsection on Transfer Operator Approaches, all describe algorithms which attempt not only to perform dimension reduction, but also to identify the space $\mathcal{X}$; the methods based on coarse-grained integration also attempt to identify appropriate coarse variables, possibly by the inclusion of further moments.

The subject is very much in its infancy and clear evaluation of the benefits and drawbacks of the different algorithms has not been carried out in the literature. A starting point for this would be to carefully compare the behaviour of algorithms when applied to the model problems described in this paper, and to more challenging problems with similar character drawn from the sciences and engineering. It is important to realize that the discussion in this section is neither comprehensive, nor do we claim to make any serious evaluation of the relative merits of the algorithms described.

9.1 POD-Based Techniques

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 POD (proper orthogonal decomposition or singular value decomposition) to extract dominant modes, and then project the equation (1.1) onto them.
Let

$$z = [z(t_1), z(t_2), \ldots, z(t_N)]$$

be a matrix formed from a large number of samples of a single path of (1.1). The SVD is used to approximate z as

$$z = U\Sigma V^* \approx U_k \Sigma_k V_k^*$$

where U_k^* projects from $\mathcal{Z}$ into a low dimensional subspace of dimension k. Σ_k contains the k leading singular values of Σ. Then the dynamics in (1.1) is approximated by

$$\dot{\xi}(t) = U_k^* f\left(U_k \xi(t)\right)$$

For applications of this approach see [BHL96], [SS89]. The usefulness of this method is sometimes 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 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*.

9.2   Model Reduction

Here the usual application domain is control theory, and most work to date concerns linear systems. With this in mind we set

$$h(z) = Az, \quad \gamma(z) = B, \quad \frac{dW}{dt}(t) = u(t)$$

in equation (1.1) and obtain

$$\frac{dz}{dt} = Az + Bu.$$

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

$$\eta = Cz.$$

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

$$\frac{dx}{dt} = \widehat{A}x + \widehat{B}u$$
$$\eta' = \widehat{C}x$$

provides a good approximation η' to η, for a range of controllers u. In the (x, y) picture of equations (2), this corresponds to finding co-ordinates 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 $\widehat{A}$, $\widehat{B}$, $\widehat{C}$ so that $\eta'(s)$ is a good approximation to $\eta(s)$, where

$$\eta(s) = C(sI - A)^{-1}B$$
$$\eta'(s) = \widehat{C}(sI - \widehat{A})^{-1}\widehat{B}.$$

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 [BG02, Fre99, ASG01].

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

9.3   Optimal Prediction

In this work [CKK98, CKL00, CHK00, CHK02] 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 α, $\beta \equiv 0$ too.

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

$$X(t) = \mathbb{E}[x(t)|x(0) = x_0], \tag{9.1}$$

where $\mathbb{E}$ is with respect to measure ν on $z(0)$. This $X(t)$ 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

$$\frac{dX}{dt} = F(X), \quad X(0) = x_0 \tag{9.2}$$

where

$$F(\zeta) = \frac{\int f(\zeta, y)\nu_\zeta(dy)}{\int \nu_\zeta(dy)}, \tag{9.3}$$

and where ν_ζ is the appropriate conditional measure.

In certain cases, where a separation of scale exists, this approximation coincides with the method of averaging. It cannot work well in general, and errors between the solution of (9.2) and the desired x solving (9.1) can grow like t [Hal99]. However, for certain model problems, this method gives an excellent prediction. One such set of problems are the particle-in-a-heat-bath models of Subsection 7.3 [Kas00].

In order to understand the limitations of this approach, consider the discussion of the Master Equation for countable state space Markov chains in Section 2. It is clear that, in general, there will be no closed equation for the first moment $X(t)$. The approach of Chorin et al. in [BCC00] to

overcome this is to put fluctuations back into the model (9.2) to understand how typical paths x might behave. This is done on an ad hoc basis by *fitting* a diffusion co-efficient α in the model

$$\frac{dX}{dt} = F(X) - \gamma X + \sqrt{2\gamma T}\, \frac{dU}{dt},$$

where U is standard Brownian motion.

9.4   Coarse-Grained Integration

The aim of this approach is, as for optimal prediction, to produce algorithms for X defined by (9.1). Again we assume α, β, $\gamma \equiv 0$ and random initial data. However the methods generalize to random driving, as for optimal prediction.

   The motivation for the details of the approach is that, if the time t map induced by the flow of the vector field F 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. It is possible, in certain circumstances, that stochastic equations for z may have a choice of coarse-grained variables x, for which X is effectively deterministic, as in Section 8, and the algorithm has been used in this context.

   Let $\rho(x, y, t; X)$ denote the probability density function for (9.2), started from measure ν, conditional on $x(0) = X$. In its simplest form the method introduced in [MMK02] is to consider the map

$$X_{n+1} = \Gamma(X_n) \tag{9.4}$$

where

$$\Gamma(X) = \int \rho(x, y, t; X)\; x\; dxdy. \tag{9.5}$$

This may be viewed as an attempt to approximate the time t map for $X(t)$ given by (9.1), and it suffers from many of the same limitations as (9.2), (9.3); however, through the time-scale parameter t, it has an additional flexibility which may be useful. Infinitesimally the method reduces to (9.2), (9.3). To see this 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

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

and so, for $\rho = \rho(x, y, 0; X)$,

$$\Gamma(X) \approx \int \rho x \, dxdy - t \left[\int \int [\nabla_x \cdot (f\rho) + \nabla_y \cdot (g\rho)] x \, dxdy \right]$$

$$= X + t \int f\rho \, dxdy.$$

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 integration by parts in x and y on the second and third terms respectively. Thus

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

with $F(X)$ as in the previous subsection. Thus (9.4) is seen to be of the form

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

an approximation of (9.2), (9.3) which recovers the solution of these equations in the limit $t \to 0$.

In situations where the approximation (9.2), (9.3) or (9.4), (9.5) fails, the approach in [MMK02] proposes a rational closure scheme, in contrast to the somewhat ad hoc closure proposed in [BCC00]. The idea in [MMK02] is to propagate a number of moments of X rather than just the mean. Let

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

and 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 (9.4), (9.5) is the map

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

where

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

By increasing k this gives a rational way of improving the approximation underlying equations (9.2), (9.3) because, under certain regularity assumptions, the moments of x do form a basis for the Master Equation.

However, the moments may form an ill-conditioned basis and, in general, a better choice of basis will give rise to improved approximation of the dynamics for $x(t)$. However, for a number of problems similar to those outlined in Section 8, including problems in chemical reactions, but also 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.

The approach outlined here may yield considerable savings when embedded in bifurcation or continuation software. The approach is studied, by applying it to the heat bath examples of Subsection 7.3, in [BKS03]. 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

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

For systems with scale separation, such as (5.6) and those in Subsection 6, Vanden-Eijnden has derived a new numerical method designed to make similar savings, exploiting decorrelation of fast variables to justify the accuracy [VE03], and building on the asymptotics based mode elimination of [MTVE01]. The method described in this section is more general than that of [VE03], but is less efficient when applied to the particular classes of scale-separated problems for which Vanden-Eijnden's method is purpose-built.

9.5   Coarse Time-Stepping

A related idea is to integrate (1.1) by a time-stepping method which does not resolve time-scales in y. In general this approach will fail because of numerical instabilities or resonances between the time-step frequency and fast unresolved-scales (see [AR99] 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 of Example 7.3 [SW99, CSSW01, HK02a]. One approach to coarse time-stepping is introduced in [TQK00], with applications to problems of the type described in Section 8 covered, for example, in [MMK02, MMPK02]. Application of related algorithmic ideas to the dynamics of a biomolecule may be

found in [HK02b]. Another approach is to use different time-steps for the fast and slow contributions to the force [GASSS98].

9.6   Transfer Operator Approach

The methods of optimal prediction and coarse-grained integration, describe algorithms which attempt to approximate the dynamics of various moments of x, given the full evolution equations for $z = (x, y)$. The transfer operator approach is a more sophisticated, and in general more expensive, attempt to represent the dynamics in x by a small number of degrees of freedom. To simplify exposition of the method, we describe an idealization studied in [HSS03]. In practice the algorithm is used in a more complex fashion and details may be found in [DHFS00].

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 $\mathcal{X}$ 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 to j 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 $m-1$, near 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. It is then possible to find a Markov chain on an $m-$dimensional state space which accurately approximates the coarse dynamics of the problem. In [HSS03] various models similar to those described in Subsection 7.2 are used to evaluate the Transfer Operator approach.

9.7   Other Literature

The *algorithmic* derivation of effective coarse-grained equations for both static and dynamic problems, and for both ODEs and PDEs, is a subject of growing importance. It would be impossible to do justice to the whole field, so we simply mention here several representative references. The asymptotics-based mode elimination of Section 6 is developed into a numerical method in

[VE03]. The work of Keverekidis, outlined in Subsection 9.4 is representative of a general philosophy, which is further generalized in [GK, KGH$^+$02]. A related approach is developed in [EE03], 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 [Hug95, EHW00]; see [Bra01] for a recent review on multiscale methods.

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 represent the low wave number modes, and y the remainder and the algorithms attempt to approximate numerically the function η. A useful reference where this is studied in an applied context is [KJT90], and a discussion of the rate of convergence of such algorithms may be found in [DMT93]. A more recent perspective on these methods, and a cheap implementation through post-processing of the standard Galerkin method, is discussed in [GANT02].

References

[AR99]   U. Ascher and S. Reich. The midpoint scheme and variants for Hamiltonian systems: advantages and pitfalls. *SIAM J. Sci. Comp.*, 21:1045–1065, 1999.

[Art02]   Z. Artstein. On singularly perturbed ordinary differential equations with measure-valued limits. *Math. Bohem.*, 127:139–152, 2002.

[AS01]   Z. Artstein and M. Slemrod. On singularly perturbed retarded functional differential equations. *J. Diff. Eq.*, 171:88–109, 2001.

[ASG01]   A.C. Antoulas, D.C. Sorensen, and S. Gugerrin. *A survey of model reduction methods for large scale systems.* Contemporary mathematics. AMS Publications, 2001.

[AV96]   Z. Artstein and A. Vigodner. Singularly perturbed ordinary differential equations with dynamic limits. *Proc. Roy. Soc. Edimburgh*, 126A:541–569, 1996.

[BCC00]   J. Bell, A.J. Chorin, and W. Crutchfield. Stochastic optimal prediction with application to averaged Euler equations. In C.A. Lin, editor, *Proc. 7th Nat. Conf. Comput. Fluid Mech.*, pages 1–13, Pingtung, Taiwan, 2000.

[Bec90]   C. Beck. Brownian motion from deterministic dynamics. *Physica A*, 169:324–336, 1990.

[BG02]   Z. Bai and G. Golub. Computation of large scale quadratic forms and transfer functions using the theory of moments, quadrature and Pade approximation. In A. Bourlioux and M.J. Gander, editors, *Modern methods in scientific computing and applications*, NATO Science Series. Kluwer Academic Pubs., 2002.

[BHL96]   G. Berkooz, P. Holmes, and J. Lumley. *Coherent structures, dynamical systems and symmetry.* Cambridge monograph on mechanics, Cambridge, 1996.

[Bil68]   P. Billingsley. *Convergence of probability measures.* John Wiley & Sons, New York, 1968.

[BKS03]   D. Barkley, I.G. Kevrekidis, and A.M. Stuart. Coarse integration for large scale dynamical systems: A numerical study. in preparation, 2003.

[BM96]   M. Bianucci and R. Mannella. Linear response of Hamiltonian chaotic systems as a function of the number of degrees of freedom. *Phys. Rev. Lett.*, 77:1258–1261, 1996.

[BO99]   C.M. Bender and S.A. Orszag. *Advanced mathematical methods for scientists and engineers.* Springer-Verlag, New York, 1999.

[Bor98]   F. Bornemann. *Homogenization in Time of Singularly Perturbed Mechanical Systems.* Springer-Verlag, New York, 1998.

[Bra01]   A. Brandt. Multiscale scientific computing: review 2001. In T.J. Barth, T.F. Chan, and R. Haimes, editors, *Multiscale and multiresolution methods: theory and applications.* Springer Verlag, Heidelberg, 2001.

[BS97]   F.A. Bornemann and Ch. Schütte. Homogenization of Hamiltonian systems with a strong constraining potential. *Physica D*, 102:57–77, 1997.

[Car]   J. Cardy. Field theory and nonequilibrium statistical physics. Notes, http://www-thphys.physics.ox.ac.uk/users/JohnCardy/home.html.

[CFNT94]   P. Constantin, C. Foias, B. Nicolaenko, and R. Temam. *Integral manifolds and inertial manifolds.* Springer, New York, 1994.

[CHK00]   A.J. Chorin, O.H. Hald, and R. Kupferman. Optimal prediction and the Mori-Zwanzig representation of irreversible processes. *Proc. Nat. Acad. Sci. USA*, 97:2968–2973, 2000.

[CHK02]   A.J. Chorin, O.H. Hald, and R. Kupferman. Optimal prediction with memory. *Physica D*, 166:239–257, 2002.

[CKK98]   A.J. Chorin, A. Kast, and R. Kupferman. Optimal prediction of underresolved dynamics. *Proc. Nat. Acad. Sci. USA*, 95:4094–4098, 1998.

[CKL00]   A.J. Chorin, R. Kupferman, and D. Levy. Optimal prediction for Hamiltonian partial differential equations. *J. Comp. Phys.*, 162:267–297, 2000.

[CSSW01]   B. Cano, A.M. Stuart, E. Süli, and J.O. Warren. Stiff oscillatory systems, delta jumps and white noise. *Found. Comp. Math.*, 1:69–100, 2001.

[DH96]   P. Deuflhard and J. Heroth. Dynamic dimension reduction in ODE models. In F. Keil, W. Mackens, H. Voss, and J. Werther, editors, *Scientific computing in chemical engineering.* Springer, 1996.

[DHFS00]   P. Deuflhard, W. Huisinga, W. Fischer, and C. Schütte. Identification of almost invariant aggregates in reversible nearly uncoupled Markov chains. *Lin. Alg. Appl.*, 315:39–59, 2000.

[DMT93]   C. Devulder, M. Marion, and E.S. Titi. On the rate of convergence of the nonlinear Galerkin methods. *Math. Comp.*, 60:495–514, 1993.

[EE03]   W. E and B. Engquist. The heterogenous multiscale method. preprint, 2003.

[EHW00]   Y.R. Efendiev, T.Y. Hou, and X.H. Wu. Convergence of a non-conformal multiscale finite element method. *SIAM J. Num. Anal.*, 37:888–910, 2000.

[EM90]   D. Evans and G. Morriss. *Statistical Mechanics of Nonequilibrium Liquids.* Academic Press, London, 1990.

[FK87]   G.W. Ford and M. Kac. On the quantum Langevin equation. *J. Stat. Phys.*, 46:803–810, 1987.

[FK91]   R.F. Fox and J. Keizer. Amplification of intrinsic fluctuations by chaotic dynamics in physical systems. *Phys. Rev. A*, 43:1709–1720, 1991.

[FKM65]   G.W. Ford, M. Kac, and P. Mazur. Statistical mechanics of assemblies of coupled oscillators. *J. Math. Phys.*, 6:504–515, 1965.

[Fre99]   R.W. Freund. Reduced-order modelling techniques based on Krylov subspaces and their use in circuit simulation. *Appl. Comp. Control, Signals and Circuits*, 1:435–498, 1999.

[GANT02]   B. Garcia-Archilla, J. Novo, and E.S. Titi. Postprocessing Fourier spectral methods: the case of smooth solutions. *Appl. Numer. Math.*, 43:191–209, 2002.

[Gar85]   C.W. Gardiner. *Handbook of stochastic methods.* Springer-Verlag, New-York, second edition, 1985.

[GASSS98]   B. Garca-Archilla, J.M. Sanz-Serna, and R.D. Skeel. Long-time-step methods for oscillatory differential equations. *SIAM J. Sci. Comp.*, 20:930–963, 1998.

[GHK03]   D. Givon, O.H. Hald, and R. Kupferman. Existence of orthogonal dynamics. submitted to SIAM J. Math. Anal., 2003.

[Gil76]   D.T. Gilespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comp. Phys.*, 22:403–434, 1976.

[Gil77]   D.T. Gilespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81:2340–2361, 1977.

[GK]   C. W. Gear and I.G. Kevrikidis. Projective methods for stiff differential equations: problems with gaps in their eigenvalue spectrum. submitted to SIAM J. Sci. Comp; also in NEC Technical Report NECI-TR 2001-029.

[GS96]   I.I Gikhman and A.V. Skorokhod. *Introduction to the theory of random processes.* Dover, Mineola, NY, 1996.

[Hal88]   J.K Hale. *Asymptotic behavior of dissipative systems*, volume 25 of *Mathematical surveys and monographs.* AMS, Providence, 1988.

[Hal99]   O.H. Hald. Optimal prediction of the Klein-Gordon equation. *Proc. Nat. Acad. Sci. USA*, 96:4774–4779, 1999.

[HK02a]   O.H. Hald and R. Kupferman. Asymptotic and numerical analyses for mechanical models of heat baths. *J. Stat. Phys.*, 106:1121–1184, 2002.

[HK02b]   G. Hummer and I.G. Kevrikidis. Coarse molecular dynamics of a peptide fragment: free energy, kinetics and long time dynamics computations. submitted to J. Chem. Phys., 2002.

[HSS03]   W. Huisinga, C. Schütte, and A.M. Stuart. Extracting macroscopic stochastic dynamics: model problems. *Comm. Pure Appl. Math.*, 56:234–269, 2003.

[Hug95]   T.J.R. Hughes. Multiscale phenomena: Green's functions, the Dirichlet-to-Neumann formulation, subgrid-scale models, bubbles and the origins of stabilized methods. *Computer Methods in Applied Mechanics and Engineering*, 127:387–401, 1995.

[HV02]   C. Haselwandter and D. Vvedensky. Fluctuations in the lattice gas for Burgers' equation. *J. Phys. A: Math. Gen.*, 35:579–584, 2002.

[JGB⁺03]   W. Just, K. Gelfert, N. Baba, A. Riegert, and H. Kantz. Elimination of fast chaotic degrees of freedom: on the accuracy of the Born approximation. *J. Stat. Phys.*, 112:277–292, 2003.

[JKRH01]   W. Just, H. Kantz, C. Rödenbeck, and M. Helm. Stochastic modelling: replacing fast degrees of freedom by noise. *J. Phys. A*, 34:3199–3213, 2001.

[JP97]   V. Jakšić and C.-A. Pillet. Ergodic properties of the non-Markovian Langevin equation. *Lett. Math. Phys.*, 41:49–57, 1997.

[Kah85]   J.P Kahane. *Some random series of functions*. Cambridge University Press, Cambridge, 1985.

[Kas00]   A.P. Kast. Optimal prediction of stiff oscillatory mechanics. *Proc. Nat. Acad. Sci. USA*, 97:6253–6257, 2000.

[KGH⁺02]   I.G. Kevrekidis, C. W. Gear, J. M. Hyman, P. G. Kevrekidis, O. Runborg, and K. Theodoropoulos. Equation-free multiscale computation: enabling microscopic simulators to perform system-level tasks. submitted to Comm. Math. Sci., 2002.

[KH95]   A. Katok and B. Hasselblatt. *Introduction to the modern theory of dynamical systems*. Cambridge University Press, Cambridge, 1995.

[Kha63]   R.Z. Khasminskii. Principle of averaging for parabolic and elliptic differential equations and for markov processes with small diffusion. *Theory. Prob. Applics.*, 8:1–21, 1963.

[Kha66]   R.Z. Khasminskii. A limit theorem for the solutions of differential equations with random right-handed sides. *Theory. Prob. Applics.*, 11:390–406, 1966.

[KJT90]   I.G. Keverekidis, M.S. Jolly, and E. S. Titi. Approximate inertial manifolds for the Kuramoto-Sivashinsky equation: Analysis and computations. *Physica D*, 44:38–60, 1990.

[KMVE03]   P.R. Kramer, A.J. Majda, and E. Vanden-Eijnden. Testing approximate closures for turbulent diffusion on some model flows. *J. Stat. Phys*, 111:565–679, 2003.

[Kre92]   H.-O. Kreiss. *Problems with different time-scales*. Acta Numerica. Cambridge University Press, Cambridge, 1992.

[Kry95]   N.V. Krylov. *Introduction to the theory of diffusion processes*, volume 142 of *AMS Translation of Monographs*. AMS, Providence, RI, 1995.

[KS03]   R. Kupferman and A.M. Stuart. Fitting SDE models to nonlinear Kac-Zwanzig heat bath models. submitted to Physica D, 2003.

[KSTT02]   R. Kupferman, A.M Stuart, J.R. Terry, and P.F. Tupper. Long term behaviour of large mechanical systems with random initial data. *Stoch. Dyn.*, 2:533–562, 2002.

[KT75]   S. Karlin and H.M. Taylor. *A first course in stochastic processes.* Academic Press, New York, 1975.

[Kur73]   T.G. Kurtz. A limit theorem for perturbed operator semigroups with applications to random evolutions. *J. Func. Anal.*, 12:55–67, 1973.

[Kur76]   T.G. Kurtz. Limit theorems and diffusion approximations for density dependent Markov chains. *Math. Prog. Stud.*, 5:67, 1976.

[Kur78]   T.G. Kurtz. Strong approximation theorems for density dependent Markov chains. *Stoch. Proc. Applics.*, 6:223, 1978.

[LM88]   P. Lochak and C. Meunier. *Multiple phase averaging for classical systems.* Springer-Verlag, New York, 1988.

[LS81]   K. Lindenberg and V. Seshadri. Dissipative contributions of internal multiplicative noise. I: mechanical oscillator. *Physica A*, 109:483–499, 1981.

[Mao97]   X. Mao. *Stochastic differential equations and applications.* Horwood, Chichester, 1997.

[MFS74]   H. Mori, H. Fujisaka, and H. Shigematsu. A new expansion of the master equation. *Prog. Theor. Phys.*, 51:109–122, 1974.

[MMK02]   A. Makeev, D. Maroudas, and Y. Kevrekidis. "Coarse" stability and bifurcation analysis using stochastic simulators: kinetic Monte Carlo examples. *J. Chem. Phys.*, 2002.

[MMPK02]   A. G. Makeev, D. Maroudas, A. Z. Panagiotopoulos, and I.G. Kevrekidis. Coarse bifurcation analysis of kinetic Monte Carlo simulations: a lattice gas model with lateral interactions. *J. Chem. Phys.*, 117:8229–8240, 2002.

[Mor65]   H. Mori. Transport, collective motion, and Brownian motion. *Prog. Theor. Phys.*, 33:423–450, 1965.

[MT00]   A. Majda and I. Timofeyev. Remarkable statistical behavior for truncated Burgers-Hopf dynamics. *Proc. Nat. Acad. Sci. USA*, 97:12413–12417, 2000.

[MTVE01]   A.J. Majda, I. Timofeyev, and E. Vanden-Eijnden. A mathematical framework for stochastic climate models. *Comm. Pure Appl. Math.*, LIV:891–947, 2001.

[Nor97]   J. Norris. *Markov chains.* Cambridge University Press, Cambridge, 1997.

[Øks98]   B. Øksendal. *Stochastic differential equations.* Springer-Verlag, Berlin, fifth edition edition, 1998.

[O'M91]   R.E. O'Malley. *Singular perturbation methods for ordinary differential equations.* Springer, New York, 1991.

[Pap74]   G. Papanicolaou. Introduction to the asymptotic analysis of stochastic equations. In R.C. DiPrima, editor, *Modern Modeling of Continuum Phenomena.* AMS, 1974.

[Pap76]   G.C. Papanicolaou. Some probabilistic problems and methods in singular perturbations. *Rocky Mtn. J. Math*, 6:653–673, 1976.

[PK74]   G.C. Papanicolaou and W. Kohler. Asymptotic theory of mixing stochastic ordinary differential equations. *Comm. Pure Appl. Math*, XXVII:641–668, 1974.

[PV73]   G.C. Papanicolaou and S.R.S. Varadhan. A limit theorem with strong mixing in Banach space and two applications to stochastic differential equations. *Comm. Pure Appl. Math.*, XXVI:497–524, 1973.

[Ris84]   H. Risken. *The Fokker-Planck equation.* Springer, New York, 1984.

[Ros76]   O.E. Rossler. An equation for continuous chaos. *Phys. Lett.*, 35a:397–398, 1976.

[RU57]   H. Rubin and P. Ungar. Motion under a strong constraining force. *Comm. Pure Appl. Math.*, 10:65–87, 1957.

[RW00]   L.C.G. Rogers and D. Williams. *Diffusions, Markov processes and martingales. Volume 2: Itô calculus.* Cambridge University Press, Cambridge, second edition, 2000.

[SS89]   L. Sirovich and C.H. Sirovich. Low dimensional description of complicated phenomena, in connection between finite and infinity dimensional flows. *Cont. Math.*, 99:277–305, 1989.

[SV85]   J.A. Sanders and F. Verhulst. *Averaging methods in nonlinear dynamical systems.* Springer-Verlag, New York, 1985.

[SW99]   A.M. Stuart and J.O. Warren. Analysis and experiments for a computational model of a heat bath. *J. Stat. Phys.*, 97:687–723, 1999.

[Tak80]   F. Takens. Motion under the influence of a strong constraining force. In Z. Nitecki and C. Robinson, editors, *Global theory of dynamical systems, Evanston 1979*. Springer, Berlin, 1980.

[Tem99]   R. Teman. *Infinite dimensional dynamical systems*. Springer-Verlag, New York, second edition, 1999.

[TQK00]   K. Theodoropoulos, Y.-H. Qian, and I.G. Kevrekidis. 'coarse" stability and bifurcation analysis using timesteppers: a reaction diffusion example. *Proc. Natl. Acad. Sci.*, 97:9840–9843, 2000.

[TVS85]   A.N. Tikhonov, A.B. Vasiléva, and A.G. Sveshnikov. *Differential equations*. Springer, Berlin, 1985.

[VE03]   E. Vanden-Eijnden. Numerical techniques for multi-scale dynamical systems with stochastic effects. *Comm. Math. Sci.*, 1:377–384, 2003.

[VEG98]   E. Vanden-Eijnden and A. Greco. Stochastic modelling of turbulence and anomalous transport in plasmas. *J. Plasma Phys.*, 59:683–694, 1998.

[Vve03]   D. Vvedensky. Edwards-Wilkinson equation from lattice transition rules. Preprint, Imperial Colledge Physics Dept., 2003.

[Wel76]   J.C. Wells. Invariant manifolds of nonlinear operators. *Pac. J. Math.*, 62:285–293, 1976.

[Wig90]   S. Wiggins. *Introduction to applied nonlinear dynamical systems and chaos*. Springer, New York, 1990.

[Zwa73]   R. Zwanzig. Nonlinear generalized Langevin equations. *J. Stat. Phys.*, 9:215–220, 1973.

[Zwa80]   R. Zwanzig. Problems in nonlinear transport theory. In L. Garrido, editor, *Systems far from equilibrium*, pages 198–225. Springer, New York, 1980.