G.A. Pavliotis
Department of Mathematics
Imperial College London
and
A.M. Stuart
Mathematics Institute
University of Warwick

# MULTISCALE METHODS: AVERAGING AND HOMOGENIZATION

August 2, 2007

Springer
Berlin Heidelberg New York
Hong Kong London
Milan Paris Tokyo

To my parents $Αργυρη$ and $Σουλτανα$ and to my brother $Γιωργο$.
Carry Home. $Γρηγορης$.

For my children Natalie, Sebastian and Isobel.
AMS.

# Contents

# Preface

The aim of these notes is to describe, in a unified fashion, a set of methods for the simplification of a wide variety of problems which all share the common feature of possessing multiple scales.[1] The mathematical methods which we study are often referred to as the methods of **averaging** and of **homogenization**. The methods apply to partial differential equations (PDEs), stochastic differential equations (SDEs), ordinary differential equations (ODEs) and Markov chains. The unifying principle underlying the collection of techniques described here is the approximation of **singularly perturbed** linear equations. The unity of the subject is most clearly visible in the application of perturbation expansions to the approximation of these singular perturbation problems. A significant portion of the notes is devoted to such perturbation expansions. In this context we use the term **Result** to describe the conclusions of a formal perturbation argument. This enables us to derive important approximation results without the burden of rigorous proof which can sometimes obfuscate the main ideas. However, we will also study a variety of tools from analysis and probability, used to place the approximations derived on a rigorous footing. The resulting theorems are proved using a range of methods, tailored to different settings. There is less unity to this part of the subject. As a consequence considerable background is required to absorb the entire rigorous side of the subject, and we devote a significant fraction of the book to this background material.

The first part of the notes is devoted to the **Background**, the second to the **Perturbation Expansions** which provide the unity of the subject matter, and the third to the **Theory** justifying these perturbative techniques. We do not necessarily recommend that the reader covers the material in this order. A natural way to get an overview of the subject is to read through Part II of the book on Perturbation Expan-

[1] In this book we will apply the general methodology to problems with two, widely separated, characteristic scales. The extension to systems with many seperated scales is fairly straightforward and will be discussed in a number of the Discussion and Bibliography sections which conclude each chapter. In all cases, the important assumption will be that of scale separation.

sions, referring back to the Background material as needed. The Theory can then be studied, after the form of the approximations is understood, on a case by case basis.

Part I (Background) contains the elements of the theory of analysis, probability and stochastic processes, as required for the material in these notes, together with basic introductory material on ODEs, Markov chains, SDEs and PDEs. Part II (Perturbation Expansions) illustrates the use of ideas from averaging and homogenization to study ODEs, Markov chains, SDEs and PDEs of elliptic, parabolic and transport type; invariant manifolds are also discussed, and viewed as a special case of averaging. Part III (Theory) contains illustrations of the rigorous methods which may be employed to establish the validity of the perturbation expansions derived in Part II. The chapters in Part III relate to those in Part II in a one-to-one fashion. It is possible to pick particular themes from this book and cover subsets of chapters devoted only to those themes. The reader interested primarily in SDEs should cover Chapters 6, 10, 11, 17 and 18. Markov chains are covered in Chapters 5, 9 and 16. The subject of homogenization for elliptic PDEs is covered in Chapters 12 and 19. Homogenization and averaging for parabolic and transport equations is covered in Chapters 13, 14, 20 and 21.

The subject matter in these set of notes has, for the most part, been known for several decades. However, the particular presentation of the material here is, we believe, particularly suited to the pedagogical goal of communicating the subject area to the wide range of mathematicians, scientists and engineers who are currently engaged in the use of these tools to tackle the enormous range of applications that require them. In particular we have chosen a setting which demonstrates quite clearly the wide applicability of the techniques to PDEs, SDEs, ODEs and Markov chains, as well as highlighting the unity of the approach. Such a wide–ranging setting is not undertaken, we believe, in existing books, or is done so less explicitly than in this text. We have chosen to use the phrasing **Multiscale Methods** in the title of the book because the material presented here forms the backbone of a significant portion of the amorphous field which now goes by that name. However we do recognize that there are vast parts of the field which we do not cover in this book. In particular, scale separation is a fundamental requirement in all of the perturbation techniques presented in this book. Many applications, however, possess a continuum of scales, with no clear separation. Furthermore, many of the problems arising in multiscale analysis are concerned with the interfacing of different mathematical models appropriate at different scales (such as quantum, molecular and continuum); the tools presented in these notes do not directly address problems arising in such applications as our starting point is a single mathematical model, in which scale separation is present.

These notes are meant to be an introduction, aimed primarily towards graduate students. Part I of the book (where we lay the theoretical foundations) and Part III of the book (where we state and prove theorems concerning simplified versions of the models that are studied in Part II) are necessarily terse; without being so it would be impossible to present the wide range of applications of the ideas, and illustrate their unity. Extensions and generalizations of the results presented in these notes, as well as references to the literature, are given in the Discussion and Bibliography section, at the end of each chapter. With the exception of Chapter 1, all chapters

are supplemented with exercises. We hope that the format of the book will make it appropriate for use both as a textbook and for self–study.

**Acknowledgements**

We are especially grateful to Konstantinos Zygalakis who read and commented upon much of the manuscript, typed parts of it, and helped to create some of the figures. Special thanks also are due to Martin Hairer, Valeriy Slastikov, Endre Süli and Roger Tribe, all of whom read portions of the book in great detail, resulting in many constructive suggestions for improvement. We are also grateful to Niklas Branstrom, Mikhail Cherdanstev, Dror Givon, Weinan E, David Epstein, Liam Jones, Raz Kupferman, Brenda Quinn, Florian Theil and David White for many helpful comments which substantially improved these lecture notes. The book grew out of courses on Multiscale Methods and on Homogenization that we taught, at Imperial College London, and at Warwick University, during the last four years; we thank all the students who attended these courses. Some of this material, in turn, grew out of the review article [125] and we are grateful to Dror Givon and Raz Kupferman for their collaboration in this area.

Parts of the book were used as a basis for two short courses on Multiscale Methods that we taught in April 2007 at the Mathematical Sciences Research Institute (MSRI), Berkeley, CA and the Mathematics Research Centre (MRC), Warwick University. We thank all the participants at these courses, together with the MSRI, MRC, London Mathematical Society and the UK Engineering and Physical Sciences Research Council (EPSRC) for administrative and financial support for these short courses.

GP has received financial support from the EPSRC and AMS has received financial support from the EPSRC and from the US Office of Naval Research during the writing of this book. This funded research has helped shape much of the material presented in the book, and the authors are grateful for the financial support.

GA Pavliotis and AM Stuart, August 2007

# Part II

## Perturbation Expansions

# 8

# Invariant Manifolds for ODEs

## 8.1 Introduction

Perhaps the simplest situation where variable reduction occurs in dynamical systems is that of attractive invariant manifolds. These manifolds *slave* one subset of the variables to another. In this chapter we describe a situation where attractive invariant manifolds can be constructed in scale separated systems, by means of perturbation expansions. In Section 8.2 we introduce the system of ODEs that we want to simplify by means of the theory of invariant manifolds, and in Section 8.3 we present the simplified equations. The simplified equations are derived in Section 8.4 and several examples are presented in Section 8.5. In Section 8.6 we describe various extensions of the results presented in this chapter, together with making bibliographical remarks. We also discuss the material in this section in relation to averaging, the subject of Chapters 9 and 10.

## 8.2 Full Equations

We consider a system of ODEs of the form (4.1.1) and write $z$ as $z = (x^T, y^T)^T$, where

$$\frac{dx}{dt} = f(x, y), \tag{8.2.1a}$$

$$\frac{dy}{dt} = \frac{1}{\varepsilon} g(x, y), \tag{8.2.1b}$$

and $\varepsilon \ll 1$. Here $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ in the notation of Chapter 4.

Let $\varphi_x^t(y)$ be the solution operator of the fast dynamics with $x$ viewed as a fixed parameter and $\varepsilon = 1$. To be precise, for any $\xi \in \mathcal{X}$, let

$$\frac{d}{dt}\varphi_\xi^t(y) = g(\xi, \varphi_\xi^t(y)), \qquad \varphi_\xi^0(y) = y. \tag{8.2.2}$$

We assume that

$$\lim_{t\to\infty} \varphi_\xi^t(y) = \eta(\xi) \tag{8.2.3}$$

exists, is independent of $y$ and that the convergence is uniform in $\xi$. Roughly speaking $y(t)$ solving (8.2.1) is given by $y(t) \approx \varphi_{x(0)}^{t/\varepsilon}(y(0))$ for times $t$ which are small compared with 1 (i.e., $t = o(1)$) so that $x(t)$ has not evolved very much. If we then look at short timescales which are nonetheless large compared with $\varepsilon$, so that $y$ is close to its equilibrium point, (for example if $t = \mathcal{O}(\varepsilon^{\frac{1}{2}})$), we deduce that then $y(t) \approx \eta(x(0))$. This is the mechanism by which $y$ becomes slaved to $x$ and we now seek to make the above heuristics more precise.

Notice that the generator $\mathcal{L}$ for (8.2.1) has the form

$$\mathcal{L} = \frac{1}{\varepsilon}\mathcal{L}_0 + \mathcal{L}_1 \tag{8.2.4}$$

where

$$\mathcal{L}_0 = g(x,y)\cdot\nabla_y, \quad \mathcal{L}_1 = f(x,y)\cdot\nabla_x$$

In particular, $\mathcal{L}_0$ is the generator of a process on $\mathcal{Y}$ for each fixed $x$.

Now consider the following PDE for $v(y,t)$ in which $x$ is viewed as a fixed parameter:

$$\frac{\partial v}{\partial t} = \mathcal{L}_0 v, \quad v(y,0) = \phi(y). \tag{8.2.5}$$

Result 4.6 shows that

$$v(y,t) = \phi(\varphi_x^t(y)).$$

Thus, by (8.2.3),

$$v(y,t) \to \phi(\eta(x)), \text{ as } \quad t\to\infty. \tag{8.2.6}$$

This is related to ergodicity, as equation (8.2.6) shows that the function $v(y,t)$ exhibits no dependence on initial data, asymptotically as $t\to\infty$, and approaches a constant in $y$. Compare with the discussion of ergodicity in Chapter 4, and Theorems 4.12 and 4.13 in particular.

Recall the Definition 4.3 of invariant set. If this set is a manifold then we refer to it as an *invariant manifold*. In this chapter we use the scale-separated form of the equations (8.2.1) to construct an approximate invariant manifold. In fact the manifold will have the structure of a *graph*: it will be represented as a function relating the $y$ coordinates to the $x$ coordinates. Invariant manifolds representible as graphs are particulary important in describing the dynamics of ODEs close to equilibria, leading to the concepts of *stable, unstable and center manifolds* .

## 8.3 Simplified Equations

We now state an approximation result that will be derived by formal perturbation arguments in the next section. Define the vector field $F_0(x)$ by

$$F_0(x) = f(x,\eta(x)). \tag{8.3.1}$$

**Result 8.1.** *For $\varepsilon \ll 1$ and time $t$ upto $\mathcal{O}(1)$, $x(t)$ solving* (8.2.1) *is approximated by $X(t)$ solving*

$$\frac{dX}{dt} = F_0(X), \tag{8.3.2}$$

*where $F_0(x)$ is given by* (8.3.1)*.*

Underlying the derivation of this result is an assumption that $y(0)$ is initalized close to $\eta(x(0))$. When this fails then further arguments are required to deal with what is termed an initial or boundary layer – see Section 8.6 for a discussion of this point.

Result 8.1 gives us the leading order approximation in $\varepsilon$. Keeping the next order yields the refined approximation

$$\frac{dX}{dt} = F_0(X) + \varepsilon F_1(X), \tag{8.3.3}$$

where

$$F_1(x) = \nabla_y f(x, \eta(x)) \Big( \nabla_y g(x, \eta(x)) \Big)^{-1} \nabla_x \eta(x) f(x, \eta(x)).$$

This approximation requires that $\nabla_y g(x, \eta(x))$ is invertible.

## 8.4 Derivation

The method used to find these simplified equations is to seek an approximate invariant manifold for the system. Furthermore, we assume that the manifold can be represented as a graph over $x$, namely $y = \Psi(x)$. The set determined by such a graph is invariant (see Definition 4.3) under the dynamics if

$$\frac{dy}{dt} = \nabla \Psi(x(t)) \frac{dx}{dt},$$

whenever $y = \Psi(x)$. This implies that $\Psi$ must solve the nonlinear PDE

$$\frac{1}{\varepsilon} g(x, \Psi(x)) = \nabla \Psi(x) f(x, \Psi(x)).$$

We seek solutions to this equation as a power series

$$\Psi(x) = \Psi_0(x) + \varepsilon \Psi_1(x) + \mathcal{O}(\varepsilon^2).$$

This is our first example of a perturbation expansion.

Substituting and equating coefficients of successive powers of $\varepsilon$ to zero yields the hierarchy

$$\begin{array}{ll} \mathcal{O}(\frac{1}{\varepsilon}) & g(x, \Psi_0(x)) = 0, \\ \mathcal{O}(1) & \nabla_y g(x, \Psi_0(x)) \Psi_1(x) = \nabla \Psi_0(x) f(x, \Psi_0(x)). \end{array}$$

Notice that equations (8.2.2),(8.2.3) together imply that $g(\xi, \eta(\xi)) = 0$ for all $\xi$. Hence the $\mathcal{O}(\frac{1}{\varepsilon})$ equation above may be satisfied by choosing $\Psi_0(x) = \eta(x)$, giving the approximation (8.3.2). Since the rate of convergence in (8.2.3) is assumed to be uniform is it natural to assume that $y = \eta(\xi)$ is a hyperbolic equilibrium point[1] of (8.2.2), so that $\nabla_y g(x, \eta(x))$ is invertible. Setting $\Psi_0(x) = \eta(x)$ in the $\mathcal{O}(1)$ equation, and inverting, yields

$$\Psi_1(x) = \nabla_y g(x, \eta(x))^{-1} \nabla\eta(x) f(x, \eta(x)).$$

Thus

$$\begin{aligned} f(x, \Psi(x)) &= f\left(x, \Psi_0(x) + \varepsilon\Psi_1(x) + \mathcal{O}(\varepsilon^2)\right) \\ &= f(x, \Psi_0(x)) + \varepsilon\nabla_y f(x, \Psi_0(x))\Psi_1(x) + \mathcal{O}(\varepsilon^2) \\ &= f(x, \eta(x)) + \varepsilon\nabla_y f(x, \eta(x))\Psi_1(x) + \mathcal{O}(\varepsilon^2), \end{aligned}$$

and the refined approximation (8.3.3) follows.

## 8.5 Applications

### 8.5.1 Linear Fast Dynamics

A structure arising in many applications is where the frozen $x$ dynamics, given by $\varphi_\xi^t(\cdot)$, is linear. As a simple example consider the equations

$$\begin{aligned} \frac{dx}{dt} &= f(x, y), \\ \frac{dy}{dt} &= -\frac{y}{\varepsilon} + \frac{\tilde{g}(x)}{\varepsilon}. \end{aligned} \tag{8.5.1}$$

Here $d = 2$ and $\mathcal{X} = \mathcal{Y} = \mathbb{R}$, $\mathcal{Z} = \mathbb{R}^2$. It is straightforward to show that

$$\begin{aligned} \varphi_\xi^t(y) &= e^{-t}y + \int_0^t e^{s-t}\tilde{g}(\xi)ds \\ &= e^{-t}y + (1 - e^{-t})\tilde{g}(\xi). \end{aligned}$$

Hence (8.2.3) is satisfied for $\eta(\cdot) = \tilde{g}(\cdot)$

The simplified equation given by Result 8.1 is hence

$$\frac{dX}{dt} = f(X, \tilde{g}(X)).$$

Using the fact that $\nabla_y g(x, y) = -1$ we see that the more refined approximation (8.3.3) is

$$\frac{dX}{dt} = f(X, \tilde{g}(X))\Big(1 - \varepsilon\frac{df}{dy}(X, \tilde{g}(X))\frac{d\tilde{g}}{dx}(X)\Big).$$

[1] A hyperbolic equilibrium point is one where the linearization of the vector field at the equlibrium point contains no spectrum on the imaginary axis.

### 8.5.2 Large Time Dynamics

The statement of the result concerning simplified dynamics concerns the approximation of $x(t)$ on $\mathcal{O}(1)$ time intervals with respect to $\varepsilon$. However in many cases the results extend naturally to the infinite time domain. The following example illustrates this idea.

Consider the equations

$$\frac{dx_1}{dt} = -x_2 - x_3, \tag{8.5.2a}$$
$$\frac{dx_2}{dt} = x_1 + \frac{1}{5}x_2, \tag{8.5.2b}$$
$$\frac{dx_3}{dt} = \frac{1}{5} + y - 5x_3, \tag{8.5.2c}$$
$$\frac{dy}{dt} = -\frac{y}{\varepsilon} + \frac{x_1 x_3}{\varepsilon}, \tag{8.5.2d}$$

so that $\mathcal{X} = \mathbb{R}^3$ and $\mathcal{Y} = \mathbb{R}$. Result 8.1 indicates that $x$ should be well approximated by $X$ solving the Rössler system

$$\frac{dX_1}{dt} = -X_2 - X_3, \tag{8.5.3a}$$
$$\frac{dX_2}{dt} = X_1 + \frac{1}{5}X_2, \tag{8.5.3b}$$
$$\frac{dX_3}{dt} = \frac{1}{5} + X_3(X_1 - 5). \tag{8.5.3c}$$

The Rössler equations are chaotic and consequently comparison of trajectories over long time-intervals is not natural. A more useful object is the attractor. A comparison of the numerically generated attractors for the two systems is shown in Figure 8.1. The first figure shows the attractor for equations (8.5.2), projected into the $x$ coordinates, for $\varepsilon = 10^{-2}$. The second shows the attractor for the Rössler equations themselves. The agreement is very strong indicating that the simplified dynamics do indeed capture behaviour over long time-intervals.

### 8.5.3 Center Manifold

The center manifold is an invariant manifold containing an equilibrium point whose linearization has neutral directions (subspaces corresponding to eigenvalues with zero real part). Consider the equations

$$\frac{dx}{dt} = \lambda x + \sum_{i=0}^{2} a_i x^i y^{2-i},$$
$$\frac{dy}{dt} = x - y + \sum_{i=0}^{2} b_i x^i y^{2-i}.$$

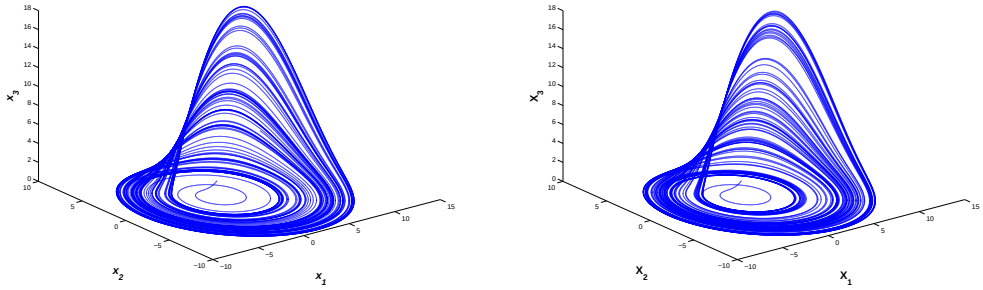

**Fig. 8.1.** Comparison between the attracting sets for (8.5.2) with $\varepsilon = 0.01$ (left) and (8.5.3) (right), projected on the $(x_1, x_2)$ and $(X_1, X_2)$ planes, respectively.

Here $\lambda \in \mathbb{R}$ and the $a_i$ and $b_i$ are also real numbers. Furthermore, for each $t$, $x(t) \in \mathbb{R}$ and $y(t) \in \mathbb{R}$. When linearized at the origin this equation becomes

$$\frac{dx}{dt} = \lambda x,$$
$$\frac{dy}{dt} = x - y.$$

If $(z = (x, y)^T$ then

$$\frac{dz}{dt} = Lz$$

with

$$L = \begin{pmatrix} \lambda & 0 \\ 1 & -1 \end{pmatrix}.$$

The eigenvalues of $L$ are $\lambda$ and $-1$. As $\lambda$ passes through $0$ the linear stability property of the origin thus changes from stable to unstable. For this reason, studying the equation in the vicinity of $\lambda = 0$ is of interest. In particular we expect to find a center manifold at $\lambda = 0$: an invariant manifold tangent to the eigenspace corresponding to eigenvalue $0$ of $L$.

To construct this manifold rescale the equations as follows: we set

$$x \to \varepsilon x, y \to \varepsilon y, \lambda \to \varepsilon\lambda, t \to \varepsilon^{-1}t.$$

This corresponds to looking for small amplitude solutions, close to the fixed point at the origin, at parameter values close to the bifurcation values. Such solutions evolve slowly and hence time is rescaled to capture non-trivial dynamics. The equations become

$$\frac{dx}{dt} = \lambda x + \sum_{i=0}^{2} a_i x^i y^{2-i},$$
$$\frac{dy}{dt} = \frac{1}{\varepsilon}(x-y) + \sum_{i=0}^{2} b_i x^i y^{2-i}.$$

A perturbation expansion gives the invariant manifold $y = x$ and we obtain the following equations for the dynamics on the invariant manifold:

$$\frac{dX}{dt} = \lambda X + AX^2,$$

with $A = \sum_{i=0}^{2} a_i$. The case $\lambda = 0$ gives the center manifold itself, and $\lambda < 0$ the stable manifold.

## 8.6 Discussion and Bibliography

The topic of invariant manifolds has a long history and is itself the subject of entire books. To do it justice here is impossible and we provide only brief pointers to the literature. From the perspective of this book, our primary motivation for covering the topic is that it provides a special case of the method of averaging introduced in the next two chapters; furthermore this case can be introduced without appeal to any arguments from ergodic theory or from the theory of stochastic processes. It hence provides a suitable inroad into the topics of this book for readers with a background in dynamical systems; conversely it provides a concrete link between averaging and dynamical systems. We discuss this perspective further in Chapter 10. Note also that the perturbation expansion that we use in this chapter is, at a high level, similar to those used in the remainder of Part II. It differs in one significant respect, however: all the remaining chapters involve perturbation expansions for the approximation of linear problems (by working with the backward equation, and rely on repeated use of the Fredholm alternative. In this chapter the strategy underlying the perturbation expansion is somewhat different, as the problem for the graph $\Psi$ is nonlinear and the Fredholm alternative is not used.

Invariant manifolds in general are described in [131] and [331]. These books have considerable emphasis on the construction of unstable, stable and center manifolds for invariant sets of the equation (4.1.1). In particular, for the case of the simplest invariant set, an equilibrium point, we may change coordinates to a frame in which the origin $0$ is an equilibrium point and (4.1.1) takes the form

$$\frac{dz}{dt} = Lz + h_1(z), \quad z(0) = z_0.$$

Here $h_1(z)$ is small compared to $z \to 0$. In the case of a hyperbolic equilibrium point the invariant subspaces of $L$ split into stable and unstable spaces. If we let $P$ denote the orthogonal projection onto the stable space, and $Q = I - P$ denote the

orthogonal projection onto the unstable space, then introducing $x = Pz, y = Qz$ we obtain the equations

$$\frac{dx}{dt} = L_1 x + f_1(x, y),$$
$$\frac{dy}{dt} = L_2 y + g_1(x, y).$$

The stable manifold is (locally near the origin) representable as a graph $y = \Theta(x)$; likewise the unstable manifold is representable as a graph $x = \Phi(y)$. The center manifold is similar to the stable manifold, but occurs when, for example, $P\mathcal{Z}$ comprises neutral directions in $L$. Center manifolds in particular are discussed in [57]. The special case where the neutral spectrum of $L$ contains a pair of complex conjugate eigenvalues leads to the Hopf bifurcation theorem; see [212].

These special invariant manifold theorems, concerning behaviour near fixed points, show the central role of graphs relating one set of variables to another in the construction of invariant manifolds. Such a graph is at the heart of our construction of what is sometimes termed a *slow manifold* for (8.2.1). Early studies of the approximation of ODE with attracting slow manifold by differential-algebraic equations includes the independent work of Levinson and of Tikhonov (see O'Malley [239] and Tikhonov et al. [317]). As mentioned in section 8.3 the simplest version of the approximation result requires the fast variable $y$ to be initialized close to the invariant manifold. However, even if it is not, an *initial layer* (sometimes termed *boundary layer*) can be introduced to extend the approximation result, and studied through the method of matched asymptotic expansions; see [324] and [71, 272].

Our construction of an invariant manifold uses the explicit slaving of $y$ to $x$ through the asymptotically stable fixed points of (8.2.2). More generally, 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., Carr [57], Wiggins [331]), slow manifolds (Kreiss [178]) and inertial manifolds (Constantin et al. [69]). In particular the inertial manifold construction shows how ideas from invariant manifold theory extend naturally to infinite dimensions in the context of dissipative PDEs.

References to numerical methods for dynamical systems, and for the computation of invariant manifolds in particular, may be found in [305]. It is also possible to construct invariant manifolds for stochastic (partial) differential equations. See, for example, [39, 38, 41, 77, 329, 328] and the references therein.

## 8.7 Exercises

1. Consider the equations

$$\frac{dx}{dt} = \lambda x + a_0 x^3 + a_1 xy,$$
$$\frac{dy}{dt} = -y + \sum_{i=0}^{2} b_i x^i y^{2-i}.$$

Here $\lambda \in \mathbb{R}$ and the $a_i$ and $b_i$ are also real numbers. For each fixed $t$ we have $x(t) \in \mathbb{R}$ and $y(t) \in \mathbb{R}$. Show that the scaling

$$x \to \varepsilon x, y \to \varepsilon^2 y, \lambda \to \varepsilon^2 \lambda, t \to \varepsilon^{-2} t$$

puts this system in a form to which the perturbation techniques of this section apply. Deduce that the center manifold has the form

$$\frac{dX}{dt} = \lambda X + AX^3$$

where $A = a_0 + a_1 b_2$.

2. Assume $\varepsilon > 0$, $A \in \mathbb{R}^{l \times l}$ and $B \in \mathbb{R}^{(d-l)\times(d-l)}$. Consider the equations

$$\frac{dx}{dt} = Ax + \varepsilon f_0(x, y),$$
$$\frac{dy}{dt} = -\frac{1}{\varepsilon} By + g_0(x, y),$$

for $\varepsilon \ll 1$ and $x \in \mathbb{R}^l$, $y \in \mathbb{R}^{d-l}$. Assume that $B$ is symmetric positive-definite. Find the first three terms in an expansion for an invariant manifold representing $y$ as a graph over $x$.

3. Assume $\varepsilon > 0$ and $B \in \mathbb{R}^{(d-l)\times(d-l)}$. Consider the equations

$$\frac{dx}{dt} = f(x, y),$$
$$\frac{dy}{dt} = -\frac{1}{\varepsilon}\Big(By - \tilde{g}(x)\Big),$$

for $\varepsilon \ll 1$ and $x \in \mathbb{R}^l$, $y \in \mathbb{R}^{d-l}$.

a) Assume that $B$ is symmetric positive-definite. Find the first term in an expansion for an invariant manifold representing $y$ as a graph over $x$.

b) Consider the case $d - l = 2$, $\tilde{g} \equiv 0$ and

$$B = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

What happens to the solution as $\varepsilon \to 0$?

# 9

# Averaging for Markov Chains

## 9.1 Introduction

Perhaps the simplest setting in which to expose variable elimination for stochastic dynamical problems is to work in the setting of Markov chains. In this context it is natural to study situations where a subset of the variables evolves rapidly compared with the remainder, and can be replaced by their averaged effect. In Section 9.2 we describe the unaveraged Markov chain and in Section 9.3 we present the averaged equations; the averaged equations are derived in Section 9.4 and an example is given in Section 9.5. In Section 9.6 we discuss various extensions of the results from this chapter and make some bibliographical remarks.

## 9.2 Full Equations

We work in the set-up of Chapter 5 and consider the backward equation

$$\frac{dv}{dt} = Qv. \tag{9.2.1}$$

Recall that this equation, with $v(0) = \phi$, has the property that

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

where $\mathbb{E}$ denotes expectation with respect to the Markov transition probabilities. We assume that the generator $Q$ [1] takes the form

$$Q = \frac{1}{\varepsilon} Q_0 + Q_1, \tag{9.2.2}$$

with $0 < \varepsilon \ll 1$. We study situations where the state space is indexed by two variables, $x$ and $y$, and the leading order contribution in $Q$, namely $Q_0$, corresponds

[1] In this chapter we denote the generator by $Q$ rather than $L$ because we use index $l$ for the state-space; thus we wish to avoid confusion with the components of the generator.

to fast ergodic dynamics in $y$, with $x$ frozen. Averaging over $y$ then gives the effective reduced dynamics for $x$.

The precise situation is as follows. Our state space is $\mathcal{I} := \mathcal{I}_x \times \mathcal{I}_y$ with $\mathcal{I}_x$, $\mathcal{I}_y \subseteq \{1, 2, \cdots\}$. We let $q((i,k),(j,l))$ denote the element of the generator associated with transition from $(i,k) \in \mathcal{I}_x \times \mathcal{I}_y$ to $(j,l) \in \mathcal{I}_x \times \mathcal{I}_y$.[2] Consider now a family of Markov chains on $\mathcal{I}_y$, indexed by $i \in \mathcal{I}_x$. We write the generator as $A_0(i)$ with entries as $a_0(k,l;i)$; the indices denote transition from $k \in \mathcal{I}_y$ to $l \in \mathcal{I}_y$ for given fixed $i \in \mathcal{I}_x$. We assume that, for each $i \in \mathcal{I}_x$, $A_0(i)$ generates an ergodic Markov chain on $\mathcal{I}_y$. Hence $A_0(i)$ has a one-dimensional null space for each fixed $i$, and [3]

$$\begin{array}{rl} \sum_l a_0(k,l;i) & = 0, \quad (i,k) \in \mathcal{I}_x \times \mathcal{I}_y, \\ \sum_k \rho^\infty(k;i) a_0(k,l;i) & = 0, \quad (i,l) \in \mathcal{I}_x \times \mathcal{I}_y. \end{array} \tag{9.2.3}$$

This is the index form of equations (5.6.2) with $L$ replaced by $A_0(i)$. Without loss of generality we choose the normalization

$$\sum_k \rho^\infty(k;i) = 1 \quad \forall i \in \mathcal{I}_x.$$

Thus $\rho^\infty(i) = \{\rho^\infty(k;i)\}_{k \in \mathcal{I}_y}$ is the invariant distribution of a Markov chain on $\mathcal{I}_y$, indexed by $i \in \mathcal{I}_x$.

Similarly to the above we introduce the generators of a Markov chain on $\mathcal{I}_x$, parameterized by $k \in \mathcal{I}_y$. We denote the generator by $A_1(k)$ with indices $a_1(i,j;k)$; the indices denote transition from $i \in \mathcal{I}_x$ to $j \in \mathcal{I}_x$, for each fixed $k \in \mathcal{I}_y$. With this notation for the $A_0, A_1$ we introduce generators $Q_0, Q_1$ of Markov chains on $\mathcal{I}_x \times \mathcal{I}_y$ by

$$\begin{array}{l} q_0((i,k),(j,l)) = a_0(k,l;i)\delta_{ij}, \\ q_1((i,k),(j,l)) = a_1(i,j;k)\delta_{kl}. \end{array} \tag{9.2.4}$$

Here $\delta_{ij}$ is the usual Kronecker delta. In the construction of $Q_0$ (resp. $Q_1$) the Kronecker delta represents the fact that no transitions are taking place in $\mathcal{I}_x$ (resp. $\mathcal{I}_y$).

To confirm that $Q_0, Q_1$ as defined are indeed generators, notice that non-diagonal entries $(i,k) \neq (j,l)$ are nonnegative because $A_0$ and $A_1$ are generators. Also

$$\begin{aligned} \sum_{j,l} q_0((i,k),(j,l)) &= \sum_{j,l} a_0(k,l;i)\delta_{ij} \\ &= \sum_l a_0(k,l;i) \\ &= 0 \end{aligned}$$

[2] In this chapter, and in Chapter 16, we will not use suffices to denote the dependence on the state space as the double-indexing makes this a cluttered notation. Hence we use $q((i,k),(j,l))$ rather than $q_{(i,k),(j,l)}$.

[3] Summation is always over indices in $\mathcal{I}_x$ or $\mathcal{I}_y$ in this chapter. It should be clear from the context which of the two sets is being summed over.

by (9.2.3). A similar calculation shows that

$$\sum_{j,l} q_1((i,k),(j,l)) = 0,$$

using the fact that

$$\sum_j a_1(i,j;k) = 0 \quad \forall\, (i,k) \in \mathcal{I}_x \times \mathcal{I}_y,$$

since $A_1(k)$ is a generator for each fixed $k$. Thus $Q_0, Q_1$ are also the generators of Markov chains. Finally note that any linear combination of generators, via positive scalar constants, will also be a generator. Hence (9.2.2) defines a generator for any $\varepsilon > 0$.

## 9.3 Simplified Equations

We define the generator $\bar{Q}_1$ of a Markov chain on $\mathcal{I}_x$ by:

$$\bar{q}_1(i,j) = \sum_k \rho^\infty(k;i) a_1(i,j;k). \tag{9.3.1}$$

Notice that $\bar{q}_1(i,j) \geqslant 0$ for $i \neq j$ because $\rho^\infty(k;i) \geqslant 0$ and $a_1(i,j;k) \geqslant 0$ for $i \neq j$. Furthermore

$$\begin{aligned}\sum_j \bar{q}_1(i,j) &= \sum_k \rho^\infty(k;i) \left( \sum_j a_1(i,j;k) \right) \\ &= 0.\end{aligned}$$

Hence $\bar{Q}_1$ is the generator of a Markov chain.

**Result 9.1.** *Consider equation (9.2.1) under assumption (9.2.2). Then for $\varepsilon \ll 1$ and times $t$ upto $\mathcal{O}(1)$ the finite dimensional distributions of $x \in \mathcal{I}_x$ are approximated by a Markov chain $X$ with generator $\bar{Q}_1$.*

We emphaszie that $x$ is not itself Markovian: only the pair $(x,y)$ is. As discussed above, $\bar{Q}_1$ is the generator of a Markov chain on $\mathcal{I}_x$ alone, and the dynamics in $\mathcal{I}_y$ has been eliminated through averaging. Thus the approximate variable $X$ is Markovian and is governed by the backward equation

$$\frac{dv_0}{dt} = \bar{Q}_1 v_0. \tag{9.3.2}$$

We now provide justification for this elimination of variables, by means of perturbation expansion.

## 9.4 Derivation

The method used is to show that the backward equation for the full Markov chain in $(x, y) \in \mathcal{I}_x \times \mathcal{I}_y$ can be approximated by the backward equation (9.3.2) for $x \in \mathcal{I}_x$ alone. We consider equation (9.2.1) under (9.2.2). We have the backward equation

$$\frac{dv}{dt} = \Big(\frac{1}{\varepsilon}Q_0 + Q_1\Big)v.$$

Unlike the previous chapter, where we approximated a nonlinear PDE containing a small parameter $\varepsilon$, here the problem is linear. In the following five chapters, all our perturbation expansions are for similar linear equations. The derivation here is hence prototypical of what follows.

We seek solutions $v = v(i, k, t)$ in the form of the multiscale expansion

$$v = v_0 + \varepsilon v_1 + \mathcal{O}(\varepsilon^2). \tag{9.4.1}$$

Substituting and equating coefficients of powers of $\varepsilon$ to zero we find

$$\mathcal{O}(\tfrac{1}{\varepsilon})\ Q_0 v_0 = 0, \tag{9.4.2a}$$

$$\mathcal{O}(1)\ Q_0 v_1 = -Q_1 v_0 + \frac{dv_0}{dt}. \tag{9.4.2b}$$

By (9.2.3) we deduce from (9.4.2a) that $v_0$ is independent of $k \in \mathcal{I}_y$. Abusing notation, we write

$$v_0(i, k, t) = v_0(i, t)1(k) \tag{9.4.3}$$

where $1(k) = 1$ for all $k \in \mathcal{I}_y$. The operator $Q_0$ is singular and hence, for (9.4.2b) to have a solution, the Fredholm alternative implies the solvability condition

$$-Q_1 v_0 + \frac{dv_0}{dt} \perp \text{Null}\,\{Q_0^T\}. \tag{9.4.4}$$

From (9.2.3) we deduce that the null space of $Q_0^T$ is characterized by

$$\sum_{k,i} \rho^\infty(k; i)c(i)q_0((i, k), (j, l)) = 0, \tag{9.4.5}$$

for any vector $c = \{c(i)\}$ on $\mathcal{I}_x$. Using (9.4.3) we find that

$$\begin{aligned}\frac{dv_0}{dt} - Q_1 v_0 &= \frac{dv_0}{dt}(i, t)1(k) - \sum_{j,l} a_1(i, j; k)\delta_{kl}v_0(j, t)1(l)\\ &= \Big(\frac{dv_0}{dt}(i, t) - \sum_j a_1(i, j; k)v_0(j, t)\Big)1(k).\end{aligned}$$

Imposing the solvability condition (9.4.4) by means of (9.4.5) we obtain

$$\sum_{k,i} \rho^\infty(k; i)c(i)\Big(\frac{dv_0}{dt}(i, t) - \sum_j a_1(i, j; k)v_0(j, t)\Big) = 0,$$

which implies that

$$\sum_i c(i)\Big(\frac{dv_0}{dt}(i,t) - \sum_j \bar{Q}_1(i,j)v_0(j,t)\Big) = 0.$$

Since $c$ is an arbitrary vector on $\mathcal{I}_x$ we deduce that each component of the sum over $i$ is zero. This yields (9.3.2).

## 9.5 Application

Consider a simple example where $\mathcal{I}_x = \mathcal{I}_y = \{1,2\}$. Thus we have a four–state Markov chain on $\mathcal{I} = \mathcal{I}_x \times \mathcal{I}_y$. We assume that the generators of the Markov chains on $\mathcal{I}_y$ and $\mathcal{I}_x$ are given by

$$A_0(i) = \begin{pmatrix} -\theta_i & \theta_i \\ \phi_i & -\phi_i \end{pmatrix}$$

and

$$A_1(k) = \begin{pmatrix} -\alpha_k & \alpha_k \\ \beta_k & -\beta_k \end{pmatrix},$$

respectively. In the first (resp. second) of these Markov chains $i \in \mathcal{I}_x$ (resp. $k \in \mathcal{I}_y$) is a fixed parameter. The parameters $\theta_i, \phi_i, \alpha_k, \beta_k$ are all non-negative.

If we order the four states of the Markov chain as $(1,1),(1,2),(2,1),(2,2)$ then the generators $Q_0$ and $Q_1$ are given by

$$Q_0 = \begin{pmatrix} -\theta_1 & \theta_1 & 0 & 0 \\ \phi_1 & -\phi_1 & 0 & 0 \\ 0 & 0 & -\theta_2 & \theta_2 \\ 0 & 0 & \phi_2 & -\phi_2 \end{pmatrix} \tag{9.5.1}$$

and

$$Q_1 = \begin{pmatrix} -\alpha_1 & 0 & \alpha_1 & 0 \\ 0 & -\alpha_2 & 0 & \alpha_2 \\ \beta_1 & 0 & -\beta_1 & 0 \\ 0 & \beta_2 & 0 & -\beta_2 \end{pmatrix}. \tag{9.5.2}$$

Note that any linear combination of $Q_0$ and $Q_1$ will have zeros along the anti-diagonal and hence the same is true of $Q$; this reflects the fact that, by construction, transitions in both $\mathcal{I}_x$ and $\mathcal{I}_y$ do not happen simultaneously.

The invariant density of the Markov chain with generator $A_0(i)$ is $\rho^\infty(i) = (\lambda_i, 1-\lambda_i)^T$ with $\lambda_i = \phi_i/(\theta_i + \phi_i)$. Recall that the averaged Markov chain on $\mathcal{I}_x$ has generator $\bar{Q}_1$ with entries

$$q_1(i,j) = \sum_k \rho^\infty(k;i) a_1(i,j;k)$$
$$= \lambda_i a_1(i,j;1) + (1-\lambda_i) a_1(i,j;2).$$

Thus

$$\bar{Q}_1 = \begin{pmatrix} -\lambda_1\alpha_1 - (1-\lambda_1)\alpha_2 & \lambda_1\alpha_1 + (1-\lambda_1)\alpha_2 \\ \lambda_2\beta_1 + (1-\lambda_2)\beta_2 & -\lambda_2\beta_1 - (1-\lambda_2)\beta_2 \end{pmatrix}. \tag{9.5.3}$$

## 9.6 Discussion and Bibliography

Two recent monographs where multiscale problems for Markov chains are studied are [335], [336]. See also [291] for a broad discussion of averaging and dimension reduction in stochastic dynamics. Markov chain approximations for SDEs, especially in the large deviation limit, are studied in [111]. Computational methods for multiscale Markov chains are discussed in [85, 86]. Diffusion limits of ODEs driven by Markov Chains are studied in [245]. See also [96] for the proof of a related diffusion limit theorem. For a connection between Markov chains and center manifolds see [262].

In this chapter we have presented averaging for Markov chains. Homogenization (i.e. central limit theorem) results for Markov chains can be found in [184].

In deriving the approximate equation we implicitly assume that the original Markov chain is prepared in a state which does not depend upon the parts of the state space in $\mathcal{I}_y$. If this is not the case then a similar analysis can still be carried out, but an initial layer must be included, over time of order $\mathcal{O}(\varepsilon)$, on which $v(t)$ adjusts from being a function on $\mathcal{I}_x \times \mathcal{I}_y$ to being a function only on $\mathcal{I}_x$, to leading order.

## 9.7 Exercises

1. Find a multiscale expansion for the invariant measure of the Markov chain with generator $Q = \frac{1}{\varepsilon}Q_0 + Q_1$ when $Q_0, Q_1$ are given by (9.5.1), (9.5.2).
2. Find the invariant measure of $\bar{Q}_1$ given by (9.5.3) and interpret your findings in the light of your answer to the previous question.
3. Consider the SDE (6.5.1). Assume that $u$ is governed by a two-state Markov chain, with states $\{-1,+1\}$. Write down the generator for the resulting Markov process in $(z,u)$, on the assumption that the generator for the Markov chain has the form
$$L(z) = \begin{pmatrix} -\theta(z) & \theta(z) \\ \phi(z) & -\phi(z) \end{pmatrix}.$$
4. Consider the same set-up as in the previous question but where the two-state Markov chain now has generator $\frac{1}{\varepsilon}L(z)$ with $L(z)$ as given in the previous question. Use the method of averaging to find the averaged SDE in $z$ in the limit $\varepsilon \to 0$, where $u$ may be eliminated.

5. Let $u$ be a two state continuous time Markov chain with generator as in the previous question. Consider the ODE

$$\frac{dz}{dt} = \lambda(u)z, \quad t \in [0, \infty).$$

Assume that $\lambda(-1) < 0$ and $\lambda(+1) > 0$. Use multiscale analysis to determine conditions under which the trajectories of $z$ do not grow.

6. Let $u$ be a Markov chain on a finite state-space with generator $Q$ taking the form

$$Q = \frac{1}{\varepsilon}Q_0 + Q_1.$$

Assume that the $Q_i$ are generators of Markov chains for $i = 0, 1$ and that $Q$ has a two-dimensional null-space:

$$\mathcal{N}(Q_0) = \text{span}\{\phi_0, \phi_1\}.$$

Derive a two state Markov chain which approximates the dynamics in this null-space.

# 10

# Averaging for ODEs and SDEs

## 10.1 Introduction

Here we take the averaging principle developed in the previous chapter for Markov chains, and apply it to ODEs and SDEs. The unifying theme is the approximate solution of the backward equation by means of an appropriate perturbation expansion, and consequent elimination of variables.

In Section 10.2 we present the equations that we will study and in Section 10.3 we present the averaged equations. Section 10.4 contains the derivation of the averaged equations; the derivation is carried out in the case where the fast process is stochastic. In Section 10.5 we study how the deterministic situation may be handled. Section 10.6 contains two illustrative examples. Extensions of the results presented in this chapter, together with bibliographical remarks, are given in Section 10.7.

## 10.2 Full Equations

We write $z$ solving (6.1.1) as $z = (x^T, y^T)^T$ and consider the case where

$$\frac{dx}{dt} = f(x,y), \quad x(0) = x_0, \tag{10.2.1a}$$

$$\frac{dy}{dt} = \frac{1}{\varepsilon} g(x,y) + \frac{1}{\sqrt{\varepsilon}} \beta(x,y) \frac{dV}{dt}, \quad y(0) = y_0, \tag{10.2.1b}$$

with $\varepsilon \ll 1$ and $V$ a standard Brownian motion. Here $x \in \mathcal{X}$, $y \in \mathcal{Y}$, $z \in \mathcal{Z}$ and the notation is as in Sections 4.1 and 6.1.

In Chapter 8 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. As in the previous chapter on Markov chains, we average out the fast variable $y$, over an appropriate invariant measure. We now make these heuristics precise. We define the generators

$$\mathcal{L}_0 = g(x,y)\cdot\nabla_y + \frac{1}{2}B(x,y) : \nabla_y\nabla_y, \tag{10.2.2a}$$

$$\mathcal{L}_1 = f(x,y)\cdot\nabla_x, \tag{10.2.2b}$$

where $B(x,y) = \beta(x,y)\beta(x,y)^T$. To carry out the averaging procedure in this section the most useful way to make an ergodicity assumption is to assume that, for each fixed $x$, $\mathcal{L}_0$ has one dimensional null space characterized by

$$\mathcal{L}_0 1(y) = 0, \tag{10.2.3a}$$

$$\mathcal{L}_0^*\rho^\infty(y;x) = 0. \tag{10.2.3b}$$

Here $1(y)$ denotes constants in $y$. In the case where $\mathcal{Y} = \mathbb{T}^d$ the operators $\mathcal{L}_0$ and $\mathcal{L}_0^*$ are equipped with periodic boundary conditions. In this case these assumptions about the null spaces of $\mathcal{L}_0$ and $\mathcal{L}_0^*$ are shown to hold if $B(x,y)$ is strictly positive-definite, uniformly in $(x,y) \in \mathcal{X}\times\mathcal{Y}$, as shown in Theorem 6.16. In more general situations, such as when $\mathcal{Y} = \mathbb{R}^d$ or when the matrix valued function $B(x,y)$ is degenerate, similar rigorous justifications are possible, but the functional setting is more complicated, typically employing weighted $L^p$–spaces which characterize the decay of the invariant density at infinity. See the remarks in Section 18.4.

## 10.3 Simplified Equations

We assume that the generator of the fast process $y(t)$, namely $\mathcal{L}_0$, satisfies (10.2.3) for every $x \in \mathcal{X}$. Define the vector field $F$ by

$$F(x) = \int_{\mathcal{Y}} f(x,y)\,\mu_x(dy). \tag{10.3.1}$$

with $\mu_x(dy) = \rho^\infty(y;x)dy$.

**Result 10.1.** *For $\varepsilon \ll 1$ and times $t$ upto $\mathcal{O}(1)$, $x(t)$ solving* (10.2.1) *is approximated by $X$ solving*

$$\frac{dX}{dt} = F(X), \quad X(0) = x_0. \tag{10.3.2}$$

*Remark 10.2.* A similar result holds even in the case where the equation for the slow variable $x$ is stochastic and has the form

$$\frac{dx}{dt} = f(x,y) + \alpha(x,y)\frac{dU}{dt}, \quad x(0) = x,$$

with $U$ a standard Brownian motion, indepenent of $V$. Under the assumptions of Result 10.1 the averaged equation becomes

$$\frac{dX}{dt} = F(X) + A(X)\frac{dU}{dt}, \quad X(0) = x,$$

where $F(X)$ is the same as above and

$$A(X)A(X)^T = \int_{\mathcal{Y}} \alpha(x,y)\alpha(x,y)^T\,\mu_x(dy).$$

See Exercise 1. □

## 10.4 Derivation

As for Markov chains, we derive the averaged equations by working with the backward Kolmogorov equation. Let

$$v(x,y,t) = \mathbb{E}\Big(\phi(x(t),y(t))|x(0)=x, y(0)=y\Big).$$

The backward equation (6.3.4) for the SDE (10.2.1) is

$$\frac{\partial v}{\partial t} = \frac{1}{\varepsilon}\mathcal{L}_0 v + \mathcal{L}_1 v. \tag{10.4.1}$$

Here $\mathcal{L}_0, \mathcal{L}_1$ are given by (10.2.2) and $z$ in (6.3.4) is $(x,y)$ here. Note that $\mathcal{L}_0$ is a differential operator in $y$, in which $x$ appears as a parameter. Thus we must equip it with boundary conditions. We simply assume that, with suitable boundary conditions imposed, (10.2.3) holds. In the case where $\mathcal{Y} = \mathbb{T}^d$ and periodic boundary conditions are used the rigorous results of Chapter 7 apply and the ergodicity assumption on the fast process is satisfied. Note, however, that other functional settings are also possible; the key in what follows is application of the Fredholm alternative to operator equations defined through $\mathcal{L}_0$.

We seek a solution to (10.4.1) in the form of the multiscale expansion

$$v = v_0 + \varepsilon v_1 + \mathcal{O}(\varepsilon^2)$$

and obtain

$$\mathcal{O}(1/\varepsilon)\ \mathcal{L}_0 v_0 = 0, \tag{10.4.2a}$$

$$\mathcal{O}(1)\quad \mathcal{L}_0 v_1 = -\mathcal{L}_1 v_0 + \frac{\partial v_0}{\partial t}. \tag{10.4.2b}$$

Equation (10.4.2a) implies that $v_0$ is in the null space of $\mathcal{L}_0$ and hence, by (10.2.3) and ergodicity, is a function only of $(x,t)$. Fix $x$. Then the Fredholm alternative for (10.4.2b), viewed as a differential equation in $y$, shows that

$$-\mathcal{L}_1 v_0 + \frac{\partial v_0}{\partial t} \perp \text{Null}\,\{\mathcal{L}_0^*\}.$$

By (10.2.3) this implies that

$$\int_{\mathcal{Y}} \rho^\infty(y;x)\Big(\frac{\partial v_0}{\partial t}(x,t) - f(x,y)\cdot\nabla_x v_0(x,t)\Big)dy = 0.$$

Since $\rho^\infty$ is a probability density we have $\int_{\mathcal{Y}} \rho^\infty(y;x)dy = 1$. Hence

$$\frac{\partial v_0}{\partial t} - \left(\int_{\mathcal{Y}} f(x,y)\mu_x(y)dy\right)\cdot\nabla_x v_0(x,t) = 0$$

so that by (10.3.1),

$$\frac{\partial v_0}{\partial t} - F(x)\cdot\nabla_x v_0 = 0$$

This is the backward equation for (10.3.2); indeed the method of characteristics as given by in Result 4.6 shows that we have the required result.

## 10.5 Deterministic Problems

In this section we provide a viewpoint on the averaged equation which is useful for two reasons: it applies when the equations (10.2.1) are deterministic; and it forms the basis of numerical methods to compute effective equations, in either the deterministic or stochastic contexts. Our starting point is to analyze the behavior of the fast dynamics in $y$ with $x$ being a fixed parameter.

Let $\varphi_x^t(y)$ be the solution operator of the fast dynamics with $x$ a fixed parameter and $\varepsilon = 1$. To be precise, for fixed $\xi$,

$$\frac{d}{dt}\varphi_\xi^t(y) = g(\xi, \varphi_\xi^t(y)) + \beta(\xi, \varphi_\xi^t(y))\frac{dV}{dt}, \qquad \varphi_\xi^0(y) = y. \tag{10.5.1}$$

As in Chapter 8, $y(t)$ solving (10.2.1b) is given by $y(t) \approx \varphi_{x(0)}^{t/\varepsilon}(y)$ for times $t$ which are $o(1)$, so that $x$ has not evolved very much. Assume that (10.5.1) is ergodic with invariant measure $\mu_\xi$. On timescales small compared to 1 and large compared to $\varepsilon$ we expect that $x(t)$ is approximately frozen and that $y(t)$ will traverse its ($x$-dependent) invariant measure on this timescale because it is evolving quickly. Thus it is natural to average $y(t)$ in the $x(t)$ equation, against the invariant measure for (10.5.1) with $\xi = x(t)$.

In the case where $\beta \equiv 0$ then $\varphi_\xi^t(y)$ coincides with the solution of (8.2.2). When $\beta \neq 0$, note that $\varphi_\xi^t(y)$ depends on the Brownian motion $\{V(s)\}_{s\in[0,t]}$ and hence is a stochastic process. Rather than assuming convergence to a fixed point, as we did in (8.2.3), we assume here that $\varphi_\xi^t(y)$ is ergodic (see Section 6.4). This implies that the measure defined by

$$\mu_x(A) = \lim_{T\to\infty} \frac{1}{T}\int_0^T I_A(\varphi_x^t(y))\, dt, \quad A \subseteq \mathbb{T}^d, \tag{10.5.2}$$

exists, for $I_A$ the indicator function of arbitrary Borel sets $A \subseteq \mathcal{Y}$. The averaged vector field $F$ in (10.3.1) can be defined using this measure.

When working with an SDE ($\beta \neq 0$) then it is natural to assume that $\mu_x(\cdot)$ has a density with respect to the Lebesgue measure so that $\mu_x(dy) = \rho^\infty(y;x)dy$. In fact, under appropriate assumptions on the coefficients $g(x,y)$ and $\beta(x,y)$ it is possible to prove that such a density exists. However, we will illustrate by means of an example arising in Hamiltonian mechanics that this assumption is not necessary. Note also that the situation in Chapter 8 corresponds to the measure $\mu_x(dy)$ being a Dirac mass characterizing the invariant manifold: $\mu_x(dy) = \delta(y - \eta(x))dy$. In this case we obtain

$$F(x) = f(x, \eta(x)).$$

This is precisely the vector field in (8.3.2) and so the simplified equations in Chapter 8 are a special case of those derived here. However, we derived Result 10.1 in the case where $\beta$ is nonzero and we assumed that the measure $\mu$ has a smooth density $\rho^\infty(y;x)$ with respect to Lebesgue measure; that is, we assumed that (10.2.3) holds and we have that $\mu_x(dy) = \rho^\infty(y;x)dy$. It is useful to have an expression for the

averaged equation which is also valid for deterministic problems, and for the numerical construction of $F$ in either deterministic or random problems. We do this by representing ergodic averages via time averages.

**Result 10.3.** *An alternative representation of $F(x)$ is via a time average:*

$$F(x) = \lim_{T\to\infty} \frac{1}{T} \int_0^T f\left(x, \varphi_x^s(y)\right)\, ds. \tag{10.5.3}$$

This representation is found by using (10.5.2) to evaluate (10.3.1). Note that, by ergodicity, the resulting average does not depend upon $y$.

## 10.6 Applications

We consider two applications of the averaging principle, the first in the context of SDEs, and the second in the context of Hamiltonian ODEs.

### 10.6.1 A Skew-Product SDE

Consider the equations

$$\begin{aligned}
\frac{dx}{dt} &= (1-y^2)x, \\
\frac{dy}{dt} &= -\frac{\alpha}{\varepsilon}y + \sqrt{\frac{2\lambda}{\varepsilon}}\frac{dV}{dt}.
\end{aligned}$$

Here $\mathcal{X} = \mathcal{Y} = \mathbb{R}$. It is of interest to know whether $x$ will grow in time, or remain bounded. We can get insight into this question in the limit $\varepsilon \to 0$ by deriving the averaged equations. Note that $y$ is a time-rescaling of the OU process from Example 6.19. The invariant measure for the ergodic process $y$ is a mean zero Gaussian: $\mathcal{N}(0, \frac{\lambda}{\alpha})$ (see Example 6.19). Note that this measure does not depend on $x$ and hence has density $\rho^\infty(y)$ only. The averaged vector field $F$ is here defined by

$$F(x) = \Big(1 - \int_{\mathbb{R}} \rho^\infty(y) y^2 dy\Big) x$$

where $\rho^\infty$ is the density associated with Gaussian $\mathcal{N}(0, \frac{\lambda}{\alpha})$. Thus

$$\int_{\mathbb{R}^d} \rho^\infty(y) y^2 dy = \frac{\lambda}{\alpha}$$

and

$$F(x) = \Big(1 - \frac{\lambda}{\alpha}\Big) x.$$

Hence the averaged equation is

$$\frac{dX}{dt} = \Big(1 - \frac{\lambda}{\alpha}\Big)X.$$

From this we deduce that trajectories of $x$ will explode if $\lambda < \alpha$ and will contract if $\lambda > \alpha$. If $\lambda = \alpha$ then the averaged vector field is zero. In this situation we need to rescale time $t \mapsto t/\varepsilon$ to obtain the problem

$$\begin{aligned}\frac{dx}{dt} &= \frac{1}{\varepsilon}(1-y^2)x,\\ \frac{dx}{dt} &= -\frac{\alpha}{\varepsilon^2}y + \sqrt{\frac{2\alpha}{\varepsilon^2}}\frac{dv}{dt}.\end{aligned}$$

On this longer timescale nontrivial dynamics occur. SDEs of this form are the topic of Chapter 11, and this specific example is considered in Section 11.7.

### 10.6.2 Hamiltonian Mechanics

[1] In many applications Hamiltonian systems with strong potential forces, responsible for fast, small amplitude oscillations around a constraining sub-manifold, are encountered. It is then of interest to describe the evolution of the slowly evolving degrees of freedom by averaging over the rapidly oscillating variables. We give an example of this. The example is interesting because it shows that the formalism of this chapter can be extended to pure ordinary differential equations, with no noise present; it also illustrates that it is possible to deal with situations where the limiting measure $\mu$ retains some memory of initial conditions – in this case the total energy of the system.

Consider a two-particle system with Hamiltonian,

$$H(x,p,y,v) = \frac{1}{2}(p^2+v^2) + \Phi(x) + \frac{\omega(x)}{2\,\varepsilon^2}y^2, \tag{10.6.1}$$

where $(x,y)$ are the coordinates and $(p,v)$ are the conjugate momenta of the two particles, $\Phi(x)$ is a nonnegative potential and $\omega(x)$ is assumed to satisfy $\omega(x) \geqslant \bar{\omega} > 0$ for all $x$. The corresponding equations of motion are

$$\begin{aligned}\frac{dx}{dt} &= p,\\ \frac{dp}{dt} &= -\Phi'(x) - \frac{\omega'(x)}{2\varepsilon^2}y^2,\\ \frac{dy}{dt} &= v,\\ \frac{dv}{dt} &= -\frac{\omega(x)}{\varepsilon^2}y.\end{aligned}$$

We let $E$ denote the value of the Hamiltonian $H$ at time $t = 0$:

[1] This example was developed in collaboration with R. Kupferman.

$$E = H(x(0), p(0), y(0), v(0)).$$

Note that $E$ is the total energy of the two–particle system. We assume that $E$ is bounded independently of $\varepsilon$. Since the Hamiltonian $H$ is conserved in time, since $\Phi$ is nonnegative and $\omega \geqslant \bar{\omega}$, equation (10.6.1) implies that

$$y^2 \leqslant 2\varepsilon^2 E/\bar{\omega}.$$

Hence the solution approaches the submanifold $y = 0$ as $\varepsilon \to 0$. Note, however, that $y$ appears in the combination $y/\varepsilon$ in the $x$ equations and in the expression for the energy $H$. Thus it is natural to make the change of variables $\eta = y/\varepsilon$. The equations then read

$$\begin{aligned}
\frac{dx}{dt} &= p, \\
\frac{dp}{dt} &= -\Phi'(x) - \frac{\omega'(x)}{2}\eta^2, \\
\frac{d\eta}{dt} &= \frac{1}{\varepsilon}v, \\
\frac{dv}{dt} &= -\frac{\omega(x)}{\varepsilon}\eta.
\end{aligned} \tag{10.6.2}$$

In these variables we recover a system of the form (10.2.1) with "slow" variables, $x \leftarrow (x, p)$, and "fast" variables, $y \leftarrow (\eta, v)$. It is instructive to write the equation in second order form as

$$\begin{aligned}
\frac{d^2x}{dt^2} + \Phi'(x) + \frac{1}{2}\omega'(x)\eta^2 &= 0, \\
\frac{d^2\eta}{dt^2} + \frac{1}{\varepsilon^2}\omega(x)\eta &= 0.
\end{aligned}$$

The fast equations represent a harmonic oscillator whose frequency $\omega^{1/2}(x)$ is modulated by the $x$ variables.

Consider the fast dynamics, with $(x, p)$ frozen. The Hamiltonian for this fast dynamics is, for $\varepsilon = 1$ and $x$ frozen,

$$H_{\text{fast}} = \frac{1}{2}v^2 + \frac{\omega(x)}{2}\eta^2.$$

The energy of the fast system, at given $(x, p)$, which is conserved whilst $(x, p)$ is frozen, is found by subtracting the energy associated with the frozen variables from the total energy of the original system. We denote the result of this calculation by

$$E_{\text{fast}} = E - \frac{1}{2}p^2 - \Phi(x).$$

For fixed $x, p$ the dynamics in $\eta, v$ is confined to the energy shell $H_{\text{fast}}(v, \eta) = E_{\text{fast}}$. We denote this energy shell by $\mathcal{Y}(x, p)$, noting that it is parameterized by the frozen variables $(x, p)$.

The harmonic oscillator is studied in Example 4.17. Using the calculations therein, it follows that the average of the kinetic energy of the fast oscillator against the ergodic measure $\mu_{x,p}$ on $\mathcal{Y}(x,p)$ is

$$\int_{\mathcal{Y}(x,p)} \frac{\omega(x)}{2}\eta^2 \mu_{x,p}(d\eta, dv) = \frac{1}{2}\left[E - \frac{1}{2}p^2 - \Phi(x)\right].$$

Thus

$$\int_{\mathcal{Y}(x,p)} \frac{1}{2}\eta^2 \mu_{x,p}(d\eta, dv) = \frac{1}{2\omega(x)}\left[E - \frac{1}{2}p^2 - \Phi(x)\right].$$

Here $(x,p)$ are viewed as fixed parameters and the total energy $E$ is specified by the initial data of the whole system. The averaging principle states that the rapidly varying $\eta^2$ in the equation (10.6.2) for $p$ can be approximated by its ergodic average, giving rise to a closed system of equations for $(X,P) \approx (x,p)$. These are

$$\begin{aligned} \frac{dX}{dt} &= P, \\ \frac{dP}{dt} &= -\Phi'(X) - \frac{\omega'(X)}{2\omega(X)}\left[E - \frac{1}{2}P^2 - \Phi(X)\right], \end{aligned} \tag{10.6.3}$$

with initial data $E$, $X(0) = X_0 = x(0)$ and $P(0) = P_0 = p(0)$. It is verified below that $(X,P)$ satisfying (10.6.3) conserve the following *adiabatic invariant*

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

Thus, (10.6.3) reduces to the Hamiltonian form

$$\frac{dX}{dt} = P, \tag{10.6.4a}$$

$$\frac{dP}{dt} = -\Phi'(X) - J_0\left[\omega^{1/2}(X)\right]', \tag{10.6.4b}$$

where $J_0$ is given by

$$J_0 = \frac{1}{\omega^{1/2}(X_0)}\left[E - \frac{1}{2}P_0^2 - \Phi(X_0)\right].$$

This means that the influence of the stiff potential on the slow variables is to induce a Hamiltonian structure, but to replace the potential $\Phi(x)$ by an effective potential,

$$\Phi_{\text{eff}}(x) = \Phi(x) + J_0\,\omega^{1/2}(x).$$

Note that the limiting equation contains memory of the initial conditions for the fast variables, through the constant $J_0$. Thus the situation differs slightly from that covered by the conjunction of Results 10.1 and 10.3.

To verify that $J$ is indeed conserved in time, note that, from the definition of $J$ and from equation (10.6.3),

$$\begin{aligned}\frac{d}{dt}\Big(\omega^{\frac{1}{2}}(X)J\Big) &= \frac{d}{dt}\Big(E - \frac{1}{2}P^2 - \Phi(X)\Big)\\ &= -P\frac{dP}{dt} - \Phi'(X)\frac{dX}{dt}\\ &= \frac{P\omega'(X)}{2\omega(X)}\Big(E - \frac{1}{2}P^2 - \Phi(X)\Big)\\ &= \frac{P\omega'(X)}{2\omega^{\frac{1}{2}}(X)}J.\end{aligned}$$

But, since $\frac{dX}{dt} = P$, we find the alternate expression,

$$\begin{aligned}\frac{d}{dt}\Big(\omega^{\frac{1}{2}}(X)J\Big) &= \frac{1}{2}\frac{\omega'(X)}{\omega^{\frac{1}{2}}(X)}\frac{dX}{dt}J + \omega^{\frac{1}{2}}(X)\frac{dJ}{dt}\\ &= \frac{P\omega'(X)}{2\omega^{\frac{1}{2}}(X)}J + \omega^{\frac{1}{2}}(X)\frac{dJ}{dt}.\end{aligned}$$

Equating the two expressions gives

$$\frac{dJ}{dt} = 0,$$

since $\omega(X)$ is strictly positive.

## 10.7 Discussion and Bibliography

Averaging is based on some form or ergodicity of the fast process; whether this process is deterministic or stochastic is not of primary importance. However it is easier, in general, to establish ergodicity for stochastic problems and this is why our general developments are confined to this case. The averaging method applied to equations (10.2.1) is analyzed in an instructive manner in [240], where the Liouville equation is used to construct a rigorous proof of the averaged limit. It is sometimes possible to obtain averaging results in the nonergodic case, when the null space of the fast process is finite dimensional, rather than one dimensional. See [246, 326].

A detailed account of the averaging method for ODEs, as well as numerous examples, can be found in [281]. See also [13]. An English language review of the Russian literature can be found in [193]. An overview of the topic of slow manifolds, especially in the context of Hamiltonian problems, may be found in [199]. The paper [321] provides an overview of variable elimination in a wealth of problems with scale separation.

**Anosov's Theorem** is the name often given to the averaging principle in the context of ODEs – (10.2.1) with $\beta \equiv 0$. This theorem requires the fast dynamics to be ergodic. Often ergodicity fails due to the presence of "resonant zones"—regions in $\mathcal{X}$ for which the fast dynamics is not ergodic. Arnold and Neistadt [193] extended Anosov's result to situations in which the ergodicity assumption fails on a sufficiently

small set of $x \in \mathcal{X}$. Those results were further generalized and extended to the stochastic framework by Kifer, who also studied the diffusive and large deviation character of the discrepancy between the effective and exact solution [169, 170, 171, 172]. See also [111, Ch. 7].

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* (see [29, 309]). Artstein and co-workers considered a class of singularly perturbed system of type (10.2.1), with attention given to the limiting behavior of both slow and fast variables. In all of the above cases the pair $(x, y)$ can be shown to converge to $(X, \mu_X)$, where $X$ is the solution of

$$\frac{dX}{dt} = \int_{\mathbb{T}^d} f(X, y)\, \mu_X(dy),$$

and $\mu_X$ is the ergodic measure on $\mathbb{T}^d$; the convergence of $y$ to $\mu_X$ is in the sense of Young measures. (In the case of a fixed point the Young measure is a Dirac mass concentrated at a point.) A general theorem along these lines is proved in [17].

There are many generalizations of this idea. The case of nonautonomous fast dynamics, as well as a case with infinite dimensions are covered in [18]. Moreover, these results still make sense even if there is no unique invariant measure $\mu_x$, in which case the slow variables can be proved to satisfy a (nondeterministic) differential inclusion [19].

In the context of SDE, an interesting generalization of (10.2.1) is to consider systems of the form

$$\frac{dx}{dt} = f(x, y) + \alpha(x, y)\frac{dU}{dt}, \tag{10.7.1a}$$

$$\frac{dy}{dt} = \frac{1}{\varepsilon} g(x, y) + \frac{1}{\sqrt{\varepsilon}} \beta(x, y)\frac{dV}{dt}. \tag{10.7.1b}$$

The simplified equation is then an SDE, not an ODE (see Remark 10.2).This situation is a subcase of the set-up we consider in the next chapter. It can be obtained by setting $f_0 = 0$ in that chapter, letting $f_1 = f$ there, and by identifying $\varepsilon$ here with $\varepsilon^2$ in that chapter.

In the application section we studied the averaging principle for a two–scale Hamiltonian system. The systematic study of Hamiltonian problems with two timescales was initiated by Rubin and Ungar [277]. More recently the ideas of Neistadt, based on normal form theory, have been applied to such problems [32]; this approach is very powerful, yielding very tight, exponential, error estimates between the original and limiting variables. A different approach to the problem, using the techniques of time-homogenization [43], is the paper [44]. The example presented in Section 10.6.2 is taken from that paper. The heuristic derivation we have given here is made rigorous in [44], using time-homogenization techniques, and it is also generalized to higher dimension. Resonances become increasingly important as the codimension, $m$, increases, limiting the applicability of the averaging approach to such two-scale Hamiltonian systems (Takens [306]).

Numerical work on multiscale ODEs and SDEs is overviewed in the next chapter.

## 10.8 Exercises

1. Derive the averaged equation resulting from the SDE (10.7.1) under the assumption that $U$ and $V$ are independent, standard Brownian motions (see Remark 10.2).
2. Let $\Phi : \mathcal{X} \times \mathcal{Y} : \mathbb{R}^+$ and consider the equations

$$\frac{dx}{dt} = -\nabla_x \Phi(x,y) + \sqrt{2\sigma}\frac{dU}{dt}$$
$$\frac{dy}{dt} = -\frac{1}{\varepsilon}\nabla_y \Phi(x,y) + \sqrt{\frac{2\sigma}{\varepsilon}}\frac{dV}{dt},$$

   where $U$ and $V$ are standard Brownian motions of appropriate dimensions. Under a Fredholm alternative assumption which you should clearly state, show that the averaged equation for $X$ has the form

$$\frac{dX}{dt} = -\nabla \Psi(X) + \sqrt{2\sigma}\frac{dW}{dt}$$

   where the *Fixman potential* $\Psi$ is given by

$$\exp\Big(-\frac{1}{\sigma}\Psi(x)\Big) = \int_{\mathcal{Y}} \exp\Big(-\frac{1}{\sigma}\Phi(x,y)\Big)dy.$$

   Here $W$ is Brownian motion of appropriate dimension. (In fact strong convergence techniques, such as those highlighted in Chapter 17, may be use used to show that $X \approx x$ strongly for $W = U$.).
3. Let $\Phi$ be as in the previous question. Write the following second order system as a system of coupled first order SDEs:

$$\frac{d^2x}{dt^2} + \frac{dx}{dt} = -\nabla_x \Phi(x,y) + \sqrt{2\sigma}\frac{dU}{dt},$$
$$\varepsilon\frac{d^2y}{dt^2} + \frac{dy}{dt} = -\frac{1}{\varepsilon}\nabla_y \Phi(x,y) + \sqrt{\frac{2\sigma}{\varepsilon}}\frac{dV}{dt}.$$

   Find the stationary distribution of the fast process $y$ explicitly. Find the averaged equation for $X$, using the previous question to guide you.
4. Derive the averaged equation from the example in Subsection 10.6.1 by use of formula (10.5.3) from Result 10.3.
5. Let $u$ be a continuous time Markov chain with generator

$$L = \begin{pmatrix} -a & a \\ b & -b \end{pmatrix}.$$

   Without loss of generality label the state-space $\mathcal{I} = \{-1, +1\}$. Define two functions $\omega : \mathcal{I} \to (0,\infty)$ and $m : \mathcal{I} \to (-\infty,\infty)$ by $\omega(\pm 1) = \omega^{\pm}$ and $m(\pm 1) = m^{\pm}$. Now consider the stochastic differential equations, with coefficients depending upon $u$, given by

$$\frac{dx}{dt} = f(x,y) + \sqrt{2\sigma}\frac{dU}{dt},$$
$$\frac{dy}{dt} = -\frac{1}{\varepsilon}\omega(u)(y - m(u)) + \sqrt{\frac{2\sigma}{\varepsilon}}\frac{dV}{dt},$$

with $U$ and $V$ standard Brownian motions of appropriate dimensions. Write the generator for the process $(x, y, u)$ and use multiscale analysis to derive the averaged coupled Markov chain and SDE of the form

$$\frac{dX}{dt} = F(X,u) + \sqrt{2\sigma}\frac{dW}{dt}$$

where $W$ is a standard Brownian motion with the same dimension as $U$.

6. Generalize the previous exercise to the case where the transition rates of the Markov chain, determined by $a$ and $b$, depend upon $x$ and $y$.
7. Find a representation for the effective coefficient matrix $A(x)$ in Remark 10.2, using time-averaging.

# 11

# Homogenization for ODEs and SDEs

## 11.1 Introduction

In this chapter we continue our study of systems of SDEs with two, widely separated, characteristic time scales. The setting is similar to the one considered in the previous chapter. The difference is that in this chapter we seek to derive an effective equation describing dynamics on the longer, *diffusive timescale* $\mathcal{O}(1/\varepsilon^2)$. This is the timescale of interest when the effective drift $F(x)$ defined in equation (10.3.1) vanishes due, for example, to the symmetries of the problem. The vanishing of the effective drift is captured in the centering condition, equation (11.2.5) below. In contrast to the case considered in the previous chapter, in the diffusive timescale the effective equation is stochastic, even when noise does not act directly on the slow variables, that is, even when $\alpha(x, y) \equiv 0$ in equation (11.2.1) below.

In Section 11.2 we present the SDEs that we will analyze in this chapter. Section 11.3 contains the simplified equations which we derive in Section 11.4. In Section 11.5 we describe various properties of the simplified equations. The derivation assumes that the fast process to be eliminated is stochastic. In Section 11.6 we show how the deterministic case can be handled. In Section 11.7 we present various applications of the theory developed in this chapter: the case where the fast process is of Ornstein–Uhlenbeck type is in Section 11.7.1 and the case where the fast process is a chaotic deterministic process is in Section 11.7.2. Deriving the Stratonovich stochastic integral as the limit of smooth approximations to white noise is considered in Section 11.7.3; Stokes' law is studied in Section 11.7.4. The Green–Kubo formula from statistical mechanics is derived in Section 11.7.5. The case where the stochastic integral in the limiting equation can be interpreted in neither the Itô nor the Stratonovich sense in considered in Section 11.7.6. Lévy area corrections are studied in Section 11.7.7. Various extensions of the results presented in this chapter, together with bibliographical remarks, are presented in Section 11.8.

## 11.2 Full Equations

Consider the SDEs

$$\frac{dx}{dt} = \frac{1}{\varepsilon} f_0(x,y) + f_1(x,y) + \alpha(x,y)\frac{dU}{dt}, \quad x(0) = x_0, \tag{11.2.1a}$$

$$\frac{dy}{dt} = \frac{1}{\varepsilon^2} g(x,y) + \frac{1}{\varepsilon}\beta(x,y)\frac{dV}{dt}, \quad y(0) = y_0. \tag{11.2.1b}$$

Here $U$ and $V$ are indepenent standard Brownian motions. Both the $x$ and $y$ equations contain fast dynamics, but the dynamics in $y$ is an order of magnitude faster than in $x$. As discussed in Sections 4.1 and 6.1 $x \in \mathcal{X}$, $y \in \mathcal{Y}$ and $\mathcal{X} \oplus \mathcal{Y} = \mathcal{Z}$.

For equation (11.2.1) the backward Kolmogorov equation (6.3.4) with $\phi = \phi(x)$[1] is,

$$\frac{\partial v}{\partial t} = \frac{1}{\varepsilon^2}\mathcal{L}_0 v + \frac{1}{\varepsilon}\mathcal{L}_1 v + \mathcal{L}_2 v, \quad \text{for } (x,y,t) \in \mathcal{X}\times\mathcal{Y}\times\mathbb{R}^+, \tag{11.2.2a}$$

$$v = \phi(x), \quad \text{for } (x,y,t) \in \mathcal{X}\times\mathcal{Y}\times\{0\}, \tag{11.2.2b}$$

where

$$\mathcal{L}_0 = g\cdot\nabla_y + \frac{1}{2}B : \nabla_y\nabla_y, \tag{11.2.3a}$$

$$\mathcal{L}_1 = f_0\cdot\nabla_x, \tag{11.2.3b}$$

$$\mathcal{L}_2 = f_1\cdot\nabla_x + \frac{1}{2}A : \nabla_x\nabla_x, \tag{11.2.3c}$$

with

$$A(x,y) := \alpha(x,y)\alpha(x,y)^T,$$
$$B(x,y) := \beta(x,y)\beta(x,y)^T.$$

By using the method of multiple scales we eliminate the $y$ dependence in this Kolmogorov equation, in order to identify a simplified equation for the dynamics of $x$ alone.

In terms of the generator $\mathcal{L}_0$, which is viewe as a differential operator in $y$, in which $x$ appears as a parameter, the natural ergodicity assumption to make for variable elimination is the statement that $\mathcal{L}_0$ has one dimensional null space characterized by

$$\mathcal{L}_0 1(y) = 0, \tag{11.2.4a}$$

$$\mathcal{L}_0^* \rho^\infty(y;x) = 0. \tag{11.2.4b}$$

Here $1(y)$ denotes constants in $y$ and $\rho^\infty(y;x)$ is the density of an ergodic measure $\mu_x(dy) = \rho^\infty(y;x)dy$. We also assume that $f_0(x,y)$ averages to zero under this measure, so that the **centering condition**

[1] For simplicity we will take the initial condition of the backward Kolmogorov equation to be independent of $y$. This is not necessary. See the discussion in Section 11.8

$$\int_{\mathcal{Y}} f_0(x,y)\mu_x(dy) = 0 \quad \forall x \in \mathcal{X} \tag{11.2.5}$$

holds. It can then be shown that the term involving $f_0$ in the $x$ equation will, in the limit $\varepsilon \to 0$, give rise to $\mathcal{O}(1)$ effective drift and noise contributions in an approximate equation for $x$.

As in the previous chapter, in the case where $\mathcal{Y} = \mathbb{T}^d$ the operators $\mathcal{L}_0$ and $\mathcal{L}_0^*$ are equipped with periodic boundary conditions. Then, assuming that $B(x,y)$ is strictly positive definite, uniformly in $(x,y) \in \mathcal{X} \times \mathbb{T}^d$, Theorem 6.16 justifies the statement that the null space of $\mathcal{L}_0^*$ is one-dimensional. In more general situations, such as when $\mathcal{Y} = \mathbb{R}^d$, or $B(x,y)$ is degenerate, similar rigorous justifications are possible, but the functional setting is more complicated, typically employing weighted $L^p$ spaces which characterize the decay of the invariant density at infinity.

When $\mathcal{Y} = \mathbb{T}^d$ and $B(x,y)$ is strictly positive definite, Theorem 7.9 also applies and we have a solvability theory for Poisson equations of the form

$$-\mathcal{L}_0\phi = h. \tag{11.2.6}$$

In particular, the equation has a solution if and only if the right hand side of the above equation is centered with respect to the invariant measure of the fast process $\mu_x(dy)$:

$$\int_{\mathbb{T}^d} h(x,y)\,\mu_x(dy) = 0 \quad \forall\, x \in \mathcal{X}. \tag{11.2.7}$$

When (11.2.7) is satisfied, the solution of (11.2.6) is unique up to a constant in the null space of $\mathcal{L}_0$. We can fix this constant by requiring that

$$\int_{\mathbb{T}^d} \phi(x,y)\mu_x(dy) = 0 \quad \forall\, x \in \mathcal{X}.$$

In more general situations, such as when $\mathcal{Y} = \mathbb{R}^d$, or $B(x,y)$ is degenerate, the question of existence and uniqueness of solutions to the Poisson equation (11.2.6) becomes more complicated; however, analogous results are posible in function space settings which enforce appropriate decay properties at infinity. See the remarks and references to the literature in Section 11.8.

## 11.3 Simplified Equations

We assume that the operator $\mathcal{L}_0$ satisfies the Fredholm alternative, Theorem 2.42, and has one-dimensional null-space characterized by (11.2.4). We define the **cell problem** [2] as follows:

$$-\mathcal{L}_0\Phi(x,y) = f_0(x,y), \qquad \int_{\mathcal{Y}} \Phi(x,y)\rho^\infty(y;x)dy = 0. \tag{11.3.1}$$

[2] The word "cell" here refers to the periodic unit cell which sets the scale for the fast variable, in the case $\mathcal{Y} = \mathbb{T}^d$. The terminology comes from the theory of periodic homogenization for PDEs.

This is viewed as a PDE in $y$, with $x$ a parameter. By the Fredholm alternative, (11.3.1) has a unique solution, since $f_0$ satisfies (11.2.5). We may then define a vector field $F$ by

$$\begin{aligned} F(x) &= \int_{\mathcal{Y}} \big(f_1(x,y) + (\nabla_x \Phi(x,y)) f_0(x,y)\big)\rho^\infty(y;x)dy \\ &= F_1(x) + F_0(x) \end{aligned} \tag{11.3.2}$$

and a diffusion matrix $A(x)$ by

$$A(x)A(x)^T = A_1(x) + \frac{1}{2}\Big(A_0(x) + A_0(x)^T\Big), \tag{11.3.3}$$

where

$$A_0(x) := 2\int_{\mathcal{Y}} f_0(x,y) \otimes \Phi(x,y)\rho^\infty(y;x)dy, \tag{11.3.4}$$

$$A_1(x) := \int_{\mathcal{Y}} A(x,y)\rho^\infty(y;x)dy. \tag{11.3.5}$$

To make sure that $A(x)$ is well defined it is necessary to prove that the sum of $A_1(x)$ and the symmetric part of $A_0(x)$ is positive semidefinite. This is done in Section 11.5.

**Result 11.1.** *For $\varepsilon \ll 1$ and times $t$ upto $\mathcal{O}(1)$, the process $x(t)$, the solution of* (11.2.1)*, is approximated by the process $X(t)$, the solution of*

$$\frac{dX}{dt} = F(X) + A(X)\frac{dW}{dt}, \quad X(0) = x_0. \tag{11.3.6}$$

*Remark 11.2.* Notice that knowledge of $AA^T$ is not sufficient to determine $A$ uniquely. As a result equation (11.3.3) does not determine the limiting SDE (11.3.6) uniquely. This is a consequence of the fact that there may be many SDEs that have the same generator. This in turn relates to the fact that the approximation of the solution to (11.2.1) by the solution to (11.3.6) is only valid in the sense of weak convergence of probability measures; see Chapter 18. □

## 11.4 Derivation

We seek a multiscale expansion for the solution of (11.2.2) with the form

$$v = v_0 + \varepsilon v_1 + \varepsilon^2 v_2 + \cdots. \tag{11.4.1}$$

Here $v_j = v_j(x,y,t)$. Substituting this expansion into (11.2.2) and equating powers of $\varepsilon$ gives a hierarchy of equations, the first three of which are

$$\mathcal{O}(1/\varepsilon^2) \quad -\mathcal{L}_0 v_0 = 0, \tag{11.4.2a}$$

$$\mathcal{O}(1/\varepsilon) \quad -\mathcal{L}_0 v_1 = \mathcal{L}_1 v_0, \tag{11.4.2b}$$

$$\mathcal{O}(1) \quad -\mathcal{L}_0 v_2 = -\frac{\partial v_0}{\partial t} + \mathcal{L}_1 v_1 + \mathcal{L}_2 v_0. \tag{11.4.2c}$$

By (11.2.4) equation (11.4.2a) implies that the first term in the expansion is independent of $y$, $v_0 = v(x,t)$. We proceed now with equation (11.4.2b). The solvability condition is satisfied for this equation since, by assumption (11.2.5), $f_0(x,y)$ is centered with respect to the invariant measure for $\varphi_x^t(\cdot)$ and, from (11.2.3b),

$$\mathcal{L}_1 v_0 = f_0(x,y) \cdot \nabla_x v_0(x,t).$$

Equation (11.4.2b) becomes

$$-\mathcal{L}_0 v_1 = f_0(x,y) \cdot \nabla_x v_0(x,t). \tag{11.4.3}$$

Since $\mathcal{L}_0$ is a differential operator in $y$ alone with $x$ appearing as a parameter, the general solution of (11.4.3) has the form

$$v_1(x,y,t) = \Phi(x,y) \cdot \nabla_x v_0(x,t) + \Phi_1(x,t). \tag{11.4.4}$$

The function $\Phi_1$ plays no role in what follows and thus we set it to zero. Thus we represent the solution $v_1$ as a linear operator acting on $v_0$. As our aim is to find a closed equation for $v_0$, this form for $v_1$ is a useful representation of the solution. Substituting for $v_1$ in (11.4.3) shows that $\Phi$ solves the cell problem (11.3.1). Condition (11.2.5) ensures that there is a solution to the cell problem and the normalization condition makes it unique. Turning now to equation (11.4.2c) we see that the right hand side takes the form

$$-\Big(\frac{\partial v_0}{\partial t} - \mathcal{L}_2 v_0 - \mathcal{L}_1\big(\Phi \cdot \nabla_x v_0\big)\Big).$$

Hence solvability of (11.4.2c) for each fixed $x$, requires

$$\begin{aligned}\frac{\partial v_0}{\partial t} &= \int_{\mathcal{Y}} \rho^\infty(y;x)\mathcal{L}_2 v_0(x,t)dy + \int_{\mathcal{Y}} \rho^\infty(y;x)\mathcal{L}_1\big(\Phi(x,y)\cdot\nabla_x v_0(x,t)\big)\,dy \\ &= I_1 + I_2. \end{aligned} \tag{11.4.5}$$

We consider the two terms on the right hand side separately. The first is

$$\begin{aligned} I_1 &= \int_{\mathcal{Y}} \rho^\infty(y;x)\big(f_1(x,y)\cdot\nabla_x + \frac{1}{2}A(x,y) : \nabla_x\nabla_x\big)v_0(x,t)dy \\ &= F_1(x)\cdot\nabla_x v_0(x,t) + \frac{1}{2}A_1(x) : \nabla_x\nabla_x v_0(x,t). \end{aligned}$$

Now for the second term $I_2$ note that

$$\mathcal{L}_1(\Phi\cdot\nabla_x v_0) = f_0 \otimes \Phi : \nabla_x\nabla_x v_0 + (\nabla_x \Phi f_0)\cdot\nabla_x v_0.$$

Hence $I_2 = I_3 + I_4$ where

$$I_3 = \int_{\mathcal{Y}} \rho^\infty(y;x)\big(\nabla_x \Phi(x,y) f_0(x,y)\big) \cdot \nabla_x v_0(x,t)\, dy$$

and

$$I_4 = \int_{\mathcal{Y}} \rho^\infty(y;x)\big(f_0(x,y) \otimes \Phi(x,y) : \nabla_x \nabla_x v_0(x,t)\big)\, dy.$$

Thus

$$I_2 = F_0(x) \cdot \nabla_x v_0(x,t) + \frac{1}{2} A_0(x) : \nabla_x \nabla_x v_0(x,t).$$

Combining our simplifications of the right hand side of (11.4.5) we obtain, since by (2.2.2) only the symmetric part of $A_0$ is required to calculate the Frobenius inner product with another symmetric matrix, the following expression:

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

This is the backward equation corresponding to the reduced dynamics given in (11.3.6).

## 11.5 Properties of the Simplified Equations

The effective SDE (11.3.6) is only well defined if $A(x)A(x)^T$ given by (11.3.3), (11.3.5) is nonnegative definite. We now prove that this is indeed the case.

**Theorem 11.3.** *Consider the case where $\mathcal{Y} = \mathbb{T}^d$ and $\mathcal{L}_0$ is equipped with periodic boundary conditions. Then*

$$\langle \xi, A_1(x)\xi + A_0(x)\xi \rangle \geqslant 0 \quad \forall x \in \mathcal{X}, \xi \in \mathbb{R}^l.$$

*Hence the real-valued matrix function $A(x)$ is well defined by* (11.3.3) *since $A(x)A(x)^T$ is non-negative definite.*

*Proof.* Let $\phi(x,y) = \xi \cdot \Phi(x,y)$. Then $\phi$ solves

$$-\mathcal{L}_0 \phi = \xi \cdot f_0.$$

By Theorem 6.12 we have

$$\begin{aligned}
&\langle \xi, A_1(x)\xi + A_0(x)\xi \rangle \\
&\qquad = \int_{\mathcal{Y}} \Big( |\alpha(x,y)^T \xi|^2 - 2(\mathcal{L}_0 \phi(x,y))\phi(x,y) \Big) \rho^\infty(y;x) dy \\
&\qquad = \int_{\mathcal{Y}} \Big( |\alpha(x,y)^T \xi|^2 + |\beta(x,y)^T \nabla_y \phi(x,y)|^2 \Big) \rho^\infty(y;x) dy \\
&\qquad \geqslant 0.
\end{aligned}$$

Thus

$$\begin{aligned}\langle \xi, AA^T\xi\rangle &= \langle \xi, A_1\xi\rangle + \frac{1}{2}\langle \xi, (A_0 + A_0^T)\xi\rangle \\ &= \langle \xi, (A_1 + A_0)\xi\rangle \geqslant 0.\square\end{aligned}$$

Two important remarks are in order.

*Remark 11.4.* Techniques similar to those used in the proof of the previous theorem, using (6.3.11) instead of the Dirichlet form itself, show that

$$\frac{1}{2}\Big(A_0(x)+A_0(x)^T\Big) = \int_{\mathcal{Y}}\Big(\nabla_y \Phi(x,y)\beta(x,y)\otimes\nabla_y\Phi(x,y)\beta(x,y)\Big)\rho^\infty(y;x)dy. \quad \square \tag{11.5.1}$$

*Remark 11.5.* By virtue of Remark 6.13 we see that the proceeding theorem can be extended to settings other than $\mathcal{Y} = \mathbb{T}^d$. $\square$

## 11.6 Deterministic Problems

As in the previous chapter, it is useful to have representations of the effective equation in terms of time averages, both for numerical purposes, and for deterministic problems. To this end, a second representation of $A_0(x)$ and $F_0(x)$ is as follows. Let $\varphi_\xi^t(y)$ solve (10.5.1) and let $\mathbb{E}^{\mu_x}$ be the product measure formed from use of $\mu_x(\cdot)$ on initial data and standard independent Wiener measure on driving Brownian motions. Using this notation we may now employ a time integral to represent the solution of the cell problem, leading to the following representation formulae.

**Result 11.6.** *Alternative representations of the vector field $F_0(x)$ and diffusion matrix $A_0(x)$ can be found through the following integrals over time and $\mathbb{E}^{\mu_x}$:*

$$A_0(x) = 2\int_0^\infty \mathbb{E}^{\mu_x}\Big(f_0(x,y)\otimes f_0(x,\varphi_x^t(y))\Big)\,dt \tag{11.6.1}$$

*and, if the generator $\mathcal{L}_0$ is independent of $x$, then*

$$F_0(x) = \int_0^\infty \mathbb{E}^{\mu_x}\Big(\nabla_x f_0(x,\varphi_x^t(y))f_0(x,y)\Big)\,dt. \tag{11.6.2}$$

*All these representations hold for any $y$, by ergodicity.*

The integral over $t$ in this result enables us to express the effective equations without explicit reference to the solution of the cell problem $\Phi$, and requires sufficiently fast *decay of correlations* in order to be well defined.

Another pair of alternative representations of $F(x)$ and $A(x)A(x)^T$ may be found by using time averaging (over $s$) to replace the expectations in the previous result. The expressions for $A_0$ and $F_0$ then involve two time integrals: the integral over

$s$ is an ergodic average, replacing averaging with respect to the stationary measure on path space; the integral over $t$ expresses the effective equations without reference to the solution of the cell problem $\Phi$ and, again, requires sufficiently fast *decay of correlations* in order to be well-defined. In fact the well posedness of the cell problem (11.3.1) implies the decay of correlations property.

**Result 11.7.** *Alternative representations of the vector field $F$ and diffusion matrix $A$ can be found through the following integrals over time:*

$$F_1(x) = \lim_{T\to\infty} \frac{1}{T} \int_0^T f_1(x, \varphi_x^s(y))\, ds,$$
$$A_1(x) = \lim_{T\to\infty} \frac{1}{T} \int_0^T A(x, \varphi_x^s(y))\, ds;$$

*and*

$$A_0(x) = 2\int_0^\infty \Big( \lim_{T\to\infty} \frac{1}{T} \int_0^T f_0(x, \varphi_x^s(y)) \otimes f_0(x, \varphi_x^{t+s}(y)) ds \Big)\, dt, \qquad (11.6.3)$$

*where $\varphi_x^t(y)$ solves* (10.5.1)*. Furthermore, if the generator $\mathcal{L}_0$ is independent of $x$, then*

$$F_0(x) = \int_0^\infty \Big( \lim_{T\to\infty} \frac{1}{T} \int_0^T \nabla_x f_0(x, \varphi_x^{t+s}(y)) f_0(x, \varphi_x^s(y)) ds \Big) dt.$$

*All these representations hold for any $y$, by ergodicity.*

The following result will be useful to us in deriving the alternate representations of $A_0(x)$ and $F_0(x)$ in the two preceding results. It uses ergodicity to represent the solution of the cell problem, and related Poisson equations, as time integrals.

**Result 11.8.** *Let $\mathcal{L}$ be the generator of the ergodic Markov process $y(t)$ on $\mathcal{Y}$ which satisfies the SDE*

$$\frac{dy}{dt} = g(y) + \beta(y)\frac{dV}{dt}, \quad y(t) = y \qquad (11.6.4)$$

*and let $\mu(dy)$ denote the unique invariant measure. Assume that $h$ is centered with respect to $\mu$ :*

$$\int_{\mathcal{Y}} h(y)\mu(dy) = 0.$$

*Then the solution $f(y)$ of the Poisson equation*

$$-\mathcal{L}f = h, \quad \int_{\mathcal{Y}} f(y)\mu(dy) = 0$$

*admits the representation formula*

$$f(y) = \int_0^\infty \big(e^{\mathcal{L}t} h\big)(y)\, dt. \qquad (11.6.5)$$

*Proof.* We apply the Itô formula to $f(y(t))$ to obtain

$$\begin{aligned} f(y(t)) - f(y) &= \int_0^t \mathcal{L} f(y(s))\, ds + \int_0^t \langle \nabla_y f(y(s)), \beta(y(s))\, dW(s) \rangle \\ &= \int_0^t -h(y(s))\, ds + \int_0^t \langle \nabla_y f(y(s)), \beta(y(s))\, dW(s) \rangle. \end{aligned}$$

We take expectation with respect to the Wiener measure and use the martingale property of stochastic integrals, and the fact that $\mathbb{E}h(y(s)|y(0) = y)$ solves the backward Kolmogorov equation, to conclude that

$$f(y) = \mathbb{E} f(y(t)) + \int_0^t \left(e^{\mathcal{L}s} h\right)(y)\, ds.$$

We take the limit $t \to \infty$ and use the ergodicity of the process $y(t)$, together with the fact that $f(y)$ is centered with respect to the invariant measure with density $\rho^\infty(y; x)$, to deduce that

$$\begin{aligned} f(y) &= \lim_{t\to\infty} \mathbb{E} f(y(t)) + \int_0^\infty \left(e^{\mathcal{L}t} h\right)(y)\, dt \\ &= \int_{\mathcal{Y}} f(y)\mu(dy) + \int_0^\infty \left(e^{\mathcal{L}t} h\right)(y)\, dt \\ &= \int_0^\infty \left(e^{\mathcal{L}t} h\right)(y)\, dt \end{aligned}$$

and the proof is complete. □

*Remark 11.9.* Notice that the preceding result implies that we can write, at least formally,

$$\mathcal{L}^{-1} = -\int_0^\infty e^{\mathcal{L}t}\, dt$$

when applied to functions centered with respect to $\mu$. Furthermore, the result is also valid for the case where the coefficients in (11.6.4) depend on a parameter $x$. □

We complete the section by deriving the alternative expressions for $A(x)$ and $F(x)$ through time integration, given in Results 11.7 and 11.6. The expressions for $F_1(x)$ and $A_1(x)$ in Result 11.7 are immediate from ergodicity, simply using the fact that the time average equals the average against $\rho^\infty$. By use of Result 11.8, the solution to the cell problem can be written as

$$\Phi(x, y) = \int_0^\infty \left(e^{\mathcal{L}_0 t} f_0\right)(x, y)\, dt = \int_0^\infty \mathbb{E} f_0(x, \varphi_x^t(y))\, dt \tag{11.6.6}$$

where $\mathbb{E}$ denotes expectation with respect to the Wiener measure. Now

$$F_0(x) = \int_{\mathcal{Y}} \rho^\infty(y; x) \nabla_x \Phi(x, y) f_0(x, y)\, dy.$$

In the case where $\mathcal{L}_0$ is $x-$independent so that $\varphi_x^t(\cdot) = \varphi^t(\cdot)$ is also $x$ independent, as are $\mu_x = \mu$ and $\rho^\infty(\,\cdot\,;x) = \rho^\infty(\cdot)$, we may use (11.6.6) to see that

$$F_0(x) = \int_{\mathcal{Y}} \rho^\infty(y;x) \int_0^\infty \mathbb{E}\nabla_x f_0(x, \varphi^t(y)) f_0(x,y)\, dt\, dy,$$

where $\mathbb{E}$ is expectation with respect to Wiener measure. Recal that $\mathbb{E}^{\mu_x}$ denotes the product measure formed from distributing $y$ in its invariant measure, together with the Brownian motion driving the equation for $\varphi^t(y)$. Changing the order of integration we find that

$$F_0(x) = \int_0^\infty \mathbb{E}^{\mu_x} \Big( \nabla_x f_0(x, \varphi_x^t(y)) f_0(x,y) \Big)\, dt \tag{11.6.7}$$

as required for the expression in Result 11.6. Now we replace averages over $\mathbb{E}^{\mu_x}$ by time averaging to obtain, for all $y$,

$$F_0(x) = \int_0^\infty \Big( \lim_{T\to\infty} \frac{1}{T} \int_0^T \nabla_x f_0(x, \varphi_x^{t+s}(y)) f_0(x, \varphi_x^s(y))\, ds \Big)\, dt,$$

and so we obtain the desired formula for Result 11.7.

A similar calculation to that yielding (11.6.7) gives (11.6.1) for $A_0(x)$ in Result 11.6. Replacing the average against $\mathbb{E}^{\mu_x}$ by time average we arrive at the desired formula for $A_0(x)$ in Result 11.7.

## 11.7 Applications

We give a number of examples illustrating the wide applicability of the ideas in this chapter.

### 11.7.1 Fast Ornstein-Uhlenbeck Noise

Consider the equations

$$\frac{dx}{dt} = \frac{1}{\varepsilon}(1-y^2)x, \tag{11.7.1}$$

$$\frac{dy}{dt} = -\frac{\alpha}{\varepsilon^2} y + \sqrt{\frac{2\alpha}{\varepsilon^2}} \frac{dV}{dt}, \tag{11.7.2}$$

where $V(t)$ is a standard one–dimensional Brownian motion. Here

$$f_0(x,y) = (1-y^2)x \text{ and } f_1(x,y) = 0.$$

Recall that the equation for $y$ is a time-rescaling of the OU process from Example 6.19, with $\lambda = \alpha$. Furthermore, these equations arise from the first application in Section 10.6, in the case where $\lambda = \alpha$, and after time rescaling to produce nonzero effects.

We have that

$$\int_{-\infty}^{\infty}(1-y^2)x\rho^{\infty}(y)\,dy=0,$$

where $\rho^{\infty}(y)$ is the invariant density of the Ornstein–Uhlenbeck process, namely a standard unit normal distribution. Thus the theory put forward in this chapter applies.

The generator of the process $\varphi_\xi^t(\cdot)=\varphi^t(\cdot)$ is

$$\mathcal{L}_0=-\alpha y\frac{\partial}{\partial y}+\alpha\frac{\partial^2}{\partial y^2} \tag{11.7.3}$$

and the cell problem (Poisson equation) (11.3.1) becomes

$$\alpha y\frac{\partial \Phi}{\partial y}-\alpha\frac{\partial^2\Phi}{\partial y^2}=(1-y^2)x.$$

The unique centered solution to this equation is

$$\Phi(y,x)=\frac{1}{2\alpha}(1-y^2)x.$$

Under the standard normal distribution the fourth and second moments take values $3$ and $1$ respectively. Hence, the coefficients in the limiting equation (11.3.6) are

$$F(x)=\int_{-\infty}^{\infty}\left(-\frac{1}{2\alpha}y^2(1-y^2)x\right)\rho^{\infty}(y)\,dy=\frac{1}{\alpha}x$$

and

$$A^2(x)=2\int_{-\infty}^{\infty}\left(-\frac{1}{2\alpha}y^2x(1-y^2)x\right)\rho^{\infty}(y)\,dy=\frac{2}{\alpha}x^2.$$

The homogenized SDE is thus

$$\frac{dX}{dt}=\frac{X}{\alpha}+\sqrt{\frac{2}{\alpha}}X\frac{dW}{dt}. \tag{11.7.4}$$

This is the geometric Brownian motion studied in Example 6.4. The solution is

$$X(t)=X(0)\exp\Big(\sqrt{\frac{2}{\alpha}}W(t)\Big).$$

It neither converges to $0$ nor to $\infty$, but subsequences in time attain both limits. This should be compared with the behaviour found in the first example in Section 10.6 which gives rise to decay (resp. growth) if $\lambda>\alpha$ (resp. $\lambda<\alpha$). Our example corresponds to the case $\lambda=\alpha$ with time rescaled to see nontrivial dynamics. It thus lies between decay and growth. Notice that we could have also taken the function in front of the white noise with a minus sign. See Remark 11.2.

Let us now obtain the coefficients of the homogenized equation by using the alternative representations (11.6.1) and (11.6.2). To this end we need to study the

variable $\varphi^t(y)$ solving (10.5.1). From the calculations presented in Example 6.19 we have that

$$\begin{aligned}\varphi^t(y) &= e^{-\alpha t}y + \sqrt{2\alpha}\int_0^t e^{-\alpha(t-s)}dV(s),\\ \varphi^t(y)^2 &= e^{-2\alpha t}y^2 + \sqrt{2\alpha}ye^{-\alpha t}\int_0^t e^{-\alpha(t-s)}dV(s) + 2\alpha\Big(\int_0^t e^{-\alpha(t-s)}dV(s)\Big)^2.\end{aligned} \tag{11.7.5}$$

In addition, by the Itô isometry,

$$\begin{aligned}\mathbb{E}\Big(\int_0^t e^{-\alpha(t-s)}dV(s)\Big)^2 &= \int_0^t e^{-2\alpha(t-s)}ds,\\ &= \frac{1}{2\alpha}\Big(1-e^{-2\alpha t}\Big).\end{aligned}$$

To construct the measure $\mathbb{E}^{\mu^x}$ we take the initial condition $y$ to be a standard unit Gaussian distribution and an independent driving Brownian motion $V$. (The measure is, in fact, independent of $x$ in this particular example). Thus, by stationarity under this initial Gaussian distribution,

$$\int \rho^\infty(y)y^2\,dy = 1, \quad \mathbb{E}^{\mu^x}\varphi^t(y)^2 = 1.$$

Furthermore

$$\begin{aligned}\mathbb{E}^{\mu^x}\left(\int \rho^\infty(y)y^2\varphi^t(y)^2\,dy\right) &= e^{-2\alpha t}\int \rho^\infty(y)y^4\,dy\\ &\quad +2\alpha\mathbb{E}^{\mu^x}\left(\int_0^t e^{-\alpha(t-s)}\,dV(s)\right)^2\\ &= 3e^{-2\alpha t} + 1 - e^{-2\alpha t}\\ &= 1 + 2e^{-2\alpha t}.\end{aligned}$$

Since $f_0(x,y) = (1-y^2)x$, combining these calculations in (11.6.2) gives

$$\begin{aligned}F_0(x) &= x\int_0^\infty \mathbb{E}^{\mu_x}\Big((1-\varphi^t(y)^2)(1-y^2)\Big)dt\\ &= x\int_0^\infty 2e^{-2\alpha t}dt\\ &= \frac{x}{\alpha}.\end{aligned} \tag{11.7.6}$$

Similarly from (11.6.1) we obtain

$$A_0(x) = \frac{2x^2}{\alpha}.$$

This confirms that the effective equation is (11.7.4).

### 11.7.2 Fast Chaotic Noise

We now consider an example which is entirely deterministic, but which behaves stochastically when we eliminate a fast chaotic variable. In this context it is essential to use the representation of the effective diffusion coefficient given in Result 11.7. This representation uses time-integrals, and makes no reference to averaging over the invariant measure (which does not have a density with respect to Lebesgue measure in this example; see Example 4.16). Consider the equations

$$\begin{aligned}
\frac{dx}{dt} &= x - x^3 + \frac{\lambda}{\varepsilon} y_2, \qquad (11.7.7)\\
\frac{dy_1}{dt} &= \frac{10}{\varepsilon^2}(y_2 - y_1),\\
\frac{dy_2}{dt} &= \frac{1}{\varepsilon^2}(28y_1 - y_2 - y_1 y_3),\\
\frac{dy_3}{dt} &= \frac{1}{\varepsilon^2}(y_1 y_2 - \frac{8}{3} y_3). \qquad (11.7.8)
\end{aligned}$$

The vector $y = (y_1, y_2, y_3)^T$ solves the Lorenz equations, at parameter values where the solution is ergodic (see Example 4.16). In the invariant meausure the component $y_2$ has mean zero. Thus the centering conition holds. The equation for $x$ is a scalar ODE driven by a chaotic signal with characteristic time $\varepsilon^2$. Because $f_0(x, y) \propto y_2$, with invariant measure shown in Figure 4.2, and because $f_1 = (x, y) = f_1(x)$ only, the candidate equation for the approximate dynamics is

$$\frac{dX}{dt} = X - X^3 + \sigma \frac{dW}{dt}, \qquad (11.7.9)$$

where $\sigma$ is a constant. Now let $\psi^t(y) = e_2 \cdot \varphi^t(y)$. Then the constant $\sigma$ can be found by use of (11.6.3) giving

$$\sigma^2 = 2\lambda^2 \int_0^\infty \frac{1}{T} \Big( \lim_{T\to\infty} \int_0^T \psi^s(y)\psi^{t+s}(y) ds \Big) dt.$$

This is the integrated autocorrelation function of $y_2$. By ergodicity we expect the value of $\sigma^2$ to be independent of $y$ and to be determined by the SRB measure for the Lorenz equations alone. Notice that the formula is expected to make sense, even though the cell problem is not well-posed in this case because the generator of the fast process is not elliptic.

Another way to derive this result is as follows. Gaussian white noise $\sigma\dot{W}$, the time-derivative of Brownian motion, may be thought of as a delta-correlated stationary process. The integral of its autocorrelation function on $[0, \infty)$ gives $\sigma^2/2$. On the assumption that $y_2$ has a correlation function which decays in time, and noting that this has timescale $\varepsilon^2$, the autocorrelation of $\frac{\lambda}{\varepsilon}\psi^{s/\varepsilon^2}(y)$ at timelag $t$ may be calculated and integrated from 0 to $\infty$; matching this with the known result for Gaussian white noise gives the desired result for $\sigma^2$.

### 11.7.3 Stratonovich Corrections

When white noise is approximated by a smooth process this often leads to Stratonovich interpretations of stochastic integrals, at least in one dimension. We use multiscale analysis to illustrate this phenomenon by means of a simple example. Consider the equations

$$\begin{aligned} \frac{dx}{dt} &= \frac{1}{\varepsilon} f(x) y, \\ \frac{dy}{dt} &= -\frac{\alpha y}{\varepsilon^2} + \sqrt{\frac{2\alpha}{\varepsilon^2}} \frac{dV}{dt}, \end{aligned} \tag{11.7.10}$$

with $V$ being a standard one-dimensional Brownian motion.

Assume for simplicity that $y(0) = 0$. Then

$$\mathbb{E}(y(t)y(s)) = e^{-\frac{\alpha}{\varepsilon^2}|t-s|}$$

and, consequently,

$$\lim_{\varepsilon \to 0} \mathbb{E}\left( \frac{y(t)}{\varepsilon} \frac{y(s)}{\varepsilon} \right) = \frac{2}{\alpha} \delta(t-s),$$

which implies the heuristic

$$\lim_{\varepsilon \to 0} \frac{y(t)}{\varepsilon} = \sqrt{\frac{2}{\alpha}} \frac{dV}{dt}. \tag{11.7.11}$$

Another way of seeing this is by solving (11.7.10) for $y/\varepsilon$:

$$\frac{y}{\varepsilon} = \sqrt{\frac{2}{\alpha}} \frac{dV}{dt} - \frac{\varepsilon}{\alpha} \frac{dy}{dt}. \tag{11.7.12}$$

If we neglect the $\mathcal{O}(\varepsilon)$ term on the right hand side then we arrive, again, at the heuristic (11.7.11).

Both of these arguments lead us to conjecture a limiting equation of the form

$$\frac{dX}{dt} = \sqrt{\frac{2}{\alpha}} f(X) \frac{dV}{dt}. \tag{11.7.13}$$

We will show that, as applied, *the heuristic gives the incorrect limit:* this is because, in one dimension, whenever white noise is approximated by a smooth process, the limiting equation should be interpreted in the Stratonovich sense, giving

$$\frac{dX}{dt} = \sqrt{\frac{2}{\alpha}} f(X) \circ \frac{dV}{dt}, \tag{11.7.14}$$

in this case. We now derive this limit equation by the techniques introduced in this chapter.

The cell problem is

$$-\mathcal{L}_0\Phi(x,y) = f(x)y$$

with $\mathcal{L}_0$ given by (11.7.3). The solution is readily seen to be

$$\Phi(x,y) = \frac{1}{\alpha}f(x)y, \quad \nabla_x\Phi(x,y) = \frac{1}{\alpha}f'(x)y.$$

The invariant density is

$$\rho^\infty(y) = \frac{1}{\sqrt{2\pi}}\exp\Big(-\frac{y^2}{2}\Big),$$

which is in the null space of $\mathcal{L}_0^*$ and corresponds to a standard unit Gaussian $\mathcal{N}(0,1)$ random variable.

From equation (11.3.2) we have

$$\begin{aligned} F(x) &= \int_{\mathbb{R}} \frac{1}{\alpha}f'(x)f(x)y^2\rho^\infty(y)dy \\ &= \frac{1}{\alpha}f'(x)f(x). \end{aligned}$$

Also (11.3.3) gives

$$\begin{aligned} A(x)^2 &= \int_{\mathbb{R}} \frac{2}{\alpha}f(x)^2y^2\rho^\infty(y)dy \\ &= \frac{2}{\alpha}f(x)^2. \end{aligned}$$

The limiting equation is therefore the Itô SDE

$$\frac{dX}{dt} = \frac{1}{\alpha}f'(X)f(X) + \sqrt{\frac{2}{\alpha}}f(X)\frac{dV}{dt}.$$

This is the Itô form of (11.7.14), by Remark 6.2. Hence the desired result is established.

### 11.7.4 Stokes' Law

The previous example may be viewed as describing the motion of a massless particle with position $x$ in a velocity field proportional to $f(x)y$, with $y$ an OU process. If the particle has mass $m$ then it is natural to study the generalized equation

$$m\frac{d^2x}{dt^2} = \frac{1}{\varepsilon}f(x)y - \frac{dx}{dt}, \tag{11.7.15a}$$

$$\frac{dy}{dt} = -\frac{\alpha y}{\varepsilon^2} + \sqrt{\frac{2\alpha}{\varepsilon^2}}\frac{dV}{dt}. \tag{11.7.15b}$$

(Note that setting $m = 0$ gives the previous example). Equation (11.7.15a) is *Stokes' law*, stating that the force on the particle is proportional to a drag force, $\frac{1}{\varepsilon}f(x)y - \frac{dx}{dt}$,

which is equal to the difference between the fluid velocity and the particle velocity. As in the previous example, $y$ is a fluctuating OU process. For simplicity we consider the case of unit mass, $m = 1$.

Using the heuristic argument from the previous section it is natural to conjecture the limiting equation

$$\frac{d^2X}{dt^2} = \sqrt{\frac{2}{\alpha}} f(X)\frac{dV}{dt} - \frac{dX}{dt}. \tag{11.7.16}$$

In contrast to the previous application, the conjecture that this is the limiting equation turns out to be correct. The reason is that, here, $x$ is smoother and the Itô and Stratonovich integrals coincide; there is no Itô correction to the Stratonovich integral. (To see this it is necessary to first write (11.7.16) as a first order system – see Exercise 2a). We verify the result by using the multiscale techniques introduced in this chapter.

We first write (11.7.15) as the first order system

$$\begin{aligned}
\frac{dx}{dt} &= r, \\
\frac{dr}{dt} &= -r + \frac{1}{\varepsilon} f(x) y, \\
\frac{dy}{dt} &= -\frac{1}{\varepsilon^2}\alpha y + \frac{1}{\varepsilon}\sqrt{2\alpha}\frac{dV}{dt}.
\end{aligned}$$

Here $(x, r)$ are slow variables ($x$ in (11.2.1)) and $y$ the fast variables ($y$ in (11.2.1)). The cell problem is now given by

$$\mathcal{L}_0 \Phi(x, r, y) = -f_0(x, r, y) = \begin{pmatrix} 0 \\ -f(x)y \end{pmatrix},$$

with $\mathcal{L}_0$ given by (11.7.3). The solution is

$$\Phi(x, r, y) = \begin{pmatrix} 0 \\ \frac{1}{\alpha} f(x) y \end{pmatrix}, \quad \nabla_{(x,r)} \Phi(x, y) = \begin{pmatrix} 0 & 0 \\ \frac{1}{\alpha} f'(x) y & 0 \end{pmatrix}.$$

Notice that $f_0$ is in the null space of $\nabla_{(x,r)}\Phi$, and hence (11.3.2) gives

$$F(X, R) = F_1(X, R) = \begin{pmatrix} R \\ -R \end{pmatrix}. \tag{11.7.17}$$

From (11.3.3) we have

$$A(X, R)A(X, R)^T = \int_{\mathbb{R}} 2 \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{\alpha} f(X)^2 y^2 \end{pmatrix} \rho^\infty(y) dy.$$

Recall that $\rho^\infty(y)$ is the density of an $\mathcal{N}(0, 1)$ Gaussian random variable. Evaluating the integral gives

$$A(X, R)A(X, R)^T = \begin{pmatrix} 0 & 0 \\ 0 & \frac{2}{\alpha} f(X)^2 \end{pmatrix}.$$

Hence a natural choice for $A(x)$ is

$$A(X,R) = \begin{pmatrix} 0 \\ \sqrt{\frac{2}{\alpha}} f(X) \end{pmatrix}.$$

Thus from (11.7.17) and (11.7.18) we obtain the limiting equation

$$\begin{aligned} \frac{dX}{dt} &= R, \\ \frac{dR}{dt} &= -R + \sqrt{\frac{2}{\alpha}} f(X) \frac{dW}{dt}, \end{aligned}$$

which, upon elimination of $R$, is seen to coincide with the conjectured limit (11.7.16).

### 11.7.5 Green–Kubo Formula

In the previous application we encountered the equation of motion for a particle with significant mass, subject to Stokes drag. Here we study the same equation of motion, but where the velocity field is steady. We also assume that the particle is subject to molecular diffusion. The equation of motion is thus

$$\frac{d^2x}{dt^2} = f(x) - \frac{dx}{dt} + \sigma \frac{dU}{dt}. \tag{11.7.18}$$

Here $U$ is a standard unit Brownian motion. We will study the effective diffusive behavior of the particle $x$ on large length and timescales, under the assumption that $f(x)$ is a mean zero periodic function. We show that, on appropriate large length and time scales, the particle performs an effective Brownian motion, and we calculate its diffusion coefficient.

To this end we rescale the equation of motion by setting $x \to x/\varepsilon$ and $t \to t/\varepsilon^2$ to obtain

$$\varepsilon^2 \frac{d^2x}{dt^2} = \frac{1}{\varepsilon} f\Big(\frac{x}{\varepsilon}\Big) - \frac{dx}{dt} + \sigma \frac{dU}{dt}.$$

Introducing the variables $y = \varepsilon \frac{dx}{dt}$ and $z = x/\varepsilon$ we obtain the system

$$\begin{aligned} \frac{dx}{dt} &= \frac{1}{\varepsilon} y, \\ \frac{dy}{dt} &= -\frac{1}{\varepsilon^2} y + \frac{1}{\varepsilon^2} f(z) + \frac{\sigma}{\varepsilon} \frac{dW}{dt}, \\ \frac{dz}{dt} &= \frac{1}{\varepsilon^2} y. \end{aligned}$$

The process $(y,z)$ is ergodic, with characteristic timescale $\varepsilon^2$, and plays the role of $y$ in (11.2.1); $x$ plays the role of $x$ in (11.2.1). The operator $\mathcal{L}_0$ is the generator of the process $(y,z)$. Furthermore

$$f_1(x,y,z) = 0, \quad f_0(x,y,z) = y.$$

Thus, since the evolution of $(y, z)$ is independent of $x$, $\Phi(x, y, z)$, the solution of the cell problem, is also $x-$independent. Hence (11.3.2) gives $F(x) = 0$. Turning now to the effective diffusivity we find that, since $\alpha(x, y) = A(x, y) = 0$, (11.3.3) gives $A(x)^2 = A_0(x)$. Now define $\psi^t(y, z)$ to be the component of $\varphi^t(y, z)$ projected onto the $y$ coordinate. By Result 11.7 we have that

$$A_0(x) = 2\int_0^\infty \Big(\lim_{T\to\infty} \frac{1}{T}\int_0^T \psi^s(y)\psi^{s+t}(y)ds\Big)dt.$$

The expression

$$C(t) = \lim_{T\to\infty} \frac{1}{T}\int_0^T \psi^s(y)\psi^{s+t}(y)ds$$

is the *velocity autocorrelation function*. Thus the effective equation is

$$\frac{dX}{dt} = \sqrt{2D}\frac{dW}{dt},$$

a Brownian motion with diffusion coefficient

$$D = \int_0^\infty C(t)dt.$$

Thus, the effective diffusion coefficient is given by the *integrated velocity autocorrelation*. This is an example of the **Green–Kubo formula**.

### 11.7.6 Neither Itô nor Stratonovich

We again use Stokes' law (11.7.15a), now for a particle of small mass $m = \tau_0\varepsilon^2$ where $\tau_0 = \mathcal{O}(1)$, and neglecting molecular diffusion. If we also assume that the velocity field of the underlying fluid is of the form $\frac{1}{\varepsilon}f(x)\eta$ where $\eta$ is solves an SDE, then we obtain

$$\tau_0\varepsilon^2\frac{d^2x}{dt^2} = -\frac{dx}{dt} + \frac{1}{\varepsilon}f(x)\eta, \tag{11.7.19a}$$

$$\frac{d\eta}{dt} = \frac{1}{\varepsilon^2}g(\eta) + \frac{1}{\varepsilon}\sqrt{2\sigma(\eta)}\frac{dW}{dt}. \tag{11.7.19b}$$

We interpret equations (11.7.19b) in the Itô sense. We assume that $g(\eta)$, $\sigma(\eta)$ are such that there exists a unique stationary solution of the Fokker-Planck equation for (11.7.19b), so that $\eta$ is ergodic.

We write (11.7.19) as a first order system,

$$\begin{aligned}
\frac{dx}{dt} &= \frac{1}{\varepsilon\sqrt{\tau_0}}v,\\
\frac{dv}{dt} &= \frac{f(x)\eta}{\varepsilon^2\sqrt{\tau_0}} - \frac{v}{\tau_0\varepsilon^2},\\
\frac{d\eta}{dt} &= \frac{g(\eta)}{\varepsilon^2} + \frac{\sqrt{2\sigma(\eta)}}{\varepsilon}\frac{dW}{dt}.
\end{aligned} \tag{11.7.20}$$

Equations (11.7.20) are of the form (11.2.1) and, under the assumption that the fast process $(v, \eta)$ is ergodic, the theory developed in this chapter applies. In order to calculate the effective coefficients we need to solve the stationary Fokker–Planck equation

$$\mathcal{L}_0^* \rho(x, v, \eta) = 0$$

and the cell problem

$$-\mathcal{L}_0 h = \frac{v}{\sqrt{\tau_0}}, \tag{11.7.21}$$

where

$$\mathcal{L}_0 = g(\eta)\frac{\partial}{\partial \eta} + \sigma(\eta)\frac{\partial^2}{\partial \eta^2} + \left(\frac{f(x)\eta}{\sqrt{\tau_0}} - \frac{v}{\tau_0}\right)\frac{\partial}{\partial v}.$$

Equation (11.7.21) can be simplified considerably: we look for a solution of the form

$$h(x, v, \eta) = \left(\sqrt{\tau_0}\, v + f(x)\widehat{h}(\eta)\right). \tag{11.7.22}$$

Substituting this expression in the cell problem we obtain, after some algebra, the equation

$$-\mathcal{L}_\eta \widehat{h} = \eta.$$

Here $\mathcal{L}_\eta$ denotes the generator of $\eta$. We assume that the unique invariant measure for $\eta(t)$ has density $\rho_\eta(\eta)$ with respect to Lebesgue measure; the centering condition which ensures the well posedness of the Poisson equation for $\widehat{h}$ is

$$\int_{\mathbb{R}} \eta \rho_\eta(\eta)\, d\eta = 0.$$

We assume that this holds. The homogenized SDE is

$$\frac{dX}{dt} = F(X) + \sqrt{D(X)}\frac{dW}{dt}, \tag{11.7.23}$$

where

$$F(x) := \int_{\mathbb{R}^2} \left(\frac{v}{\sqrt{\tau_0}}\widehat{h}(\eta) f'(x)\right) \rho(x, v, \eta)\, dv d\eta$$

and

$$D(x) := 2\int_{\mathbb{R}^2} \left(v^2 + \frac{v}{\sqrt{\tau_0}}\widehat{h}(\eta) f(x)\right) \rho(x, v, \eta)\, dv d\eta.$$

In the case where $\eta(t)$ is the Ornstein–Uhlenbeck process

$$\frac{d\eta}{dt} = -\frac{\alpha}{\varepsilon^2}\eta + \sqrt{\frac{2\lambda}{\varepsilon^2}}\frac{dW}{dt} \tag{11.7.24}$$

we can compute the homogenized coefficients $D(X)$ and $B(X)$ explicitly. The effective SDE is

$$\frac{dX}{dt} = \frac{\lambda}{\alpha^2(1 + \tau_0 \alpha)} f(X) f'(X) + \sqrt{\frac{2\lambda}{\alpha^2}} f(X)\frac{dW}{dt}. \tag{11.7.25}$$

Note that in the limit $\tau_0 \to \infty$ we recover the Itô stochastic integral, as in Subsection 11.7.4, whereas in the limit $\tau_0 \to 0$ we recover the Itô interpretation of the Stratonovich stochastic integral as in Subsection 11.7.3. For $\tau_0 \in (0, \infty)$ the limiting equation is of neither the Itô nor the Stratonovich form. In fact the equation (11.7.25) can be written in the form

$$X(t) = x_0 + \int_0^t \frac{2\lambda}{\alpha^2} f(X) \widehat{\circ} dW(t),$$

where the definition of the stochastic integral through Riemann sums depends on the value of $\tau_0$. The fact that we recover this interesting limit is very much tied to the scaling of the mass as $\mathcal{O}(\varepsilon^2)$. This scaling ensures that the timescale of the ergodic process $\eta$ and the relaxation time of the particle are the same. Resonance between these timescales gives the desired effect.

### 11.7.7 The Lévy Area Correction

[3] In Section 11.7.3 we saw that smooth approximation to white noise in one dimension leads to the Stratonovich stochastic integral. This is not true in general, however, in the multidimensional case: an additional drift can appear in the limit. This extra drift contribution is related to the properties of the Lévy area of the limit process (see the discussion in Section 11.8).

Consider the fast–slow system

$$\dot{x}_1 = \frac{1}{\varepsilon} y_1, \tag{11.7.26a}$$

$$\dot{x}_2 = \frac{1}{\varepsilon} y_2, \tag{11.7.26b}$$

$$\dot{x}_3 = \frac{1}{\varepsilon} \left( x_1 y_2 - x_2 y_1 \right), \tag{11.7.26c}$$

$$\dot{y}_1 = -\frac{1}{\varepsilon^2} y_1 - \alpha \frac{1}{\varepsilon^2} y_2 + \frac{1}{\varepsilon} \dot{W}_1, \tag{11.7.26d}$$

$$\dot{y}_2 = -\frac{1}{\varepsilon^2} y_2 + \alpha \frac{1}{\varepsilon^2} y_1 + \frac{1}{\varepsilon} \dot{W}_2, \tag{11.7.26e}$$

where $\alpha > 0$. Here $W_1, W_2$ are standard independent Brownian motions.

Notice that equations (11.7.26d) and (11.7.26e) may be written in the form

$$\dot{y} = -\frac{1}{\varepsilon^2} y + \frac{1}{\varepsilon^2} \alpha J y + \frac{1}{\varepsilon} \dot{W},$$

where $y = (y_1, y_2)$, $W = (W_1, W_2)$ and $J$ is the antisymmetric (symplectic) matrix

$$J = \begin{pmatrix} 0 & -1 \\ 1 & 0. \end{pmatrix}.$$

[3] This section was written in collaboration with M. Hairer.

Applying the heuristic that

$$y \approx \varepsilon(I - \alpha J)^{-1} \frac{dW}{dt}$$

leads to the conjectured limiting equations

$$\dot{x}_1 = \frac{1}{1+\alpha^2}\left(\dot{W}_1 - \alpha \dot{W}_2\right), \tag{11.7.27a}$$

$$\dot{x}_2 = \frac{1}{1+\alpha^2}\left(\dot{W}_2 + \alpha \dot{W}_1\right), \tag{11.7.27b}$$

$$\dot{x}_3 = \frac{1}{1+\alpha^2}\left((\alpha x_1 - x_2)\dot{W}_1 + (\alpha x_2 + x_1)\dot{W}_2\right). \tag{11.7.27c}$$

We know from Subsections 11.7.3 and 11.7.6 that we must take care in conjecturing such a limit as typically smooth approximations of white noise give rise to the Stratonovich stochastic integral. However in this case Itô and Stratonovich coincide so this issue does not arise. Nonetheless, the conjectured limit equation is wrong.

Multiscale techniques, as described in this chapter, lead to the correct homogenized system:

$$\dot{x}_1 = \frac{1}{1+\alpha^2}\left(\dot{W}_1 - \alpha \dot{W}_2\right), \tag{11.7.28a}$$

$$\dot{x}_2 = \frac{1}{1+\alpha^2}\left(\dot{W}_2 + \alpha \dot{W}_1\right), \tag{11.7.28b}$$

$$\dot{x}_3 = \frac{1}{1+\alpha^2}\left((\alpha x_1 - x_2)\dot{W}_1 + (\alpha x_2 + x_1)\dot{W}_2\right) + \frac{\alpha}{1+\alpha^2}. \tag{11.7.28c}$$

Notice the additional constant drift that appears in equation (11.7.28c). It is the antisymmetric part in the equation for the fast process $y$ which is responsible for the presence of the additional drift in the homogenized equation. In particular, when $\alpha = 0$ the homogenized equation becomes

$$\dot{x}_1 = \dot{W}_1,$$

$$\dot{x}_2 = \dot{W}_2,$$

$$\dot{x}_3 = -x_2\dot{W}_1 + x_1\dot{W}_2$$

which agrees with the original (in general incorrect) conjectured limit (11.7.27).

## 11.8 Discussion and Bibliography

The perturbation approach adopted in this chapter, and more general related ones, is covered in a series of papers by Papanicolaou and coworkers – see [244, 241, 242, 240], building on original work of Khasminkii [165, 166]. See [155, 154, 31, 205, 244, 242, 240, 155, 154] for further material. We adapted the general analysis to the

simple case where $\mathcal{Y} = \mathbb{T}^d$. This may be extended to, for example $\mathbb{R}^d$, by working in the appropriate functional setting; see [249, 250, 251].

The basic perturbation expansion outlined in this chapter can be rigorously justified and weak convergence of $x$ to $X$ proved as $\varepsilon \to 0$; see Kurtz [181] and Chapter 18. The perturbation expansion which underlies the approach is clearly exposed in [241]. See also [117, Ch. 6], [321]. and [291]. Similar problems are analyzed in [271, Ch. 8], by using eigenfunction expansions for the Fokker–Planck operator of the fast process. Projection operator techniques are also often employed in the physics literature as a method for eliminating fast variables. See [117, Ch. 6] and the references therein.

Studying the derivation of effective stochastic models when the original system is an ODE is a subject investigated in some generality in [242]. The specific example in Section 11.7.2 relies on the ergodicity of the Lorenz equations, something establishe in [318, 319]. Use of the integrated autocorrelation function to calculate the effective diffusion coefficient numerically is highlighted in [322]; a different approach to finding the effective diffusion coefficient is described in [125]. The program described in that example is carried out in discrete time by Beck [31] who uses a skew-product structure to facilitate an analysis; the ideas can then be rigorously justified in some cases. A skew-product set-up is also employed in [322] and [125]. A rigorous limit theorem for ODEs driven by a fast mixing system is proved in [225], using the large deviation principle for dynamical systems developed in [224]. In the paper [208] the idea that fast chaotic motion can introduce noise in slow variables is pursued for an interesting physically motivated problem where the fast chaotic behavior arises from the Burgers bath of [204]. Further numerical experiments on the Burgers bath are reported in [209].

Related work can be found in [124] and similar ideas in continuous time are addressed in [155, 154] for differential equations; however, rather than developing a systematic expansion in powers of $\varepsilon$, they find the exact solution of the Fokker-Planck equation, projected into the space $\mathcal{X}$, by use of the Mori-Zwanzig formalism [65], and then make power series expansions in $\varepsilon$ of the resulting problem.

In Section 11.7.5 we derived a formula for the effective diffusion coefficient in terms of the integral of the velocity autocorrelation function, giving the Green–Kubo formula. This calculates a *transport coefficient* via the time integral of an autocorrelation function. The Green–Kubo formula, and other transport coefficients, are studied in many books on statistical mechanics. See, for example, [28, Ch. 11], [269].

Applications of multiscale analysis to climate models, where the atmosphere evolves quickly relative to the slow oceanic variations, are surveyed in Majda et al. [205, 202]. Further applications to the atmospheric sciences may be found in [206, 207]. See also [78]. Stokes' law, equation (11.7.15a) is a phenomenological model for the motion of inertial particles in fluids; see [217]. Models of the form (11.7.15), where the velocity field of the fluid in which the particles are immersed is taken to be a Gaussian Markovian random field, were developed in [288, 289] and analyzed further in [254]. Similar Gaussian models for passive tracers were studied in [55, 56].

The fact that smooth approximations to white noise in one dimension lead, in the limit as we remove the regularization, to Stratonovich stochastic integrals (see Section 11.7.3) is often called the Wong–Zakai theorem after [332]. Whether one should interpret the stochastic integral in the sense of Itô or Stratonovich is usually called the Itô versus Stratonovich problem. In cases where more than one fast timescale is present, as in the example considered in Section 11.7.6, the correct interpretation of the stochastic integral in the limiting SDE depends on the order with which we take the limits. See [109, 280]. As was shown in Section 11.7.6, there are instances where the stochastic integral in the limiting SDE can be interpreted in neither the Itô nor the Stratonovich sense. See [129, 180, 255]. A similar phenomenon for the case where the fast process is a discrete deterministic chaotic map was observed in [124]. An interesting set-up to consider in this context is the Stokes law (11.7.15) in the case where the mass is small:

$$\varepsilon^a \frac{d^2x}{dt^2} = \frac{1}{\varepsilon} f(x)y - \frac{dx}{dt} + \sigma \frac{dU}{dt},$$
$$\frac{dy}{dt} = -\frac{\alpha y}{\varepsilon^2} + \sqrt{\frac{2\alpha}{\varepsilon^2}} \frac{dV}{dt}.$$

Setting $\varepsilon = 0$ in the first equation, and invoking a white noise approximation for $y/\varepsilon$ leads to the conjecture that the limit $X$ of $x$ satisfies a first order SDE. The question then becomes the interpretation of the stochastic integral. In [180] multiscale expansions are used to derive the limiting equation satisfied by $x$ in the cases $a = 1, 2$ and $3$. The case $a = 1$ leads to the Itô equation in the limit, the case $a = 3$ to the Stratonovich equation and $a = 2$ to an intermediate limit between the two.

In higher dimensions smooth approximations to white noise result (in general, and depending of the type of regularization) in an additional drift–apart from the Stratonovich stochastic integral–which is related to the commutator between the row vectors of the diffusion matrix. See [151]. A rigorous framework for understanding examples such as that presented in Section 11.7.7, based on the theory of rough paths, can be found in [198].

In this chapter we have considered equations of the form (11.2.1) where $U$ and $V$ are independent Brownian motions. Frequently applications arise where the noise in the two processes are correlated. We will cover such situations in Chapter 13 where we study homogenization for parabolic PDEs. The structure of the linear equations considered will be general enough to subsume the form of the backward Kolmogorov equation which arises from (11.2.1) when $U$ and $V$ are correlated – in fact they are identical. The main change over the derivation in this chapter is that the operator $\mathcal{L}_1$ has additional terms arising from the correlation in the noises – see Exercises 5 and 1.

When writing the backward Kolmogorov equation for the full system, equation (11.2.2), we assumed that the initial conditions depended only on the slow variable $x$. This assumption simplifies the analysis but it is not necessary. If the initial condition is a function of both $x$ and $y$, then an initial (or boundary) layer appears that has to be resolved. This can be achieved by adding appropriate terms in the two–scale

expansion which decay exponentially fast in time. This is done in [336] for continuous time Markov chains and in [167] for SDEs. In this case the initial conditions for the homogenized SDE are obtained by averaging the initial conditions of the original SDE with respect to the invariant measure of the fast process.

In this chapter we have studied homogenization for finite dimensional stochastic systems. Similar results can be proved for infinite dimensional stochastic systems, SPDEs. See [40] for an application of the techniques developed in this chapter to the stochastic Burgers equation.

The use of the representations in Result 11.1 is discussed in [241]. The representations in Results 11.7 and 11.6 for the effective drift and diffusion can be used in the design of coarse time-stepping algorithms – see [322]. In general the presence of two widely separated characteristic timescales in the SDEs (11.2.1) renders their numerical solution a formidable task. New numerical methods have been developed which aim at the efficient numerical solution of such problems. In the context of averaging for Hamiltonian systems the subject is described in [116]; the subject is revisited, in a more general setting, in [93]. Many of these methods exploit the fact that for $\varepsilon$ sufficiently small the solution of (11.2.1a) can be approximated by the solution of the homogenized equation (11.3.6). The homogenized coefficients are computed through formulae of the form (11.6.3) or (11.6.1), integrating equation (11.2.1b) over short time intervals; see [322, 81, 84, 123]. An ambitious program to numerically compute a subset of variables from a (possibly stochastic) dynamical system is outlined in [162]; this approach does not use scale-separation explicitly and finds application in a range of different problems; see [163, 164, 149, 30, 190, 278, 334]. Numerical methods for multiscale problems are overviewed in [83]. For work on parameter estimation for multiscale SDEs see [258]. For other (partly computational) work on dimension reduction in stochastic systems see [59, 148, 273].

## 11.9 Exercises

1. Find the homogenized equation for the SDEs

$$\begin{aligned}\frac{dx}{dt} &= \frac{1}{\varepsilon}f_0(x,y) + f_1(x,y) + \alpha_0(x,y)\frac{dU}{dt} + \alpha_1(x,y)\frac{dV}{dt}, \quad x(0) = x_0,\\ \frac{dy}{dt} &= \frac{1}{\varepsilon^2}g(x,y) + \frac{1}{\varepsilon}g_1(x,y) + \frac{1}{\varepsilon}\beta(x,y)\frac{dV}{dt}, \quad y(0) = y_0,\end{aligned}$$

   assuming that $f_0$ satisfies the centering condition and that $U$ and $V$ are independent Brownian motions.

2. a. Let $\mathcal{Y}$ denote either $\mathbb{T}^d$ or $\mathbb{R}^d$. What is the generator $\mathcal{L}$ for the process $y \in \mathcal{Y}$ given by

$$\frac{dy}{dt} = g(y) + \frac{dV}{dt}?$$

   In the case where $g(y) = -\nabla\Psi(y)$ find a function in the null space of $\mathcal{L}^*$.

b. Find the homogenized SDE arising from the system

$$\frac{dx}{dt} = \frac{1}{\varepsilon} f(x,y),$$
$$\frac{dy}{dt} = \frac{1}{\varepsilon^2} g(y) + \frac{1}{\varepsilon}\frac{dV}{dt},$$

in the case where $g = -\nabla\Psi(y)$.

c. Define the cell problem, giving appropriate conditions to make the solution unique in the case $\mathcal{Y} = \mathbb{T}^d$. State clearly any assumptions on $f$ that are required in the preceding derivation.

3. Use the Itô formula to derive the solution to the SDE (11.7.4). Convert this SDE into Stratonovich form. What do you observe?

4. a. Let $\mathcal{Y}$ be either $\mathbb{T}^d$ or $\mathbb{R}^d$. Write down the generator $\mathcal{L}_0$ for the process $y \in \mathcal{Y}$ given by:

$$\frac{dy}{dt} = g(y) + \frac{dV}{dt}.$$

In the case where $g$ is divergence free, find a function in the null space of $\mathcal{L}_0^*$.

b. Find the averaged SDE arising from the system

$$\frac{dx}{dt} = f(x,y),$$
$$\frac{dy}{dt} = \frac{1}{\varepsilon} g(y) + \frac{1}{\sqrt{\varepsilon}}\frac{dV}{dt},$$

in the case where $g$ is divergence free.

c. Find the homogenized SDE arising from the system

$$\frac{dx}{dt} = \frac{1}{\varepsilon} f(x,y),$$
$$\frac{dy}{dt} = \frac{1}{\varepsilon^2} g(y) + \frac{1}{\varepsilon}\frac{dV}{dt},$$

in the case where $g$ is divergence-free.

d. Define the cell problem, giving appropriate conditions to make the solution unique in the case $\mathcal{Y} = \mathbb{T}^d$. Clearly state any assumptions on $f$ that are required in the preceding derivation.

5. Consider the equation of motion

$$\frac{dx}{dt} = f(x) + \sigma\frac{dW}{dt},$$

where $f(x)$ is divergence-free and periodic with mean zero. It is of interest to understand how $x$ behaves on large length and timescales. To this end, rescale the equation of motion by setting $x \to x/\varepsilon$ and $t \to t/\varepsilon^2$ and introduce $y = x/\varepsilon$. Write down a pair of coupled SDEs for $x$ and $y$. Use the methods developed in Exercise 1 to enable elimination of $y$ to obtain an effective equation for $x$.

6. Carry out the analysis presented in Section 11.7.6 in arbitrary dimensions. Does the limiting equation have the same structure as in the one dimensional case?
7. Derive equation (11.7.25) from (11.7.23) when $\eta(t)$ is given by (11.7.24).
8. (The Kramers to Smoluchowski limit.) Consider the Langevin equation

$$\varepsilon^2 \frac{d^2x}{dt^2} = b(x) - \frac{dx}{dt} + \sqrt{2\sigma}\frac{dW}{dt}, \tag{11.9.1}$$

   where the particle mass is assumed to be small, $m = \varepsilon^2$.
   a. Write (11.9.1) as a first order system by introducing the variable $y = \varepsilon\dot{x}$.
   b. Use multiscale analysis to show that, when $\varepsilon \ll 1$ the solution of (11.9.1) is well approximated by the solution of the Smoluchowski equation

$$\frac{dX}{dt} = b(X) + \sqrt{2\sigma}\frac{dW}{dt}.$$

   c. Calculate the first correction to the Smoluchowski equation.
9. Write equations (11.7.16) as a first order system and show that the Itô and Stratonovich forms of the equation coincide.

# 12

# Homogenization for Elliptic PDEs

## 12.1 Introduction

In this chapter we use multiscale expansions in order to study the problem of homogenization for second order uniformly elliptic PDEs in divergence form. At a purely formal level the calculations used to derive the homogenized equations are very similar to those used in the previous chapter to study homogenization for SDEs. The primary difference is that there is no time dependence in the linear equations that we study.

In Section 12.2 we present the boundary value problem studied in this chapter. Section 12.3 contains the simplified (homogenized) equations, and their derivation is given in Section 12.4. Section 12.5 studies the structure of the simplified equation, showing that it inherits ellipticity from the original equation. In Section 12.6 we describe two applications of the theory, both explicitly solvable, a one dimensional example, and a two dimensional layered material.

## 12.2 Full Equations

We study uniformly elliptic PDEs in divergence form, with Dirichlet boundary conditions:

$$-\nabla \cdot \left( A^{\varepsilon} \nabla u^{\varepsilon} \right) = f \quad \text{for } x \in \Omega, \tag{12.2.1a}$$

$$u^{\varepsilon} = 0 \quad \text{for } x \in \partial\Omega. \tag{12.2.1b}$$

Here $u^{\varepsilon} = u^{\varepsilon}(x)$ is an unknown scalar field, to be determined, $A^{\varepsilon} = A(x/\varepsilon)$ a given matrix field and $f = f(x)$ a given scalar field. Unlike the problems in the previous four chapters, there are not two different explicit variables $x$ and $y$. We will introduce $y = x/\varepsilon$ to create a setting similar to that in the previous chapters. Our goal is then to derive a homogenized equation in which $y$ is eliminated, in the limit $\varepsilon \to 0$. Furthermore, we study various properties of the homogenized coefficients.

We take $\Omega \subset \mathbb{R}^d$, open, bounded with smooth boundary. We will assume that the matrix–valued function $A(y)$ is smooth, 1–periodic and uniformly positive definite.

This assumption implies that the differential operator that appears on the left hand side of (12.2.1a) is uniformly elliptic (see Chapter 7). Furthermore, we take the function $f(x)$ to be smooth and independent of $\varepsilon$. To summarize, we make the following assumptions:

$$f \in C^\infty(\mathbb{R}^d, \mathbb{R}); \tag{12.2.2a}$$

$$A \in C^\infty_{per}(\mathbb{T}^d, \mathbb{R}^{d\times d}); \tag{12.2.2b}$$

$$\exists \alpha > 0 : \langle \xi, A(y)\xi \rangle \geqslant \alpha |\xi|^2, \ \ \forall\, y \in \mathbb{T}^d \ \forall\, \xi \in \mathbb{R}^d. \tag{12.2.2c}$$

Notice that our assumptions on $A$ imply that $A^\varepsilon \in M(\alpha, \beta, \Omega)$ for some appropriate $\beta$ and $\alpha$ independent of $\varepsilon$. The regularity assumptions are more stringent than is necessary; we make them at this point in order to carry out the formal calculations that follow. Allowing minimal regularity assumptions is an important issue, however: in many applications one expects that the coefficient $A(y)$ will have jumps when passing from one material phase to the other. Our proofs of homogenization theorems in Chapter 19 will weaken the regularity assumptions that we make here.

Let $\mathcal{A}_0 = -\nabla_y \cdot (A\nabla_y)$ equipped with periodic boundary conditions on the unit torus and with $A = A(y)$. This operator will play a central role in the following. It was studied in Example 7.12: there it was shown that it has a one dimensional null space, comprising constants in $y$. Furthermore, use of the Fredholm Alternative, shows that the Poisson equation

$$\mathcal{A}_0 v = h, \quad v \text{ is 1–periodic}, \tag{12.2.3}$$

has a solution if and only if

$$\int_{\mathbb{T}^d} h(y)dy = 0. \tag{12.2.4}$$

The solution is unique up to an additive constant. Among all solutions of (12.2.3) which satisfy the solvability condition we will choose the unique solution whose integral over $\mathbb{T}^d$ vanishes:

$$\mathcal{A}_0 v = h, \quad v \text{ is 1–periodic}, \int_{\mathbb{T}^d} v(y)\, dy = 0.$$

Equations of the form (12.2.3) will play a central role in what follows.

## 12.3 Simplified Equations

Define the **effective diffusion tensor** by the formula

$$\overline{A} = \int_{\mathbb{T}^d} \Big( A(y) + A(y)\nabla\chi(y)^T \Big)\, dy \tag{12.3.1}$$

where the vector field $\chi : \mathbb{T}^d \to \mathbb{R}^d$ satisfies the **cell problem**

$$-\nabla_y \cdot \big(\nabla_y \chi A^T\big) = \nabla_y \cdot A^T, \ \ \chi \text{ is 1–periodic}. \tag{12.3.2}$$

**Result 12.1.** *For $0 < \varepsilon \ll 1$ the solution $u^\varepsilon$ of equation* (12.2.1) *is approximately given by the solution $u$ of the homogenized equation*

$$-\nabla \cdot \left(\overline{A}\nabla u\right) = f \quad \textit{for } x \in \Omega, \tag{12.3.3a}$$

$$u = 0 \quad \textit{for } x \in \partial\Omega. \tag{12.3.3b}$$

Notice that the field $\chi$ is determined up to a constant vector. However, since only $\nabla_y \chi$ enters into the formula for the homogenized matrix $\overline{A}$ appearing in the homogenized equation, the value of this constant is irrelevant. For definiteness, however, we work with the unique solution $\chi$ found by imposing the normalization

$$\int_{\mathbb{T}^d} \chi(y) dy = 0. \tag{12.3.4}$$

The cell problem can be written in an alternative, sometimes useful, form by writing an equation for each component of $\chi$ :

$$\mathcal{A}_0 \chi_\ell = \nabla_y \cdot a_\ell, \quad \ell = 1, \ldots, d\,, \tag{12.3.5}$$

where $a_\ell = Ae_\ell, \ell = 1, \ldots, d$ and $\{e_\ell\}_{\ell=1}^d$ is the standard basis on $\mathbb{R}^d$. Thus $a_\ell$ is the $\ell^{th}$ column of $A$.

*Remark 12.2.* Since the Hessian $\nabla_x \nabla_x u$ is symmetric, it follows from property (2.2.2) applied to (12.3.1) that the following expression for $\overline{A}$ is equally valid:

$$\overline{A} = \int_{\mathbb{T}^d} \left(A(y)^T + \nabla_y \chi(y) A(y)^T\right)\, dy.$$

Indeed this expression and (12.3.1) may be combined (for example averaged) to obtain other equally valid expressions for $\overline{A}$ (for example symmetric). □

## 12.4 Derivation

Since a small parameter $\varepsilon$ appears in equation (12.2.1), it is natural to look for a solution in the form of a power series expansion in $\varepsilon$:

$$u^\varepsilon = u_0 + \varepsilon u_1 + \varepsilon^2 u_2 + \ldots.$$

The basic idea behind the method of multiple scales is to assume that all terms in the above expansion depend explicitly on *both* $x$ and $y = \frac{x}{\varepsilon}$. Furthermore, since the coefficients of our PDE are periodic functions of $\frac{x}{\varepsilon}$ it is reasonable to require that all terms in the expansion are periodic functions of $\frac{x}{\varepsilon}$. Hence, we assume the following ansatz for the solution $u^\varepsilon$:

$$u^\varepsilon(x) = u_0\left(x, \frac{x}{\varepsilon}\right) + \varepsilon\, u_1\left(x, \frac{x}{\varepsilon}\right) + \varepsilon^2\, u_2\left(x, \frac{x}{\varepsilon}\right) + \ldots, \tag{12.4.1}$$

where $u_j(x,y)$, $j=0,1,\ldots$, are periodic in $y$.

The variables $x$ and $y=\frac{x}{\varepsilon}$ represent the "slow" (macroscopic) and "fast" (microscopic) scales of the problem, respectively. For $\varepsilon \ll 1$ the variable $y$ changes much more rapidly than $x$ and we can think of $x$ as being a constant, when looking at the problem at the microscopic scale. This is where scale separation is exploited: *we will treat $x$ and $y$ as independent variables*. Justifying the validity of this assumption as $\varepsilon \to 0$ is one of the main issues in the rigorous theory of homogenization. See Chapter 19.

The fact that $y=\frac{x}{\varepsilon}$ implies that the partial derivatives with respect to $x$ become

$$\nabla \to \nabla_x + \frac{1}{\varepsilon}\nabla_y.$$

In other words, the *total derivative* of a function $g^\varepsilon(x) := g\left(x, \frac{x}{\varepsilon}\right)$ can be expressed as

$$\nabla g^\varepsilon(x) = \nabla_x g(x,y)\Big|_{y=\frac{x}{\varepsilon}} + \frac{1}{\varepsilon}\nabla_y g(x,y)\Big|_{y=\frac{x}{\varepsilon}},$$

where the notation $h(x,y)|_{y=z}$ means that the function $h(x,y)$ is evaluated at $y=z$.

We use the above to rewrite the differential operator

$$\mathcal{A}^\varepsilon := -\nabla\cdot(A(y)\nabla)$$

in the form

$$\mathcal{A}^\varepsilon = \frac{1}{\varepsilon^2}\mathcal{A}_0 + \frac{1}{\varepsilon}\mathcal{A}_1 + \mathcal{A}_2, \tag{12.4.2}$$

where

$$\mathcal{A}_0 := -\nabla_y\cdot(A(y)\nabla_y), \tag{12.4.3a}$$
$$\mathcal{A}_1 := -\nabla_y\cdot(A(y)\nabla_x) - \nabla_x\cdot(A(y)\nabla_y), \tag{12.4.3b}$$
$$\mathcal{A}_2 := -\nabla_x\cdot(A(y)\nabla_x). \tag{12.4.3c}$$

Notice that the coefficients in all the operators defined above are periodic functions of $y$. We equip $\mathcal{A}_0$ with periodic boundary conditions on $\mathbb{T}^d$.

Equation (12.2.1) becomes, on account of (12.4.2),

$$\left(\frac{1}{\varepsilon^2}\mathcal{A}_0 + \frac{1}{\varepsilon}\mathcal{A}_1 + \mathcal{A}_2\right)u^\varepsilon = f \quad \text{for } (x,y)\in\Omega\times\mathbb{T}^d, \tag{12.4.4a}$$

$$u^\varepsilon = 0 \quad \text{for } (x,y)\in\partial\Omega\times\mathbb{T}^d. \tag{12.4.4b}$$

We substitute (12.4.1) into (12.4.4) to deduce:

$$\frac{1}{\varepsilon^2}\mathcal{A}_0u_0 + \frac{1}{\varepsilon}\left(\mathcal{A}_0u_1 + \mathcal{A}_1u_0\right) + \left(\mathcal{A}_0u_2 + \mathcal{A}_1u_1 + \mathcal{A}_2u_0\right) + \mathcal{O}(\varepsilon) = f. \tag{12.4.5}$$

We equate coefficients of equal powers of $\varepsilon$ to zero in the above equation and disregard all terms of order higher than $1$ to obtain the following sequence of problems:

$$\mathcal{O}(1/\varepsilon^2) \quad \mathcal{A}_0 u_0 = 0, \tag{12.4.6a}$$

$$\mathcal{O}(1/\varepsilon) \quad \mathcal{A}_0 u_1 = -\mathcal{A}_1 u_0, \tag{12.4.6b}$$

$$\mathcal{O}(1) \quad \mathcal{A}_0 u_2 = -\mathcal{A}_1 u_1 - \mathcal{A}_2 u_0 + f. \tag{12.4.6c}$$

Here $u_j(x,y)$ are $1-$periodic in their second argument.

Notice that $\mathcal{A}_0$ is a differential operator in $y$ and that $x$ appears in equations (12.4.6b) and (12.4.6c) merely as a parameter. From (12.4.6a) we deduce that $u_0(x,y) = u(x)$ – thus the first term in the multiscale expansion is independent of $y$. The remaining two equations are of the form (12.2.3) with $v = v(x,y)$ and similarly $h = h(x,y)$; thus $x$ enters as a parameter.

Let us proceed now with (12.4.6b) which becomes

$$\mathcal{A}_0 u_1 = \left(\nabla_y \cdot A^T\right) \cdot \nabla_x u, \quad u_1(x,\cdot) \text{ is 1–periodic, } \int_{\mathbb{T}^d} u_1 \, dy = 0. \tag{12.4.7}$$

The solvability condition (12.2.4) is satisfied because

$$\begin{aligned} \int_{\mathbb{T}^d} \left(\nabla_y \cdot A^T\right) \cdot \nabla_x u \, dy &= \nabla_x u \cdot \int_{\mathbb{T}^d} \nabla_y \cdot A^T \, dy \\ &= 0, \end{aligned}$$

by the divergence theorem and periodicity of $A(\cdot)$; see Remark 7.13. We seek a solution of (12.4.7) using separation of variables:

$$u_1(x,y) = \chi(y) \cdot \nabla_x u(x). \tag{12.4.8}$$

Upon substituting (12.4.8) into (12.4.7) we obtain the cell problem (12.3.2) for the vector field $\chi(y)$. The field $\chi(y)$ is called the **first order corrector**. Notice that the periodicity of the coefficients implies that the right hand side of equation (12.3.2) averages to zero over the unit cell and consequently the cell problem is well posed. We ensure the uniqueness of solutions to (12.3.2) by requiring the corrector field to have zero average – condition (12.3.4).

Now we consider equation (12.4.6c). By (12.2.4) we see that, in order for this equation to be well posed, it is necessary and sufficient for the right hand side to average to zero over $\mathbb{T}^d$. Since we have assumed that the function $f(x)$ is independent of $y$ the solvability condition implies:

$$\int_{\mathbb{T}^d} \left(\mathcal{A}_2 u_0 + \mathcal{A}_1 u_1\right) \, dy = f. \tag{12.4.9}$$

The first term on the left hand side of the above equation is

$$\begin{aligned} \int_{\mathbb{T}^d} \mathcal{A}_2 u_0 \, dy &= \int_{\mathbb{T}^d} -\nabla_x \cdot \left(A(y) \nabla_x u\right) \, dy \\ &= -\nabla_x \cdot \left[ \left( \int_{\mathbb{T}^d} A(y) \, dy \right) \nabla_x u(x) \right] \\ &= -\left( \int_{\mathbb{T}^d} A(y) \, dy \right) : \nabla_x \nabla_x u(x). \end{aligned} \tag{12.4.10}$$

Moreover

$$\begin{aligned}\int_{\mathbb{T}^d} \mathcal{A}_1 u_1 \, dy \; &= \; \int_{\mathbb{T}^d} \left(-\nabla_y \cdot (A(y)\nabla_x u_1) - \nabla_x \cdot (A(y)\nabla_y u_1)\right) dy \\ &=: I_1 + I_2.\end{aligned}$$

The first term $I_1 = 0$ by periodicity and Remark 7.13. Now we consider $I_2$:

$$\begin{aligned}I_2 &= \int_{\mathbb{T}^d} -\nabla_x \cdot (A(y)\nabla_y u_1) \; dy \\ &= -\int_{\mathbb{T}^d} A(y) : \nabla_x \nabla_y \left(\chi \cdot \nabla_x u\right) \, dy \\ &= -\left(\int_{\mathbb{T}^d} (A(y)\nabla_y \chi(y)^T) \, dy\right) : \nabla_x \nabla_x u. \qquad (12.4.11)\end{aligned}$$

We substitute (12.4.11) and (12.4.10) in (12.4.9) to obtain the homogenized equation of Result 12.1 where the homogenized coefficient $\overline{A}$ is given by the formula (12.3.1). This completes the derivation.

## 12.5 Properties of the Simplified Equations

In this section we study some basic properties of the effective coefficients. In particular, we show that the matrix of homogenized coefficients $\overline{A}$ is positive definite, which implies that the homogenized differential operator is uniformly elliptic and that, consequently, the homogenized equation is well posed. Furthermore, we show that symmetry is preserved under homogenization: the homogenized matrix is symmetric if $A(y)$ is. We also show that the homogenization process can create anisotropies: even if the matrix $A(y)$ is diagonal, the matrix of homogenized coefficients $\overline{A}$ need not be.

In order to study the matrix of homogenized coefficients it is useful to find an alternative representation for $\overline{A}$. To this end, we introduce the bilinear form

$$a_1(\psi, \phi) = \int_{\mathbb{T}^d} \langle \nabla_y \phi, A(y)\nabla_y \psi \rangle \, dy, \qquad (12.5.1)$$

defined for all functions $\phi, \; \psi \in C^1(\mathbb{T}^d)$. Notice that this is the bilinear from associated with the operator $\mathcal{A}_0$, in the sense that

$$\int_{\mathbb{T}^d} \phi \mathcal{A}_0 \psi \, dy = a_1(\phi, \psi) \quad \forall \, \phi, \psi \in C^1_{per}(\mathbb{T}^d). \qquad (12.5.2)$$

Note that, whenever $A$ is symmetric, so is the bilinear form $a_1(\cdot, \cdot)$. We start by obtaining an alternative, equivalent formulation for the cell problem. The formulation is closely related to the weak formulation of elliptic PDEs in divergence–form introduced in Chapter 7. In the rest of this section we will assume that the solution of

the cell problem is smooth enough to justify the calculations that follow. It will be enough to assume that each component of the corrector field $\chi(y)$ is continuously differentiable and periodic: $\chi_\ell(y) \in C^1_{per}(\mathbb{T}^d),\ \ell = 1, \ldots, d$.

Recall that $e_\ell$ denotes the unit vector with $i^{th}$ entry $\delta_{il}$. Also let $y_\ell$ denote the $\ell^{th}$ component of the vector $y$. Note that $e_\ell = \nabla_y y_\ell$ and recall that $a_\ell = Ae_\ell$, the $\ell^{th}$ column of $A$. Using these two elementary facts we can obtain the following useful lemma.

**Lemma 12.3.** *The cell problem* (12.3.2) *can be written in the form*

$$a_1(\phi, \chi_\ell + y_\ell) = 0 \quad \forall \phi \in C^1_{per}(\mathbb{T}^d),\ \ell = 1, \ldots d. \tag{12.5.3}$$

*Proof.* From (12.3.5) we deduce that

$$\mathcal{A}_0 \chi_\ell = \nabla_y \cdot (Ae_\ell) = \nabla_y \cdot (A \nabla_y y_\ell) = -\mathcal{A}_0 y_\ell.$$

Consequently, the cell problem can be written in the form

$$\mathcal{A}_0(\chi_l + y_l) = 0, \quad l = 1, \ldots, d,$$

with periodic boundary conditions. We multiply the cell problem as formulated above by an arbitrary function $\phi \in C^1_{per}(\mathbb{T}^d)$. Integrating over the unit cell, using Remark 7.13 and equations (12.5.1) and (12.5.2), we obtain (12.5.3). □

Using this lemma we give an alternative representation formula for the homogenized coefficients. The lemma shows that $\overline{A}$ is symmetric, whenever $A(y)$ is.

**Lemma 12.4.** *The effective matrix $\overline{A}$ has components given by*

$$\overline{a}_{ij} = a_1(\chi_j + y_j, \chi_i + y_i), \quad i, j = 1, \ldots, d. \tag{12.5.4}$$

*In particular, symmetry of $A(y)$ implies symmetry of $\overline{A}$.*

*Proof.* Notice first that the previous lemma implies that, since $\chi_i(y) \in C^1_{per}(\mathbb{T}^d)$,

$$a_1(\chi_i, \chi_j + y_j) = 0, \quad \forall\, i, j, = 1, \ldots, d. \tag{12.5.5}$$

We now use formula (12.3.1), together with (12.5.5) to obtain

$$\begin{aligned}
\overline{a}_{ij} &= e_i \cdot \overline{A} e_j \\
&= \int_{\mathcal{Y}} \Big( e_i \cdot Ae_j + e_i \cdot A \nabla_y \chi^T e_j \Big) dy \\
&= \int_{\mathcal{Y}} \Big( \nabla_y y_i \cdot A \nabla_y y_j + \nabla_y y_i \cdot A \nabla_y \chi_j \Big) dy \\
&= \int_{\mathcal{Y}} \Big\langle \nabla_y y_i, A\Big( \nabla_y (y_j + \chi_j) \Big) \Big\rangle \, dy \\
&= a_1(y_i, \chi_j + y_j) \\
&= a_1(y_i + \chi_i, \chi_j + y_j).
\end{aligned}$$

This proves (12.5.4). Assume now that $A(y) = A(y)^T$. This implies that the bilinear form $a_1(\cdot,\cdot)$ is symmetric. Thus

$$\begin{aligned}\overline{a}_{ij} &= a_1(y_i + \chi_i, \chi_j + y_j)\\ &= a_1(y_j + \chi_j, \chi_i + y_i)\\ &= \overline{a}_{ji},\end{aligned}$$

which shows that the homogenized matrix is symmetric. □

We now show that the homogenized matrix $\overline{A}$ is positive definite. This implies that the homogenized equation is a well posed elliptic PDE.

**Theorem 12.5.** *The matrix of homogenized coefficients $\overline{A}$ is positive definite.*

*Proof.* Let $\xi \in \mathbb{R}^d$ be an arbitrary vector. We need to show that there exists a constant $\overline{\alpha} > 0$ such that

$$\langle \xi, \overline{A}\xi \rangle \geqslant \overline{\alpha}|\xi|^2, \quad \forall \xi \in \mathbb{R}^d.$$

We use the representation formula (12.5.4) to deduce that:

$$\langle \xi, \overline{A}\xi \rangle = a_1(w, w),$$

with $w = \xi \cdot (\chi + y)$. We now use the uniform positive definiteness of $A(y)$ to obtain

$$a_1(w, w) \geqslant \alpha \int_{\mathbb{T}^d} |\nabla_y w|^2 \, dy \geqslant 0.$$

Thus $\overline{A}$ is nonnegative.

To show that it is actually positive definite we argue as follows. Let us assume that

$$\langle \xi, \overline{A}\xi \rangle = 0$$

for some $\xi$. Then, since $\alpha > 0$, $\nabla_y w = 0$ and $w = c$, a constant vector; consequently

$$\xi \cdot y = c - \xi \cdot \chi.$$

The right hand side of this equation is 1–periodic and continuous in $y$ and consequently the left hand side should also be. The only way this can happen is if $\xi = 0$. This completes the proof of the lemma. □

The above theorem shows that uniform ellipticity is a property that is preserved under the homogenization procedure. In particular, this implies that the homogenized equation is well posed, since it is a uniformly elliptic PDE with constant coefficients.

*Remark 12.6.* Note that homogenization does not preserve isotropy. In particular, even if the diffusion matrix $A$ has only diagonal non–zero elements, the homogenized diffusion matrix $\overline{A}$ will in general have non–zero off–diagonal elements. To see this, let us assume that $a_{ij} = 0$, $i \neq j$. Then the off–diagonal elements of the homogenized diffusion matrix are given by the formula (no summation convention here)

$$\overline{a}_{ij} = \int_{\mathbb{T}^d} a_{ii} \frac{\partial \chi_j}{\partial y_i}\, dy,\ i \neq j.$$

This expression is not necessarily equal to zero and leads to the surprising result that an isotropic composite material can behave, in the limit as the microstructure becomes finer and finer, like an anisotropic homogeneous material. □

## 12.6 Applications

We present two useful illustrative examples, for which explicit solutions may be found. Essentially, the one–dimensional case is the only general setting in which the cell problem can be solved analytically and an explicit formula for the effective diffusivity can be obtained. In higher dimensions, explicit formulae for the effective diffusivities can be obtained only when the specific structure of the problem under investigation enables us to reduce the calculation of the homogenized coefficients to consideration of one dimensional problems. Such a reduction is possible in the case of layered materials, the second example that we consider.

### 12.6.1 The One–Dimensional Case

Let $d = 1$ and take $\Omega = [0, L]$. Then the Dirichlet problem (12.2.1a) reduces to a two–point boundary value problem:

$$-\frac{d}{dx}\left(a\left(\frac{x}{\varepsilon}\right)\frac{du^\varepsilon}{dx}\right) = f \quad \text{for } x \in (0, L), \tag{12.6.1a}$$

$$u^\varepsilon(0) = u^\varepsilon(L) = 0. \tag{12.6.1b}$$

We assume that $a(y)$ is smooth, periodic with period 1. We also assume that there exist constants $0 < \alpha \leqslant \beta < \infty$ such that

$$\alpha \leqslant a(y) \leqslant \beta, \quad \forall y \in [0, 1]. \tag{12.6.2}$$

We also assume that $f$ is smooth.

The cell problem becomes a boundary value problem for an ordinary differential equation with periodic boundary conditions.

$$-\frac{d}{dy}\left(a(y)\frac{d\chi}{dy}\right) = \frac{da(y)}{dy}, \quad \text{for } y \in (0, 1), \tag{12.6.3a}$$

$$\chi \text{ is 1–periodic}, \qquad \int_0^1 \chi(y)\, dy = 0. \tag{12.6.3b}$$

Since $d = 1$ we only have one effective coefficient which is given by the one dimensional version of (12.3.1), namely

$$\begin{aligned}\overline{a} &= \int_0^1 \left( a(y) + a(y)\frac{d\chi(y)}{dy} \right) dy \\ &= \left\langle a(y) \left( 1 + \frac{d\chi(y)}{dy} \right) \right\rangle. \end{aligned} \tag{12.6.4}$$

Here, and in the remainder of this chapter, we employ the notation

$$\langle f(y) \rangle := \int_{\mathbb{T}^d} f(y)\, dy,$$

for the average over $\mathbb{T}^d$.

Equation (12.6.3a) can be solved exactly. Integration from $0$ to $y$ gives

$$a(y)\frac{d\chi}{dy} = -a(y) + c_1. \tag{12.6.5}$$

The constant $c_1$ is undetermined at this point. The inequality (12.6.2) allows us to divide (12.6.5) by $a(y)$ since it implies that $a$ is strictly positive. We then integrate once again from $0$ to $y$ to deduce:

$$\chi(y) = -y + c_1 \int_0^y \frac{1}{a(y)}\, dy + c_2.$$

In order to determine the constant $c_1$ we use the fact that $\chi(y)$ is a periodic function. Thus $\chi(0) = \chi(1)$ and we deduce that

$$c_1 = \frac{1}{\int_0^1 \frac{1}{a(y)}\, dy} = \langle a(y)^{-1} \rangle^{-1}.$$

Thus, from (12.6.5),

$$1 + \frac{d\chi}{dy} = \frac{1}{\langle a(y)^{-1} \rangle a(y)}.$$

(Notice that $c_2$ is not required for the calculation of $\overline{a}$.) We substitute this expression into equation (12.6.4) to obtain

$$\overline{a} = \langle a(y)^{-1} \rangle^{-1}. \tag{12.6.6}$$

This is the formula which gives the homogenized coefficient in one dimension. It shows clearly that, even in this simple one–dimensional setting, the homogenized coefficient is not found by simply averaging the unhomogenized coefficients over a period of the microstructure. Rather, the homogenized coefficient is the inverse of the average of the inverse of the unhomogenized coefficient – the *harmonic average*. It is quite easy to show that the homogenized coefficient which is given by the harmonic average (12.6.6) is bounded from above by the average of $a(y)$. See Exercise 12.

### 12.6.2 Layered Materials

We consider problem (12.2.1), with assumptions (12.2.2) satisfied, in two dimensions. We assume that the domain $\Omega \subset \mathbb{R}^2$ represents a *layered material*: the properties of the material change only in one direction. Hence, the coefficients $A(y)$ are functions of one variable: for $y = (y_1, y_2)^T$ we have

$$a_{ij} = a_{ij}(y_1), \quad i, j = 1, 2. \tag{12.6.7}$$

The fact that the coefficients are functions of $y_1$ implies the right hand side of the cell problem (12.3.2) is a function of $y_1$ alone. As a consequence the solution of the cell problem is also a function of $y_1$ alone and takes the form

$$\chi_\ell = \chi_\ell(y_1), \quad \ell = 1, 2. \tag{12.6.8}$$

Upon substituting this into (12.3.2) we conclude that the cell problem becomes

$$-\frac{d}{dy_1}\left(a_{11}(y_1)\frac{d\chi_\ell(y_1)}{dy_1}\right) = \frac{da_{1\ell}(y_1)}{dy_1}, \quad \ell = 1, 2 \tag{12.6.9}$$

with periodic boundary conditions. Similarly, the formula for the homogenized coefficients (12.3.1) becomes:

$$\overline{a}_{ij} = \int_0^1 \left(a_{ij}(y_1) + a_{i1}(y_1)\frac{d\chi_j(y_1)}{dy_1}\right) dy_1, \quad i, j = 1, 2. \tag{12.6.10}$$

Let us now solve equations (12.6.9). These are ordinary differential equations and we can solve them in exactly the same way that we solved the one–dimensional problems in the preceding subsection. To this end, we integrate from $0$ to $y$ and divide through by $a_{11}(y_1)$ to obtain

$$\frac{d\chi_\ell}{dy_1} = -\frac{a_{1\ell}}{a_{11}} + c_1\frac{1}{a_{11}}, \quad \ell = 1,\ 2 \tag{12.6.11}$$

where the constant $c_1$ is to be determined. We have to consider the cases $\ell = 1$ and $\ell = 2$ separately. We start with $\ell = 1$. In this case the above equation simplifies to

$$\frac{d\chi_1}{dy_1} = -1 + c_1\frac{1}{a_{11}},$$

which is precisely the equation that we considered in Section 12.6.1. Thus, we have:

$$\frac{d\chi_1}{dy_1} = -1 + \frac{1}{\langle a_{11}(y)^{-1}\rangle a_{11}(y)}. \tag{12.6.12}$$

Now we consider equation (12.6.11) for the case $\ell = 2$:

$$\frac{d\chi_2}{dy_1} = -\frac{a_{12}}{a_{11}} + c_1\frac{1}{a_{11}}.$$

We integrate the above equation once again and then determine the coefficient $c_1$ by requiring $\chi_2(y_1)$ to be periodic. The final result is

$$\frac{d\chi_2(y_1)}{dy_1} = -\frac{a_{12}(y_1)}{a_{11}(y_1)} + \frac{\langle a_{12}(y_1)/a_{11}(y_1)\rangle}{\langle a_{11}^{-1}(y_1)\rangle}\frac{1}{a_{11}(y_1)}. \tag{12.6.13}$$

Now we can compute the homogenized coefficients. We start with $\overline{a}_{11}$. The calculation is the same as in the one–dimensional case:

$$\overline{a}_{11} = \langle a_{11}(y_1)^{-1}\rangle^{-1}. \tag{12.6.14}$$

We proceed with the calculation of $\overline{a}_{12}$. We substitute (12.6.13) into (12.6.10) with $i = 1$, $j = 2$ to deduce:

$$\begin{aligned}
\overline{a}_{12} &= \int_0^1 \left( a_{12}(y_1) + a_{11}(y_1)\frac{d\chi_2(y_1)}{dy_1}\right) dy \\
&= \int_0^1 \left( a_{12}(y_1) + a_{11}(y_1)\left( -\frac{a_{12}(y_1)}{a_{11}(y_1)} + \frac{\langle a_{12}(y_1)/a_{11}(y_1)\rangle}{\langle a_{11}^{-1}(y_1)\rangle}\frac{1}{a_{11}(y_1)}\right)\right) dy \\
&= \int_0^1 \left( a_{12}(y_1) - a_{12}(y_1) + \frac{\langle a_{12}(y_1)/a_{11}(y_1)\rangle}{\langle a_{11}^{-1}(y_1)\rangle}\right) dy \\
&= \frac{\langle a_{12}(y_1)/a_{11}(y_1)\rangle}{\langle a_{11}^{-1}(y_1)\rangle}.
\end{aligned}$$

Hence

$$\overline{a}_{12} = \left\langle \frac{a_{12}(y_1)}{a_{11}(y_1)}\right\rangle \langle a_{11}^{-1}(y_1)\rangle^{-1}. \tag{12.6.15}$$

Similarly,

$$\overline{a}_{21} = \left\langle \frac{a_{21}(y_1)}{a_{11}(y_1)}\right\rangle \langle a_{11}^{-1}(y_1)\rangle^{-1}. \tag{12.6.16}$$

Finally we consider $\overline{a}_{22}$ :

$$\begin{aligned}
\overline{a}_{22} &= \int_0^1 \left( a_{22}(y_1) + a_{21}(y_1)\frac{d\chi_2(y_1)}{dy_1}\right) dy \\
&= \int_0^1 \left( a_{22}(y_1) + a_{21}(y_1)\left( -\frac{a_{12}(y_1)}{a_{11}(y_1)} + \frac{\langle a_{12}(y_1)/a_{11}(y_1)\rangle}{\langle a_{11}^{-1}(y_1)\rangle}\frac{1}{a_{11}(y_1)}\right)\right) dy \\
&= \int_0^1 \left( a_{12}(y_1) - \frac{a_{12}(y_1)a_{21}(y_1)}{a_{11}(y_1)} + \frac{a_{21}(y_1)}{a_{11}(y_1)}\frac{\langle a_{12}(y_1)/a_{11}(y_1)\rangle}{\langle a_{11}^{-1}(y_1)\rangle}\right) dy \\
&= \left\langle \frac{a_{21}(y_1)}{a_{11}(y_1)}\right\rangle \left\langle \frac{a_{12}(y_1)}{a_{11}(y_1)}\right\rangle \langle a_{11}^{-1}(y_1)\rangle^{-1} + \left\langle a_{22}(y_1) - \frac{a_{12}(y_1)a_{21}(y_1)}{a_{11}(y_1)}\right\rangle.
\end{aligned}$$

Consequently:

$$\overline{a}_{22} = \left\langle \frac{a_{21}(y_1)}{a_{11}(y_1)}\right\rangle \left\langle \frac{a_{12}(y_1)}{a_{11}(y_1)}\right\rangle \langle a_{11}^{-1}(y_1)\rangle^{-1} + \left\langle a_{22}(y_1) - \frac{a_{12}(y_1)a_{21}(y_1)}{a_{11}(y_1)}\right\rangle. \tag{12.6.17}$$

It is evident from formulae (12.6.14), (12.6.15), (12.6.16) and (12.6.17) that the homogenized coefficients depend on the original ones in a very complicated, highly nonlinear way.

## 12.7 Discussion and Bibliography

The method of multiple scales was developed by various researchers in the 70s with significant contributions from Keller, Babuska, Sanchez–Palenzia, Bensoussan, Lions, Papanicolaou and others. See [158, 159, 26, 25, 24, 23, 91] and the references therein. A first systematic exposition of the method of multiple scales is contained in [33], where references to the earlier literature can be found. See also the book [279]. Rigorous convergence results for elliptic PDEs with rapidly oscillating coefficients were proved before the development of the method of multiple scales. See [73, 296] and the text [153]. However the power of the method of multiple scales is its wide applicability to a variety of differing settings. In contrast, rigorous results tend to apply on a case by case basis and their proofs differ substantially between different PDEs, and between Markov chains, ODEs and SDEs. (See Part III of this book). In most cases, however, an appropriate Poisson equation (the cell problem) plays a prominent role in the analysis.

The one dimensional problem (see Section 12.6.1) was studied in [296], without using the method of multiple scales. In the one dimensional case it is possible to derive the homogenized equation using the method of multiple scales even in the nonperiodic setting; see [143, Ch. 5], [66, Ch. 5]. The homogenized equation is a second order uniformly elliptic PDE in the case of nonperiodic fast oscillatory coefficients. However this result is most naturally obtained via the theory of $H-$ and $\Gamma-$convergence, rather than multiple-scale expansions. See [296], [308]. In the general setting of nonperiodic, deterministic, homogenization the homogenized coefficients cannot be expressed in terms of solutions to appropriate Poisson equations and there are no explicit formulae for them. In this case, the best one can hope for is to obtain bounds on the homogenized coefficients.

The homogenized equation for layered materials (see Section 12.6.2) was derived rigorously by Murat and Tartar without any appeal to the method of multiple scales; see [232] and the references to the original papers therein. The two–dimensional case that we treated in subsection 12.6.2 can be easily extended to the $d$–dimensional one, $d \geqslant 2$, i.e. to the case where $a_{ij}(y) = a_{ij}(y_1)$, $i, j = 1, \ldots, d$. See [232].

The elliptic boundary value problem (12.2.1) is a Dirichlet problem. However, an inspection of the analysis presented in Section 12.4 reveals that the boundary conditions did not play any role in the derivation of the homogenized equation. In particular, the two–scale expansion (12.4.1) that we used in order to derive the homogenized equation did not contain any information concerning the boundary conditions of the problem under investigation. Indeed, the boundary conditions become somewhat irrelevant in the homogenization procedure. Exactly the same calculations enable us to obtain the homogenized equation for Neumann or mixed boundary conditions. This is not surprising since the derivation of the homogenized equation is based on the

analysis of local problems of the form (12.2.3). This local problem cannot really see the boundary – this is the key property of scale separation.

However, the boundary conditions become very important when trying to prove the homogenization theorem. The fact that the two–scale expansion (12.4.1) does not satisfy the boundary conditions of our PDE exactly but, rather, only up to $\mathcal{O}(\varepsilon)$, introduces boundary layers [143, ch. 3]. [1] Boundary layers affect the convergence rate at which $u^\varepsilon(x)$ converges to $u(x)$ as $\varepsilon \to 0$. We can solve this problem by modifying the two–scale expansion (12.4.1), adding additional terms which take care of the boundary layer and vanish exponentially fast as we move away from the boundary so that they do not affect the solution in the interior. We refer to [27] for details.

The discussion in Remark 12.2 is further elaborated in [33] and in [66]. Different expressions for the effective diffusion tensor can be useful for the proof of various properties of the the effective diffusion tensor.

From the point of view of continuum mechanics, the method of homogenization enables us to obtain macroscopic *constitutive laws* for composite materials. Macroscopic constitutive laws have been derived using homogenization theory for various types of composite materials. See, e.g. [46, 108]. An alternative approach is presented in [230, 133]. The theory of composite materials is presented in the excellent monograph [229].

In the Dirichlet problem that we analyzed in Section 12.4 we assumed that the matrix $A^\varepsilon(x)$ depends only on the microscale, i.e.

$$A^\varepsilon(x) = A\left(\frac{x}{\varepsilon}\right),$$

with $A(y)$ being a 1–periodic matrix valued function. However, the method of multiple scales is also applicable to the case where the coefficients depend explicitly on the macroscale as well as the microscale:

$$A^\varepsilon(x) = A\left(x, \frac{x}{\varepsilon}\right),$$

with $A(x, y)$ being 1–periodic in $y$ and smooth in $x$. When the coefficients have this form they are called *locally periodic* or *non–uniformly periodic* . Analysis similar to the one presented in Section 12.4 enables us to obtain the homogenized equation for the Dirichlet problem

$$-\nabla \cdot (A^\varepsilon \nabla u^\varepsilon) = f \quad \text{for } x \in \Omega, \tag{12.7.1a}$$

$$u^\varepsilon = 0 \quad \text{for } x \in \partial\Omega, \tag{12.7.1b}$$

where $A^\varepsilon(x) = A(x, x/\varepsilon)$. Now the homogenized coefficients $\overline{A}$ are functions of $x$:

$$-\nabla \cdot \left(\overline{A}\nabla u\right) = f \quad \text{for } x \in \Omega \tag{12.7.2a}$$

[1] The presence of boundary and initial layers is a common feature in all problems of singular perturbations. See the bibliographical discussions in other chapters from Part II, and [143] and [161], for further details.

$$u = 0 \ \text{ for } x \in \partial\Omega, \tag{12.7.2b}$$

and the cell problem is parameterized by $x$ since $A = A(x, y)$:

$$-\nabla_y \cdot \left(\nabla_y \chi A^T\right) = \nabla_y \cdot A^T, \quad y \in \mathbb{T}^d. \tag{12.7.3}$$

The homogenized coefficients are given by the formula:

$$\overline{A}(x) = \int_{\mathbb{T}^d} \left(A(x,y) + A(x,y)\nabla_x \chi(x,y)^T\right)\, dy. \tag{12.7.4}$$

We emphasize the fact that the "macroscopic variable" $x$ enters in the above two equations as a parameter. Consequently, in order to compute the effective coefficients we need to solve the cell problem (12.7.3) and evaluate the integrals in (12.7.4) at all points $x \in \Omega$.

The method of multiple scales can also be applied to semilinear elliptic PDEs with rapidly oscillating coefficients – equations of the form

$$-\nabla \cdot \left(A^\varepsilon \nabla u^\varepsilon\right) = f(u^\varepsilon) \ \text{ for } x \in \Omega, \tag{12.7.5a}$$

$$u^\varepsilon = 0 \ \text{ for } x \in \partial\Omega. \tag{12.7.5b}$$

The homogenized equation takes the form

$$-\nabla \cdot \left(\overline{A}\nabla u\right) = f(u) \ \text{ for } x \in \Omega, \tag{12.7.6a}$$

$$u = 0 \ \text{ for } x \in \partial\Omega, \tag{12.7.6b}$$

with $\overline{A}$ as in (12.3.1).

In section (12.2) we obtained the first two terms in the two–scale expansion for the Dirichlet problem (12.2.1). The second term is proportional–up to an unknown function of $x$– to the gradient of the first term in the expansion which solves the homogenized equation, i.e

$$u_1\left(x, \frac{x}{\varepsilon}\right) = \chi\left(\frac{x}{\varepsilon}\right) \cdot \nabla_x u(x) + \widehat{u}_1(x), \tag{12.7.7}$$

where $\chi(y)$ solves the cell problem. We can also solve higher order equations and obtain higher order terms in the two–scale expansion. For example, we can solve equation (12.4.6) and compute the third term in the expansion $u_2(x, y)$:

$$u_2(x,y) = \Theta(y) : \nabla_x \nabla_x u(x) + \widehat{u}_2(x) \tag{12.7.8}$$

where the *second order corrector field* $\Theta(y)$ is a matrix valued function which satisfies the boundary value problem

$$\mathcal{A}_0 \Theta = B. \tag{12.7.9}$$

Here $B(y)$ is given by

$$B(y) := -\overline{A} + A(y) + A(y)\nabla_y \chi(y)^T + \nabla_y \chi(y) A(y) + \chi(y) \otimes \left(\nabla_y \cdot A(y)^T\right).$$

All higher order equations are of the form

$$\mathcal{A}_0 u_{k+2} = -\mathcal{A}_1 u_{k+1} - \mathcal{A}_0 u_k, \quad k = 1, 2, \ldots$$

It turns out that $u_k(x)$ is proportional to the $k^{th}$ order derivatives of $u(x)$. See [27].

The method of multiple scales can be extended to situations where there are $k$ length scales in the problem, i.e. when the matrix $A^\varepsilon(x)$ has the form

$$A^\varepsilon(x) = A\left(\frac{x}{\varepsilon}, \frac{x}{\varepsilon^2}, \ldots, \frac{x}{\varepsilon^k}\right),$$

and $A$ is 1–periodic in all of its arguments. This is known as *reiterated homogenization* – [33, Sec. 1.8]. A rigorous analysis of reiterated homogenization in a quite general setting is presented in [8]. Reiterated homogenization has recently found applications in the problem of advection and diffusion of passive tracers in fluids. See, for example, [253, 219, 220] for details. When there are infinitely many scales in the problem, without a clear separation, the homogenization result breaks down, in the sense that the homogenized coefficient can be $0$. See [16].

In general it is not possible to compute the homogenized coefficients analytically; indeed, their calculation requires the solution of the cell problem and the calculation of the integrals in (12.3.1). In most cases this can be done only numerically. It is possible, however, to obtain bounds on the magnitude of the effective coefficients. Various tools for obtaining bounds have been developed; for example it is possible to obtain a variational characterization of the homogenized coefficients. We refer to [229, 311, 107] for various results in this direction. Many of these techniques apply to the nonperiodic setting.

The method developed in this chapter readily extends to intial/boundary value problem such as the following parabolic PDE:

$$\frac{\partial u^\varepsilon}{\partial t} - \nabla \cdot (A^\varepsilon \nabla u^\varepsilon) = f^\varepsilon \ \text{ in } \Omega \times (0, T), \tag{12.7.10a}$$

$$u^\varepsilon = 0 \ \text{ on } \partial\Omega \times (0, T) \tag{12.7.10b}$$

$$u^\varepsilon = u_{in}(x) \ \text{ in } \overline{\Omega} \times \{0\} \tag{12.7.10c}$$

under various assumptions concerning the $\varepsilon$ dependence in $A^\varepsilon$ and $f^\varepsilon$. A time-dependent situation of interest arises when the coefficients of the evolution PDE oscillate in time as well as space, i.e. $A^\varepsilon = A\left(\frac{x}{\varepsilon}, \frac{t}{\varepsilon^k}\right)$, $k > 0$ with the matrix valued function $A(y, \tau)$ being 1–periodic in both $y$ and $\tau$. This means that we have to introduce two fast variables: $y = \frac{x}{\varepsilon}$ and $\tau = \frac{t}{\varepsilon^k}$. More information on homogenization for evolution equations with space–time dependent coefficients can be found in [33, Ch. 3]. We study homogenization for parabolic PDEs using the method of multiple scales in Chapters 11, 13 and 14.

One can also study the problem of homogenization for hyperbolic (wave) equations:

$$\frac{\partial^2 u^\varepsilon}{\partial t^2} - \nabla \cdot (A^\varepsilon \nabla u^\varepsilon) = f \ \text{ in } \Omega \times (0,T), \tag{12.7.11a}$$

$$u^\varepsilon = 0 \ \text{ on } \partial\Omega \times (0,T), \tag{12.7.11b}$$

$$u^\varepsilon = u_{in} \ \text{ in } \overline{\Omega} \times \{0\}, \tag{12.7.11c}$$

$$\frac{\partial u^\varepsilon}{\partial t} = v_{in}(x) \ \text{ in } \overline{\Omega} \times \{0\}. \tag{12.7.11d}$$

The method of multiple scales can be used to obtain a homogenized equation, which is a wave equation with constant coefficients and the same initial and boundary conditions. However there is a fundamental difference between this and the parabolic case: for parabolic problems the dissipation drives the solution to lie near to the null space of the leading order operator $\mathcal{L}_0$, no matter how the initial data is chosen. For the wave equation this does not happen and it is necessary to chose initial data close to the desired subspace. We will not study homogenization for wave equations in this book. We refer the interested reader to [66, Ch. 12], [33, ch. 2], [160, 47]. Related problems arise for the Schrödinger equation with multiple scales – see [316]. Homogenization result for the Schrödinger equation and their connection to effective mass theorems are presented in [10].

The numerical evaluation of homogenized coefficients, in the periodic setting, can be performed efficiently using a spectral method. On the other hand, the numerical solution of the original boundary value problem (12.2.1) when $\varepsilon$ is small is a very hard problem. Special methods, which in one way or another are based on homogenization, have been developed over the last few years. We refer to [145, 76, 2, 82, 9, 52, 61, 89, 90, 92, 231] and the references therein on this topic. The development and analysis of finite element methods for elliptic PDEs with a multiscale structure, and related problems arising in geophysical applications, is discussed in [60, 88, 145, 146]. Numerical methods for elliptic PDEs subject to stochastic forcing, or with stochastic coefficients, is described in [3, 141, 216, 215, 286, 287].

## 12.8 Exercises

1. Consider the problem of homogenization for (12.2.1) when the coefficients matrix $A(y)$ has different period in each direction

$$A(y + \lambda_k e_k) = A(y), \quad k = 1, \ldots,$$

with $\lambda_k > 0$, $k = 1, \ldots d$. Write down the formulas for the homogenized coefficients.

2. Consider the two–scale expansion (12.4.1) for problem (12.2.1). In this chapter we calculated the first three terms in the two–scale expansion: $u_0$ solves the homogenized equation, $u_1$ is given by (12.7.7) and $u_2$ by (12.7.8). Verify the expression for $u_2$, and the form of the higher order cell problem (12.7.9).

3. Consider the Dirichlet problem (12.2.1) for a $d$–dimensional layered material, i.e.
$$a_{ij}(y) = a_{ij}(y_1), \quad \text{1–periodic in } y_1, \quad i,j = 1,\dots,d.$$
We solved this problem in Subsection 12.6.2 in the case $d = 2$. Now solve the corresponding cell problem and obtain formulas for the homogenized coefficients for $d \geqslant 3$, arbitrary.
4. Consider the problem of homogenization for second order uniformly elliptic PDE in 1 dimension, i.e. the problem studied in Section 12.6.1.
   a. Calculate $\overline{a}$ for the case
   $$a(y) = \begin{cases} a_1 & : \quad y \in [0, \frac{1}{2}], \\ a_2 & : \quad y \in (\frac{1}{2}, 1], \end{cases}$$
   where $a_1,\ a_2$ are positive constants.
   b. Now calculate $\overline{a}$ for the case
   $$a(y) = \frac{1}{2 + \sin(2\pi y)}.$$
5. Consider the Dirichlet problem (12.2.1) for a $d$–dimensional isotropic material, i.e.
$$a_{ij}(y) = a(y)\delta_{ij}, \quad \text{1–periodic}, \quad i,j = 1,\dots,d,$$
where $\delta_{ij}$ stands for Kronecker's delta.
   a. Use the specific structure of $A(y)$ to simplify the cell problem as much as you can.
   b. Let $d = 2$ and assume that $a(y)$ is of the form
   $$a(y) = Y_1(y_1)Y_2(y_2).$$
   Solve the two components of the cell problem and obtain formulae for the homogenized coefficients (hint: use separation of variables).
6. Consider the boundary value problem (12.7.1). Assume that $A^\varepsilon = A(x, \frac{x}{\varepsilon})$ where $A(x,y)$ is smooth, 1–periodic in $y$ and uniformly elliptic and that, furthermore, $f$ is smooth. Use the method of multiple scales to obtain generalizations of the homogenized equation (12.7.2), the cell problem (12.7.3) and the formula for the homogenized coefficients (12.7.4). Verify that the results of section 12.5 still hold.
7. Consider the Dirichlet problem
$$-\nabla \cdot \left( A\left(\frac{x}{\varepsilon}, \frac{x}{\varepsilon^2}\right) \nabla u^\varepsilon \right) = f \quad \text{for } x \in \Omega \tag{12.8.1a}$$
$$u^\varepsilon(x) = 0, \quad \text{for } x \in \partial\Omega. \tag{12.8.1b}$$
where the coefficients $A(y,z)$ are periodic in both $y$ and $z$ with period 1. Use the 3–scale expansion
$$u^\varepsilon(x) = u_0\left(x, \frac{x}{\varepsilon}, \frac{x}{\varepsilon^2}\right) + \varepsilon u_1\left(x, \frac{x}{\varepsilon}, \frac{x}{\varepsilon^2}\right) + \varepsilon^2 u_2\left(x, \frac{x}{\varepsilon}, \frac{x}{\varepsilon^2}\right) + \dots$$

to derive an effective homogenized equation, together with the formula for the homogenized coefficients and two cell problems.

8. Repeat the previous exercise by homogenizing first with respect to $z = y/\varepsilon$ and then with respect to $y$:
   a. Homogenize the equation

$$-\nabla \cdot \left( A\left(y, \frac{y}{\varepsilon}\right) \nabla u^{\varepsilon} \right) = f, \quad \text{for } x \in \Omega \tag{12.8.2a}$$

$$u^{\varepsilon}(x) = 0, \quad \text{for } x \in \partial\Omega \tag{12.8.2b}$$

   by treating $y$ as a parameter.
   b. Homogenize the equation

$$-\nabla \cdot \left( \overline{A}\left(\frac{x}{\varepsilon}\right) \nabla \overline{u}^{\varepsilon} \right) = f, \quad \text{for } x \in \Omega \tag{12.8.3a}$$

$$u^{\varepsilon}(x) = 0, \quad \text{for } x \in \partial\Omega, \tag{12.8.3b}$$

   where $\overline{A}(y)$ is given by the expression derived in the preceding section of the question.
9. Derive the homogenized equation, together with the cell problem and the formula for the homogenized coefficients, by applying the method of multiple scales to the heat equation (12.7.10), with $A^{\varepsilon} = A(\frac{x}{\varepsilon})$.
10. Consider the initial boundary value problem (12.7.10) with $A^{\varepsilon} = A(\frac{x}{\varepsilon}, \frac{t}{\varepsilon^k})$. Explain why it is natural for the period of oscillations in time to be characterized by $k = 2$. Carry out homogenization for the cases $k = 1, 2, 3..$[2]
11. Use the method of multiple scales to derive the homogenized equation from (12.7.11).
12. Prove that the homogenized coefficient $\overline{a}$ for equation (12.6.1) under (12.6.2) has the same upper and lower bound as $a(y)$:

$$\alpha \leqslant \overline{a} \leqslant \beta.$$

    Moreover, show that it is bounded from above by the average of $a(y)$:

$$\overline{a} \leqslant \langle a(y) \rangle.$$

13. Show that the equation (12.7.5) can be homogenized to obtain the effective equation (12.7.6).
14. Let $A : \mathbb{T}^d \to \mathbb{R}^{d \times d}$ be smooth and periodic and consider the eigenvalue problem

$$-\nabla \cdot \left( A^{\varepsilon} \nabla u^{\varepsilon} \right) = \lambda^{\varepsilon} u^{\varepsilon} \quad \text{for } x \in \Omega$$
$$u^{\varepsilon} = 0, \quad x \in \partial\Omega,$$

    where $A^{\varepsilon}(x) = A(x/\varepsilon)$. Use a multiscale expansion to find an approximation to the eigenvalue problem in which $\varepsilon \to 0$ is eliminated.

[2] See [33, ch.3] and [253] for further details on the derivation of the homogenized equations using the method of multiple scales.

15. a. Consider the eigenvalue problem

$$-\Delta u^\varepsilon + \frac{1}{\varepsilon} V^\varepsilon u^\varepsilon = \lambda^\varepsilon u^\varepsilon, \quad x \in \Omega$$
$$u^\varepsilon = 0, \quad x \in \partial\Omega.$$

Assume that $V : \mathbb{T}^d \to \mathbb{R}$ is smooth and periodic, that

$$\int_{\mathbb{T}^d} V(y)dy = 0$$

and that $V^\varepsilon(x) = V(x/\varepsilon)$. Use a multiscale expansion to find an approximation to the eigenvalue problem in which $\varepsilon \to 0$ is eliminated.

b. (ii) Are the resulting eigenvalues smaller or larger than the eigenvalues which arise when $V \equiv 0$?

# 13

# Homogenization for Parabolic PDEs

## 13.1 Introduction

In this chapter we use multiscale techniques to investigate the long time behavior of solutions to parabolic PDEs. The techniques employed are almost identical to those used in the study of homogenization for SDEs in Chapter 11. This connection will be made more explicit at the end of the chapter.

In Section 13.2 we present the full equations that we will analyze. Section 13.3 contains the simplified equations that are derived by use of the method of multiple scales in Section 13.4. Section 13.5 is devoted to various properties of the simplified equations. In Section 13.6 we study two applications of the general theory, to gradient flows (Section 13.6.1) and to divergence free flows (Section 13.6.2). The connection between homogenization for parabolic PDEs and asymptotic problems for SDEs is made in Section 13.7. Extensions and bibliographical remarks appear in Section 13.8.

## 13.2 Full Equations

We study the following initial value (Cauchy) problem

$$\frac{\partial u}{\partial t} = b \cdot \nabla u + D\Delta u \quad \text{for } (x,t) \in \mathbb{R}^d \times \mathbb{R}^+, \tag{13.2.1a}$$

$$u = u_{in} \quad \text{for } (x,t) \in \mathbb{R}^d \times \{0\}, \tag{13.2.1b}$$

with $D > 0$. In our analysis we will assume that the vector $b(x)$ is smooth and periodic in space with period 1 in all spatial directions. Furthermore, we assume that the initial conditions are slowly varying, so that

$$u_{in}(x) = g^{\varepsilon}(x) := g(\varepsilon x), \tag{13.2.2}$$

with $0 < \varepsilon \ll 1$. Since the initial data is slowly varying and so is the solution, it is natural to look at large length and time scales to see the effective behavior of the

PDE (13.2.1). If the vector field $b$ averages to zero in an appropriate sense then, as we will show in this chapter, the effective behavior of $u$ is that of a pure diffusion.

To see this effect we redefine the variables $x,\ t$ through the rescaling

$$x = \varepsilon^{-1}x, \quad t \to \varepsilon^{-2}t \tag{13.2.3}$$

and relabel $u$ to $u^\varepsilon$ to emphasize this rescaling. This particular scaling of space and time, known as the **diffusive scaling**, is appropriate whenever the advective effects, created by $b$, are expected to average out; it is then appropriate to scale time on an even longer scale than space, and seek purely diffusive effects. We will be precise about the condition that $b$ averages out at the end of this section.

The rescaled field $u^\varepsilon(x,t)$ satisfies the equation

$$\frac{\partial u^\varepsilon}{\partial t} = \frac{1}{\varepsilon}b^\varepsilon \cdot \nabla u^\varepsilon + D\Delta u^\varepsilon \quad \text{for } (x,t) \in \mathbb{R}^d \times \mathbb{R}^+, \tag{13.2.4a}$$

$$u^\varepsilon = g \quad \text{for } (x,t) \in \mathbb{R}^d \times \{0\}. \tag{13.2.4b}$$

Here $b^\varepsilon(x) = b(x/\varepsilon)$. This equation will be the object of our study in this chapter.

Let us define the operator

$$\mathcal{L}_0 = b(y) \cdot \nabla_y + D\Delta_y \tag{13.2.5}$$

with periodic boundary conditions on $[0,1]^d$ and its $L^2$–adjoint $\mathcal{L}_0^*$, also with periodic boundary conditions. We refer to $D$ as the **molecular diffusivity.** Note that $\mathcal{L}_0$ is the generator of the Markov process $y(t)$ which is the solution of the SDE

$$\frac{dy}{dt} = b(y) + \sqrt{2D}\frac{dW}{dt}$$

on the unit torus $\mathbb{T}^d$. Hence it is natural to define the **invariant distribution** $\rho(y)$ to be the stationary solution of the adjoint equation:

$$\mathcal{L}_0^*\rho = 0. \tag{13.2.6}$$

By Theorem 6.16 there is a unique solution to this equation, up to normalization, and the normalization may be chosen so that the solution is positive. In the sequel we will normalize the solution to (13.2.6) according to

$$\int_{\mathbb{T}^d} \rho(y)\, dy = 1.$$

Notice that this choice turns the measure $\mu(dy) = \rho(y)\, dy$ into a probability measure on $\mathbb{T}^d$.

In order to derive the homogenized equation for (13.2.4) we need to study equations of the form

$$-\mathcal{L}_0 v = h \tag{13.2.7}$$

with periodic boundary conditions and with $h$ being a smooth periodic function of $y$. It is straightforward to check that the assumptions of Theorem 7.9 are satisfied and

hence the operator $\mathcal{L}_0$ satisfies the Fredholm alternative. This implies, in particular, that $\mathcal{L}_0$ has a one-dimensional null space, comprising constants in $y$. It also implies that $\mathcal{L}_0^*$ has a one-dimensional null space, as stated above, and spanned by $\rho$. Furthermore, equation (13.2.7) has a solution if and only if the right hand side of the equation is centered with respect to the invariant distribution:

$$\int_{\mathbb{T}^d} h(y)\rho(y)\,dy = 0.$$

In this case the solution of (13.2.7) is unique up to constants. In the case where $h = b$, the vector field arising in the PDE (13.2.1), the condition is

$$\int_{\mathbb{T}^d} b(y)\rho(y)\,dy = 0. \tag{13.2.8}$$

We call this the **centering condition**. We fix the free constant in the solution (13.2.7) by requiring that the solution of (13.2.7) satisfies

$$\int_{\mathbb{T}^d} v(y)\rho(y)\,dy = 0. \tag{13.2.9}$$

When the centering condition is not satisfied it is necessary to rescale the original problem in a different fashion, to see effective *advective* behavior. In particular (13.2.3) is replaced by the **advective scaling**

$$x \to \varepsilon^{-1}x, \quad t \to \varepsilon^{-1}t. \tag{13.2.10}$$

Then averaging is used to find the effective equation, which is now of transport type. See Chapter 14.

## 13.3 Simplified Equations

Assume that the vector field $b(y)$ satisfies the centering condition (13.2.8). Define the vector field $\chi(y)$ to be the solution of the *cell problem*

$$-\mathcal{L}_0\chi = b, \ \ \chi \text{ is } 1\text{–periodic}, \ \int_{\mathbb{T}^d} \chi(y)\rho(y)dy = 0. \tag{13.3.1}$$

The *effective diffusion tensor* (or *effective diffusivity*) is defined as

$$\mathcal{K} = DI + 2D\int_{\mathbb{T}^d} \nabla_y\chi(y)^T\rho(y)\,dy + \int_{\mathbb{T}^d} \Big(b(y)\otimes\chi(y)\Big)\rho(y)\,dy. \tag{13.3.2}$$

**Result 13.1.** *Assume that* (13.2.8) *holds. For* $0 < \varepsilon \ll 1$ *and times* $t$ *of* $\mathcal{O}(1)$ *the solution* $u^\varepsilon$ *of* (13.2.4) *is approximated by* $u$*, the solution of the homogenized equation*

$$\frac{\partial u}{\partial t} = \mathcal{K} : \nabla_x\nabla_x u \quad \textit{for} \ \ (x,t) \in \mathbb{R}^d \times \mathbb{R}^+, \tag{13.3.3a}$$

$$u = g \quad \textit{for} \ \ (x,t) \in \mathbb{R}^d \times \{0\}. \tag{13.3.3b}$$

*Remark 13.2.* Since the Hessian $\nabla_x \nabla_x u$ is symmetric, it follows from property (2.2.2) applied to (13.3.2) that the following expression for $\mathcal{K}$ is equally valid:

$$\mathcal{K} = DI + D\int_{\mathbb{T}^d} \left(\nabla_y \chi(y) + \nabla_y \chi(y)^T\right) \rho(y)\, dy + \frac{1}{2}\int_{\mathbb{T}^d} \Big(b(y) \otimes \chi(y) + \chi(y) \otimes b(y)\Big)\rho(y)\, dy. \tag{13.3.4}$$

Many variants on this idea are possible. □

## 13.4 Derivation

Our goal now is to use the method of multiple scales in order to analyze the behavior of $u^\varepsilon(x,t)$, the solution of (13.2.4), in the limit as $\varepsilon \to 0$. In particular, we want to derive Result 13.1.

We introduce the auxiliary variable $y = x/\varepsilon$. [1] Let $\phi = \phi(x, x/\varepsilon)$ be scalar-valued. The chain rule gives

$$\nabla\phi = \nabla_x\phi + \frac{1}{\varepsilon}\nabla_y\phi \quad \text{and} \quad \Delta\phi = \Delta_x\phi + \frac{2}{\varepsilon}\nabla_x \cdot \nabla_y\phi + \frac{1}{\varepsilon^2}\Delta_y\phi.$$

The partial differential operator that appears on the right hand side of equation (13.2.4) now becomes

$$\mathcal{L} = \frac{1}{\varepsilon^2}\mathcal{L}_0 + \frac{1}{\varepsilon}\mathcal{L}_1 + \mathcal{L}_2,$$

where

$$\begin{aligned}\mathcal{L}_0 &= b(y)\cdot\nabla_y + D\Delta_y,\\ \mathcal{L}_1 &= b(y)\cdot\nabla_x + 2D\nabla_x\cdot\nabla_y,\\ \mathcal{L}_2 &= D\Delta_x.\end{aligned}$$

In terms of $x$ and $y$ equation (13.2.4a) becomes

$$\frac{\partial u^\varepsilon}{\partial t} = \left(\frac{1}{\varepsilon^2}\mathcal{L}_0 + \frac{1}{\varepsilon}\mathcal{L}_1 + \mathcal{L}_2\right) u^\varepsilon.$$

We seek a solution in the form of a multiple scales expansion

$$u^\varepsilon(x,t) = u_0(x,y,t) + \varepsilon u_1(x,y,t) + \varepsilon^2 u_2(x,y,t) + \ldots \tag{13.4.1}$$

where $u_j(x,y,t),\ j = 1, 2\ldots$, are periodic in $y$ with period 1. We substitute (13.4.1) and equate terms of equal powers in $\varepsilon$. We obtain the following sequence of equations:

[1] As in the elliptic case, this is where the assumption of scale separation is exploited : **we treat $x$ and $y$ as independent variables**. Justifying this assumption as $\varepsilon \to 0$ is one of the main issues in the rigorous theory of homogenization. See Chapter 20.

$$\mathcal{O}(1/\varepsilon^2) \quad -\mathcal{L}_0 u_0 = 0, \tag{13.4.2a}$$

$$\mathcal{O}(1/\varepsilon) \quad -\mathcal{L}_0 u_1 = \mathcal{L}_1 u_0, \tag{13.4.2b}$$

$$\mathcal{O}(1) \quad -\mathcal{L}_0 u_2 = \mathcal{L}_1 u_1 + \mathcal{L}_2 u_0 - \frac{\partial u_0}{\partial t}. \tag{13.4.2c}$$

Note that $\mathcal{L}_0$, which is a differential operator in $y$ only, is equipped with periodic boundary conditions.

Since $\mathcal{L}_0$ has a one-dimensional null-space, equation (13.4.2a) implies that the first term in the expansion is independent of $y$, so that $u_0 = u(x,t)$ only. Notice that

$$\mathcal{L}_1 u_0 = b(y) \cdot \nabla_x u(x,t).$$

The centering condition (13.2.8) ensures that (13.4.2b) has a solution, by the Fredholm alternative. Since $\mathcal{L}_0$ is a differential operator in $y$ only, we may use separation of variables to write the solution as

$$u_1(x,y,t) = \chi(y) \cdot \nabla_x u(x,t).$$

Then $\chi(y)$ solves the cell problem (13.3.1) . Our assumptions imply that there exists a unique, smooth solution to the cell problem.

Now we proceed with the analysis of the $\mathcal{O}(1)$ equation (13.4.2c). The solvability condition (13.2.8) reads

$$\int_{\mathbb{T}^d} \left( \frac{\partial u_0}{\partial t} - \mathcal{L}_2 u_0 - \mathcal{L}_1 u_1 \right) \rho \, dy = 0.$$

The fact that $u_0 = u(x,t)$ is independent of $y$ enables us to rewrite the above equation in the form

$$\frac{\partial u}{\partial t} = D\Delta u + \int_{\mathbb{T}^d} \Big( \mathcal{L}_1 u_1 \Big) \rho \, dy. \tag{13.4.3}$$

Now we have

$$\begin{aligned} \mathcal{L}_1 u_1 &= \big( b \cdot \nabla_x (\chi \cdot \nabla_x u) + 2D \nabla_x \cdot \nabla_y (\chi \cdot \nabla_x u) \big) \\ &= \big( b \otimes \chi + 2D \nabla_y \chi^T \big) : \nabla_x \nabla_x u. \end{aligned}$$

In view of the above calculation, equation (13.4.3) becomes

$$\frac{\partial u}{\partial t} = \mathcal{K} : \nabla_x \nabla_x u,$$

which is the homogenized equation (13.3.3a). The effective diffusivity $\mathcal{K}$ is given by formula (13.3.2).

## 13.5 Properties of the Simplified Equations

In this section we show that the effective diffusivity is positive definite. This implies that the homogenized equation is well posed. To prove this we need to calculate the Dirichlet form associated with the operator $\mathcal{L}_0$. The following is a direct consequence of Theorem 6.12 in the case of additive noise. Recall that $\rho$ is the invariant distribution, a non-negative $L^1(\mathbb{T}^d)$ function in the null-space of $\mathcal{L}_0^*$.

**Lemma 13.3.** *Let $f(y) \in C^2_{per}(\mathbb{T}^d)$. Then*

$$\int_{\mathbb{T}^d} (-\mathcal{L}_0 f(y)) f(y) \rho(y)\, dy = D \int_{\mathbb{T}^d} |\nabla_y f(y)|^2 \rho(y)\, dy. \tag{13.5.1}$$

*Remark 13.4.* Let $L^2_\rho(\mathbb{T}^d)$ be the $L^2$ space weighted by the invariant distribution $\rho(y)$ and denote the inner product and corresponding norm by $(\cdot,\cdot)_{L^2_\rho}$ and $\|\cdot\|_{L^2_\rho}$, respectively. Then, by equation (6.3.13), the result of Lemma 13.3 can be expressed in the form

$$(-\mathcal{L}_0 f, f)_{L^2_\rho} = D\|\nabla_y f\|^2_{L^2_\rho}. \quad \square$$

The main result of this section is that the effective diffusivity is a positive definite matrix. In particular, we have the following.

**Theorem 13.5.** *Let $\xi \in \mathbb{R}^d$ be an arbitrary vector and let $\chi_\xi(y) := \chi(y) \cdot \xi$. Then*

$$\langle \xi, \mathcal{K}\xi \rangle = D \int_{\mathbb{T}^d} |\xi + \nabla_y \chi_\xi(y)|^2 \rho(y)\, dy.$$

*Furthermore,*

$$\alpha |\xi|^2 \leqslant \langle \xi, \mathcal{K}\xi \rangle \quad \forall\, \xi \in \mathbb{R}^d. \tag{13.5.2}$$

*with*

$$\alpha = D \left( \int_{\mathbb{T}^d} \rho^{-1}(y)\, dy \right)^{-1}. \tag{13.5.3}$$

*Proof.* Note that $-\mathcal{L}_0 \chi_\xi = \xi \cdot b$. We use the definition of $\mathcal{K}$ and Lemma 13.3 to calculate

$$\begin{aligned}
\langle \xi, \mathcal{K}\xi \rangle &= D|\xi|^2 + 2D \int_{\mathbb{T}^d} \xi \cdot \nabla_y \chi_\xi(y) \rho(y)\, dy + \int_{\mathbb{T}^d} (\xi \cdot b) \chi_\xi(y) \rho(y)\, dy \\
&= D|\xi|^2 + 2D \int_{\mathbb{T}^d} \xi \cdot \nabla_y \chi_\xi(y) \rho(y)\, dy + D \int_{\mathbb{T}^d} |\nabla_y \chi_\xi(y)|^2 \rho(y)\, dy \\
&= D \int_{\mathbb{T}^d} |\xi + \nabla_y \chi_\xi(y)|^2 \rho(y)\, dy.
\end{aligned}$$

The fact that the effective diffusivity in nonnegative definite follows immediately from the above equation. To show that $\mathcal{K}$ is positive definite we use the fact that the integral of derivatives of periodic functions over $\mathbb{T}^d$ is 0, together with the Cauchy–Schwarz inequality and the fact that $\rho(y)$ is everywhere positive, to calculate:

$$\begin{aligned}
D|\xi|^2 &= D \left| \int_{\mathbb{T}^d} (\xi + \nabla_y \chi_\xi)\ dy \right|^2 \\
&= D \left| \int_{\mathbb{T}^d} (\xi + \nabla_y \chi_\xi)\, \rho^{\frac{1}{2}}(y) \rho^{-\frac{1}{2}}(y)\, dy \right|^2 \\
&\leqslant D \left( \int_{\mathbb{T}^d} |\xi + \nabla_y \chi_\xi|^2 \rho(y)\, dy \right) \left( \int_{\mathbb{T}^d} \rho^{-1}(y)\, dy \right) \\
&= \langle \xi, \mathcal{K}\xi \rangle \left( \int_{\mathbb{T}^d} \rho^{-1}(y)\, dy \right),
\end{aligned}$$

from which the lower bound immediately follows. □

It is of interest to know how the effective diffusion tensor $\mathcal{K}$ compares with the original diffusion tensor $DI$. It turns out that $\mathcal{K}$ can be either greater or smaller than $D$ (in the sense of matrices). This issue is discussed in detail in the next section where we will show that the effective diffusivity is smaller than $D$ for gradient vector fields $b$ and that it is greater than $D$ for divergence–free vector fields $b$.

## 13.6 Applications

In this section we will consider two particular choices for the drift term $b$ in (13.2.4a), *gradient* and *divergence–free* fields. In both cases it is possible to perform explicit calculations which yield considerable insight. In particular, we will be able to obtain a formula for the (unique) invariant distribution and, consequently, to simplify the centering condition (13.2.8). Furthermore we will be able to compare the effective diffusivity with the original diffusivity $D$. We will see that the effective diffusivity is smaller than $D$ for gradient vector fields $b$, and that it is greater than $D$ for divergence–free vector fields $b$. We also study two particular cases of gradient and divergence–free flows for which we can derive closed formulae for the effective diffusivity.

There are at least two reasons why it is interesting to consider gradient and divergence–free flows. On the one hand, parabolic PDEs of the form (13.2.1) with $b$ being either the gradient of a scalar field or divergence–free appear frequently in applications: when $b = -\nabla V$ then equation (13.2.1) describes Brownian motion in a periodic potential. On the other hand, when $b$ is divergence–free equation (13.2.1) becomes the advection diffusion equation which describes mixing processes in incompressible fluids. According to the Hodge decomposition theorem, every smooth vector field on $\mathbb{T}^d$ can be decomposed into the sum of a gradient and a divergence–free field:

$$b(y) = -\nabla V(y) + v(y), \quad \nabla \cdot v(y) = 0,$$

with

$$(-\nabla V(y), v(y))_{L^2(\mathbb{T}^d)} = 0.$$

Hence, by studying gradient and divergence–free flows we study the two extreme cases of this decomposition.

### 13.6.1 Gradient Vector Fields

We consider the case where the vector field $b(y)$ in equation (13.2.4a) is the gradient of a smooth, scalar periodic function,

$$b(y) = -\nabla_y V(y). \tag{13.6.1}$$

The function $V$ is called the *potential*. In this case it is straightforward to derive a formula for the solution $\rho$ of the stationary adjoint equation (13.2.6) with periodic boundary conditions.

**Lemma 13.6.** *Assume that the vector field $b$ is a gradient given by* (13.6.1)*. Let $\mathcal{L}_0^*$ denote the adjoint of $\mathcal{L}_0$ defined in* (13.2.5)*. Then the equation*

$$\mathcal{L}_0^* \rho = 0, \quad \int_{\mathbb{T}^d} \rho(y) dy = 1, \tag{13.6.2}$$

*subject to periodic boundary conditions on $\mathbb{T}^d$ has a unique solution given by*

$$\rho(y) = \frac{1}{Z} e^{-V(y)/D}, \quad Z = \int_{\mathbb{T}^d} e^{-V(y)/D}\, dy. \tag{13.6.3}$$

*Proof.* Equation (13.6.2), in view of equation (13.6.1), becomes

$$\nabla_y \cdot \big( \nabla_y V(y) \rho(y) + D \nabla_y \rho(y) \big) = 0. \tag{13.6.4}$$

We immediately check that $\rho(y)$ given by (13.6.3) satisfies

$$\nabla_y V(y) \rho(y) + D \nabla_y \rho(y) = 0,$$

and hence it satisfies (13.6.4). Furthermore, by construction we have that

$$\int_{\mathbb{T}^d} \frac{1}{Z} e^{-V(y)/D}\, dy = 1,$$

and hence $\rho(y)$ is correctly normalized. Thus we have constructed a solution of equation (13.6.2). Uniqueness follows by the ergodicity of the stochastic process with generator $\mathcal{L}_0$ (see Theorem 6.16). □

*Remark 13.7.* The positive function $\rho$ defined in (13.6.3) is called the **Gibbs distribution** and the probability measure $\rho(y)dy$ the **Gibbs measure**. The normalization constant $Z$ is called the **partition function**. □

In the case of gradient flows the centering condition (13.2.8) is satisfied identically for any potential.

**Lemma 13.8.** *Consider the operator $\mathcal{L}_0$ given by* (13.2.5) *with periodic boundary conditions and assume that $b(y) = -\nabla_y V(y)$ with $V \in C^1_{per}(\mathbb{T}^d)$. Then the centering condition* (13.2.8) *is always satisfied.*

*Proof.* We use the divergence theorem to calculate

$$\begin{aligned} \int_{\mathbb{T}^d} b(y) \rho(y)\, dy &= \frac{1}{Z} \int_{\mathbb{T}^d} -\nabla_y V(y) e^{-V(y)/D}\, dy \\ &= \frac{D}{Z} \int_{\mathbb{T}^d} \nabla_y e^{-V(y)/D}\, dy \\ &= 0. \end{aligned}$$

□

In the case of gradient flows the operator $\mathcal{L}_0$ defined in (13.2.5) equipped with periodic boundary conditions becomes symmetric in the appropriate function space. We have the following.

**Lemma 13.9.** *Assume that condition* (13.6.1) *is satisfied and let* $\rho$ *denote the Gibbs distribution* (13.6.3)*. Then the operator* $\mathcal{L}_0$ *given in* (13.2.5) *satisfies*

$$\int_{\mathbb{T}^d} f(y)\Big(\mathcal{L}_0 h(y)\Big)\rho(y)\,dy = \int_{\mathbb{T}^d} h(y)\Big(\mathcal{L}_0 f(y)\Big)\rho(y)\,dy, \tag{13.6.5}$$

*for all* $f,\ h \in C^2_{per}(\mathbb{T}^d)$.

*Proof.* Using the divergence theorem we have

$$\begin{aligned}\int_{\mathbb{T}^d} f\mathcal{L}_0 h\rho\,dy &= \frac{1}{Z}\int_{\mathbb{T}^d} f\big(-\nabla_y V\cdot\nabla_y h\big)e^{-V/D}\,dy + \frac{D}{Z}\int_{\mathbb{T}^d} f\Delta_y h e^{-V/D}\,dy\\ &= \frac{D}{Z}\int_{\mathbb{T}^d} f\nabla_y h\cdot\nabla_y\left(e^{-V/D}\right)\,dy - \frac{D}{Z}\int_{\mathbb{T}^d}\left(\nabla_y f\cdot\nabla_y h\right)e^{-V/D}\,dy\\ &\quad -\frac{D}{Z}\int_{\mathbb{T}^d} f\nabla_y h\cdot\nabla_y\left(e^{-V/D}\right)\,dy\\ &= -D\int_{\mathbb{T}^d}\Big(\nabla_y f\cdot\nabla_y h\Big)\rho\,dy.\end{aligned}$$

The expression in the last line is symmetric in $f,\ h$ and hence (13.6.5) follows. □

*Remark 13.10.* The symmetry of $\mathcal{L}_0$ arises quite naturally from the identity (6.3.11) used in proving Theorem 6.12. Furthermore, the calculation used in the proof of the above lemma gives us the following useful formula

$$\int_{\mathbb{T}^d} f(-\mathcal{L}_0 h)\rho\,dy = D\int_{\mathbb{T}^d}\Big(\nabla_y f\cdot\nabla_y h\Big)\rho\,dy \tag{13.6.6}$$

for all $f,\ h \in C^2_{per}(\mathbb{T}^d)$. The Dirichlet form Lemma 13.3 follows from this upon setting $f = h$. Now let $\phi,\ \psi \in C^2_{per}(\mathbb{T}^d;\mathbb{R}^d)$. In view of (13.6.6) we also have

$$\int_{\mathbb{T}^d}\big(\phi\otimes(-\mathcal{L}_0\psi)\big)\rho\,dy = D\int_{\mathbb{T}^d}\Big(\nabla_y\phi\otimes\nabla_y\psi\Big)\rho\,dy \tag{13.6.7}$$

□

*Remark 13.11.* Using the notation introduced in Remark 13.4 we can express the result of Lemma 13.9 by saying that $\mathcal{L}_0$ is symmetric as an operator from $L^2_\rho$ to $L^2_\rho$. Furthermore, identity (13.6.6) can be written in the form

$$(f,-\mathcal{L}_0 h)_{L^2_\rho} = D\,(\nabla_y f,\nabla_y h)_{L^2_\rho}\,.$$

Ergodic Markov processes whose generator is a symmetric operator in $L^2_\rho$ are called **reversible** . Thus we have shown that SDEs with additive noise and with a drift which is a gradient field are reversible. □

Now we are ready to prove various properties of the effective diffusivity. For this we will need the following integration by parts formula, which follows from the divergence theorem and the periodicity of $\chi$ and $\rho$ :

$$\int_{\mathbb{T}^d} \big(\nabla_y \chi\big)\rho \, dy = \int_{\mathbb{T}^d} \big(\nabla_y(\chi\rho) - \chi \otimes \nabla_y \rho\big) \, dy = -\int_{\mathbb{T}^d} (\chi \otimes \nabla_y \rho) \, dy. \quad (13.6.8)$$

**Theorem 13.12.** *Assume that $b(y)$ is a gradient so that* (13.6.1) *holds and let $\rho(y)$ denote the Gibbs distribution* (13.6.3)*. Then the effective diffusivity* (13.3.2) *satisfies the upper and lower bounds*

$$\frac{D}{Z\widehat{Z}} \leqslant \langle \xi, \mathcal{K}\xi \rangle \leqslant D|\xi|^2 \quad \forall \xi \in \mathbb{R}^d, \quad (13.6.9)$$

*where*

$$\widehat{Z} = \int_{\mathbb{T}^d} e^{V(y)/D} \, dy.$$

*In particular, diffusion is always depleted when compared to molecular diffusivity. Furthermore, the effective diffusivity is symmetric.* [2]

*Proof.* The lower bound follows from the general lower bound (13.5.2), equation (13.5.3) and the formula for the Gibbs measure. To establish the upper bound, we use (13.6.8) and (13.6.7) to obtain

$$\begin{aligned}
\mathcal{K} &= DI + 2D \int_{\mathbb{T}^d} (\nabla \chi)^T \rho \, dy + \int_{\mathbb{T}^d} -\nabla_y V \otimes \chi \rho \, dy \\
&= DI - 2D \int_{\mathbb{T}^d} \nabla_y \rho \otimes \chi \, dy + \int_{\mathbb{T}^d} -\nabla_y V \otimes \chi \rho \, dy \\
&= DI - 2 \int_{\mathbb{T}^d} -\nabla_y V \otimes \chi \rho \, dy + \int_{\mathbb{T}^d} -\nabla_y V \otimes \chi \rho \, dy \\
&= DI - \int_{\mathbb{T}^d} -\nabla_y V \otimes \chi \rho \, dy \\
&= DI - \int_{\mathbb{T}^d} \big(-\mathcal{L}_0 \chi\big) \otimes \chi \rho \, dy \\
&= DI - D \int_{\mathbb{T}^d} \big(\nabla_y \chi \otimes \nabla_y \chi\big)\rho \, dy. \qquad (13.6.10)
\end{aligned}$$

Hence, for $\chi_\xi = \chi \cdot \xi$,

$$\begin{aligned}
\langle \xi, \mathcal{K}\xi \rangle &= D|\xi|^2 - D \int_{\mathbb{T}^d} |\nabla_y \chi_\xi|^2 \rho \, dy \\
&\leqslant D|\xi|^2.
\end{aligned}$$

This proves depletion. The symmetry of $\mathcal{K}$ follows from (13.6.10). □

[2] Notice that the Cauchy-Schwarz inequality shows that $Z\widehat{Z} \geqslant 1$.

**The One Dimensional Case**

The one dimensional case is always in gradient form: $b(y) = -\partial_y V(y)$. Furthermore in one dimension we can solve the cell problem (13.3.1) in closed form and calculate the effective diffusion coefficient explicitly–up to quadratures. We start with the following calculation concerning the structure of the diffusion coefficient.

$$\begin{aligned}\mathcal{K} &= D + 2D\int_0^1 \partial_y\chi\rho\,dy + \int_0^1 -\partial_y V\chi\rho\,dy \\ &= D + 2D\int_0^1 \partial_y\chi\rho\,dy + D\int_0^1 \chi\partial_y\rho\,dy \\ &= D + 2D\int_0^1 \partial_y\chi\rho\,dy - D\int_0^1 \partial_y\chi\rho\,dy \\ &= D\int_0^1 \big(1+\partial_y\chi\big)\rho\,dy. \end{aligned} \tag{13.6.11}$$

The cell problem (13.3.1) in one dimension is

$$D\partial_{yy}\chi - \partial_y V\partial_y\chi = \partial_y V. \tag{13.6.12}$$

We multiply equation (13.6.12) by $e^{-V(y)/D}$ to obtain

$$\partial_y\left(\partial_y\chi e^{-V(y)/D}\right) = -\partial_y\left(e^{-V(y)/D}\right).$$

We integrate this equation from 0 to $y$ and multiply by $e^{V(y)/D}$ to obtain

$$\partial_y\chi(y) = -1 + c_1 e^{V(y)/D}.$$

Another integration yields

$$\chi(y) = -y + c_1\int_0^y e^{V(y)/D}\,dy + c_2.$$

The periodic boundary conditions imply that $\chi(0) = \chi(1)$, from which we conclude that

$$-1 + c_1\int_0^1 e^{V(y)/D}\,dy = 0.$$

Hence

$$c_1 = \frac{1}{\widehat{Z}}, \quad \widehat{Z} = \int_0^1 e^{V(y)/D}\,dy.$$

We deduce that

$$\partial_y\chi = -1 + \frac{1}{\widehat{Z}}e^{V(y)/D}.$$

We substitute this expression into (13.6.11) to obtain

$$\begin{aligned}\mathcal{K} &= \frac{D}{Z}\int_0^1 (1+\partial_y\chi(y))\, e^{-V(y)/D}\, dy \\ &= \frac{D}{Z\widehat{Z}}\int_0^1 e^{V(y)/D} e^{-V(y)/D}\, dy \\ &= \frac{D}{Z\widehat{Z}}, \end{aligned} \tag{13.6.13}$$

with

$$Z = \int_0^1 e^{-V(y)/D}\, dy, \quad \widehat{Z} = \int_0^1 e^{V(y)/D}\, dy. \tag{13.6.14}$$

Notice that in the one–dimensional case the formula for the effective diffusivity is precisely the lower bound in (13.6.9). This shows that the lower bound is sharp.

*Example 13.13.* Consider the potential

$$V(y) = \begin{cases} a_1 & : \quad y \in [0, \frac{1}{2}], \\ a_2 & : \quad y \in (\frac{1}{2}, 1], \end{cases} \tag{13.6.15}$$

where $a_1,\ a_2$ are positive constants.[3]

It is straightforward to calculate the integrals in (13.6.14) to obtain the formula

$$\mathcal{K} = \frac{D}{\cosh^2\left(\frac{a_1-a_2}{D}\right)}. \tag{13.6.16}$$

In Figure 13.1 we plot the effective diffusivity given by (13.6.16) as a function of the molecular diffusivity $D$, on a log scale. We observe that $\mathcal{K}$ decays exponentially fast in the limit as $D \to 0$. $\square$

### 13.6.2 Divergence–Free Fields

In this section we consider the problem of homogenization for (13.2.4a) in the case where the vector field $b(y)$ is divergence–free (or incompressible):

$$\nabla \cdot b(y) = 0. \tag{13.6.17}$$

The incompressibility of $b(y)$ simplifies the analysis considerably because the advection operator

$$\widehat{\mathcal{L}}_0 = b(y) \cdot \nabla_y,$$

with periodic boundary conditions is antisymmetric in $L^2(\mathbb{T}^d)$:

[3] Of course, this potential is not even continuous, let alone smooth, and the theory as developed in this chapter does not apply. It is possible, however, to consider a regularized version of this discontinuous potential and then homogenization theory applies.

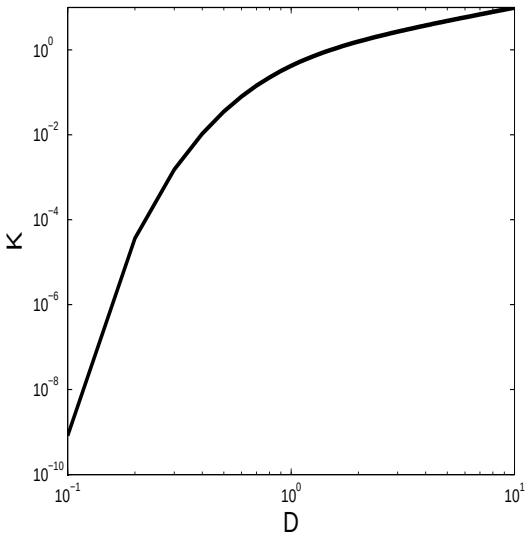


**Fig. 13.1.** Log-log plot of the effective diffusivity versus molecular diffusivity for the potential (13.6.15).

**Lemma 13.14.** *Let $b(y) \in C^1_{per}(\mathbb{T}^d;\mathbb{R}^d)$ satisfy* (13.6.17)*. Then for all $f(y),\ h(y) \in C^1_{per}(\mathbb{T}^d)$ we have*

$$\int_{\mathbb{T}^d} f(y)\,(b(y)\cdot\nabla_y h(y))\ dy = -\int_{\mathbb{T}^d} h(y)\,(b(y)\cdot\nabla_y f(y))\ dy.$$

*In particular,*

$$\int_{\mathbb{T}^d} f(y)\,(b(y)\cdot\nabla_y f(y))\ dy = 0. \tag{13.6.18}$$

*Proof.* We use the incompressibility of $b(y)$, together with the periodicity of $f(y)$, $h(y)$ and $b(y)$ to calculate

$$\begin{aligned}\int_{\mathbb{T}^d} f(y)\,(b(y)\cdot\nabla_y h(y))\ dy &= \int_{\mathbb{T}^d} f(y)\nabla_y\cdot\big(b(y)h(y)\big)\,dy\\ &= -\int_{\mathbb{T}^d} \nabla_y f(y)\cdot(b(y)h(y))\ dy\\ &= -\int_{\mathbb{T}^d} h(y)\big(b(y)\cdot\nabla_y f(y)\big)\,dy.\end{aligned}$$

Equation (13.6.18) follows from the above calculation upon setting $f = h$. □

Using the previous lemma it is easy to prove that the unique invariant measure of the fast process is the Lebesgue measure.

**Lemma 13.15.** *Let $\mathcal{L}_0$ denote the operator defined in* (13.2.5) *with periodic boundary conditions and with $b(y)$ satisfying* (13.6.17)*. Let $\mathcal{L}_0^*$ denote the $L^2$–adjoint of $\mathcal{L}_0$. Then the adjoint equation*

$$\mathcal{L}_0^* \rho = 0, \quad \int_{\mathbb{T}^d} \rho(y) dy = 1, \tag{13.6.19}$$

*with periodic boundary conditions on $\mathbb{T}^d$ has a unique classical solution given by*

$$\rho(y) = 1. \tag{13.6.20}$$

*Proof.* Lemma 13.14 implies that the $L^2$–adjoint of $\mathcal{L}_0$ is

$$\mathcal{L}_0^* = -b(y) \cdot \nabla_y + D\Delta_y, \tag{13.6.21}$$

with periodic boundary conditions. Let $\rho(y)$ be a solution of equation (13.6.19). We multiply the equation by $\rho(y)$, integrate over $\mathbb{T}^d$ and use Lemma 13.14 to obtain

$$\int_{\mathbb{T}^d} |\nabla_y \rho(y)|^2 \, dy = 0, \tag{13.6.22}$$

from which we deduce that $\rho(y)$ is a constant. Hence, the unique normalized solution of (13.6.19) is given by (13.6.20). □

*Remark 13.16.* The solution $\rho(y) = 1$ can be seen to be in the null space of (13.6.21) by inspection. Uniqueness can then be proved by appealing to ergodicity of the process with generator $\mathcal{L}_0$ (see Theorem 6.16), or by use of the maximum principle. □

*Remark 13.17.* An immediate corollary of Proposition 13.15 is that for divergence–free fields the solvability condition (13.2.8) becomes

$$\int_{\mathbb{T}^d} b(y) \, dy = 0.$$

Thus, it is straightforward to check whether a given periodic divergence–free field satisfies the solvability condition – the field must average to zero over the unit torus. □

Now let $\chi(y)$ be the solution of the cell problem (13.3.1) with $b(y)$ satisfying (13.6.17). The periodicity of $\chi(y)$, together with (13.6.20) imply that the second term on the right hand side of equation (13.3.2) vanishes and the formula for the effective diffusivity reduces to

$$\mathcal{K} = DI + \int_{\mathbb{T}^d} b(y) \otimes \chi(y) \, dy. \tag{13.6.23}$$

The effective diffusivity as given in (13.3.2) is symmetric for gradient flows. This is not true for divergence–free flows. However, only the symmetric part of $\mathcal{K}$ enters into

the homogenized equation by Remark 13.2. For this reason we redefine the effective diffusivity to be the symmetric part of $\mathcal{K}$:

$$\mathcal{K} := DI + \frac{1}{2}\int_{\mathbb{T}^d} \big(b(y)\otimes\chi(y) + \chi(y)\otimes b(y)\big)\,dy. \tag{13.6.24}$$

Our goal now is to show that the homogenization procedure enhances diffusion, i.e. that the effective diffusivity is always greater than the molecular diffusivity $D$. For this we will need an alternative representation formula for $\mathcal{K}$.

**Theorem 13.18.** *The effective diffusivity $\mathcal{K}$ given by the expression* (13.6.24) *can be written in the form*

$$\mathcal{K} = DI + D\int_{\mathbb{T}^d} \nabla_y\chi(y)\otimes\nabla_y\chi(y)\,dy. \tag{13.6.25}$$

*Proof.* We take the outer product of the cell problem (13.3.1) with $\chi(y)$ to the left and integrate over the unit cell to obtain

$$-D\int_{\mathbb{T}^d}\chi(y)\otimes\Delta_y\chi(y)\,dy - \int_{\mathbb{T}^d}\chi(y)\otimes\big(\nabla_y\chi(y)b(y)\big)\,dy = \int_{\mathbb{T}^d}\chi(y)\otimes b(y)\,dy.$$

We apply the divergence theorem to the *two integrals* on the left hand side of the above equation, using periodicity and the fact that $b$ is divergence–free, to obtain

$$D\int_{\mathbb{T}^d}\nabla_y\chi(y)\otimes\nabla_y\chi(y)\,dy + \int_{\mathbb{T}^d}\big(\nabla\chi(y)b(y)\big)\otimes\chi(y)\,dy = \int_{\mathbb{T}^d}\chi(y)\otimes b(y)\,dy. \tag{13.6.26}$$

Alternatively we may take the outer product with $\chi$ in (13.3.1) to the right and use the divergence theorem only on the first integral, to obtain

$$D\int_{\mathbb{T}^d}\nabla_y\chi(y)\otimes\nabla_y\chi(y)\,dy - \int_{\mathbb{T}^d}\big(\nabla\chi(y)b(y)\big)\otimes\chi(y)\,dy = \int_{\mathbb{T}^d} b(y)\otimes\chi(y)\,dy. \tag{13.6.27}$$

We add equations (13.6.26) and (13.6.27) to obtain:

$$\frac{1}{2}\int_{\mathbb{T}^d}\big(b(y)\otimes\chi(y)+\chi(y)\otimes b(y)\big)\,dy = D\int_{\mathbb{T}^d}\nabla_y\chi(y)\otimes\nabla_y\chi(y)\,dy.$$

Equation (13.6.25) now follows upon substituting the above expression into equation (13.6.24). □

We can now obtain upper and lower bounds for the effective diffusivity.

**Theorem 13.19.** *Assume that $b(y)$ is divergence–free. Then the effective diffusivity satisfies the upper and lower bounds*

$$D|\xi|^2 \leqslant \langle\xi,\mathcal{K}\xi\rangle \leqslant \left(D+\frac{C}{D}\right)|\xi|^2, \tag{13.6.28}$$

*where $C = C(b,\Omega) > 0$ is explicitly computable.* [4] *The lower bound becomes an equality for all $\xi$ only when $b(y)\equiv 0$.*

[4] Indeed $C = (C_p\|b\|_{L_2})^2$ where $C_p$ is the Poincaré constant from inequality (2.4.7).

*Proof.* The lower bound follows from the general bound (13.5.2), equation (13.5.3) and the fact that for divergence–free flows $\rho(y) = 1$. Furthermore, equation (13.6.25) implies that

$$\langle \xi, \mathcal{K}\xi \rangle := D|\xi|^2 + D \int_{\mathbb{T}^d} |\nabla_y \chi_\xi(y)|^2 \, dy, \tag{13.6.29}$$

where $\chi_\xi = \chi \cdot \xi$. Clearly the equality $\langle \xi, \mathcal{K}\xi \rangle = D|\xi|^2$ for all $\xi$ implies that $\chi_\xi = 0$ for all $\xi$ implying that $\chi(y) \equiv 0$. By (13.3.1) this implies that $b \equiv 0$.

For the upper bound we take the inner product of the cell problem with an arbitrary vector $\xi \in \mathbb{R}^d$ to obtain

$$-\mathcal{L}_0 \chi_\xi = b \cdot \xi.$$

We multiply this equation with $\chi_\xi$, integrate over $\mathbb{T}^d$, use Lemma 13.14 and the Poincaré inequality to calculate

$$\begin{aligned} D\|\nabla_y \chi^\xi\|_{L^2}^2 &= (-\mathcal{L}_0 \chi_\xi, \chi_\xi) = (b \cdot \xi, \chi_\xi) \\ &\leqslant \|b \cdot \xi\|_{L^2} \|\chi_\xi\|_{L^2} \\ &\leqslant C_p \|b\|_{L^2} \|\nabla_y \chi_\xi\|_{L^2} |\xi|, \end{aligned}$$

where $C_p$ is the Poincaré constant on $\mathbb{T}^d$. From the above estimate we deduce that

$$\|\nabla_y \chi_\xi\|_{L^2} \leqslant \frac{\sqrt{C}}{D} |\xi|$$

with $C = \left(C_p \|b\|_{L^2}\right)^2$. The result follows from (13.6.29). □

**Shear Flow in 2D**

In this section we study an example of a divergence–free flow for which the cell problem can be solved in closed form, that of a *shear flow*. The structure of a shear velocity field is such that the cell problem becomes an ordinary differential equation.

Let $y = (y_1, y_2)^T$. We consider the problem of homogenization for (13.2.4a) in two dimensions for the following velocity field:

$$b(y) = (0, b_2(y_1))^T, \tag{13.6.30}$$

where $b_2(y_1)$ is a smooth, 1–periodic function with mean zero. Notice that the velocity field (13.6.30) is incompressible:

$$\nabla \cdot b(y) = \frac{\partial b_1}{\partial y_1} + \frac{\partial b_2}{\partial y_2} = \frac{\partial b_2(y_1)}{\partial y_2} = 0.$$

The two components of the cell problem satisfy

$$-D\Delta_y \chi_1(y) - b_2(y_1) \frac{\partial \chi_1(y)}{\partial y_2} = 0, \tag{13.6.31a}$$

$$-D\Delta_y\chi_2(y) - b_2(y_1)\frac{\partial\chi_2(y)}{\partial y_2} = b_2(y_1), \tag{13.6.31b}$$

as well as periodicity and the normalization condition that $\chi$ integrates to zero over the unit cell $\mathcal{Y}$.

If we multiply the first equation (13.6.31a) by $\chi_1(y)$, integrate by parts over $\mathbb{T}^d$ then we deduce that

$$\int_{\mathbb{T}^d} |\nabla_y\chi_1(y)|^2\, dy = 0.$$

Hence $\chi_1(y) = 0$, since we impose the normalization $\langle\chi(y)\rangle = 0$ with $\langle\cdot\rangle := \int_{\mathbb{T}^d}\cdot dy$. On the other hand, since the right hand side of (13.6.31b) depends only on $y_1$, it is reasonable to assume that the solution $\chi_2(y)$ is independent of $y_2$; we seek a solution of this form and then, provided that we can find such a solution, uniqueness of solutions to the cell problem implies that it is the only solution. Equation (13.6.31b) becomes:

$$-D\frac{d^2\chi_2(y_1)}{dy_1^2} = b_2(y_1). \tag{13.6.32}$$

If $\psi$ is a periodic solution to

$$-\frac{d^2\psi(y_1)}{dy_1^2} = b_2(y_1), \tag{13.6.33}$$

then $\psi$ is independent of $D$ and $\chi_2 = \psi/D$.

By (13.6.24) the effective diffusivity $\mathcal{K}$ is the following $2\times 2$ matrix:

$$\begin{aligned}\mathcal{K} &= \begin{pmatrix} D + \int_{\mathbb{T}^2}(b_1\chi_1)\, dy & \frac{1}{2}\int_{\mathbb{T}^2}(b_2\chi_1 + b_1\chi_2)\, dy \\ \frac{1}{2}\int_{\mathbb{T}^2}(b_2\chi_1 + b_1\chi_2)\, dy & D + \int_{\mathbb{T}^2}(b_2\chi_2)\, dy \end{pmatrix} \\ &= \begin{pmatrix} D & 0 \\ 0 & \mathcal{K}_{22}\end{pmatrix},\end{aligned}$$

where we have used the fact that $b_1 = \chi_1 = 0$. Using the fact that $b_2$, $\chi_2$ depend only on $y_1$ we obtain

$$\begin{aligned}\mathcal{K}_{22} &:= D + \int_0^1 b_2\chi_2 dy_1 \\ &= D + \int_0^1 -D\frac{d^2\chi_2}{dy_1^2}\chi_2 dy_1 \\ &= D + D\int_0^1 \Big|\frac{d\chi_2}{dy_1}\Big|^2 dy_1 \\ &= D + \frac{1}{D}\int_0^1 \Big|\frac{d\psi}{dy_1}\Big|^2 dy_1.\end{aligned}$$

Notice the remarkable fact that, since $\psi$ is independent of $D$, the formula shows that the effective diffusion coefficient scales as $D^{-1}$ as the original molecular diffusion coefficient $D$ tends to zero. This demonstrates that the upper bound in Theorem

13.19 is sharp. The intuition behind this scaling is that, for small $D$, the equation is approximately a transport equation in the direction $x_2$. The direction of transport is slowly modulated, leading to overall diffusive behaviour, but on long time-scales the predominant effect is transport. This enhances the diffusivity.

It is possible to express $\psi$ as an integral operator acting on $b_2$ and show that

$$\mathcal{K}_{22} = D + \frac{1}{D}\|b_2\|^2_{H^{-1}_{per}(0,1)}. \tag{13.6.34}$$

See Exercise 10.

*Example 13.20.* Consider the case

$$b_2(y_1) = \sin(2\pi y_1). \tag{13.6.35}$$

We use formula (13.6.34) and Exercise 10 to obtain

$$\mathcal{K}_{22} = D + \frac{1}{8\pi^2 D}. \tag{13.6.36}$$

In Figure 13.2 we plot the effective diffusivity given by (13.6.36) as a function of

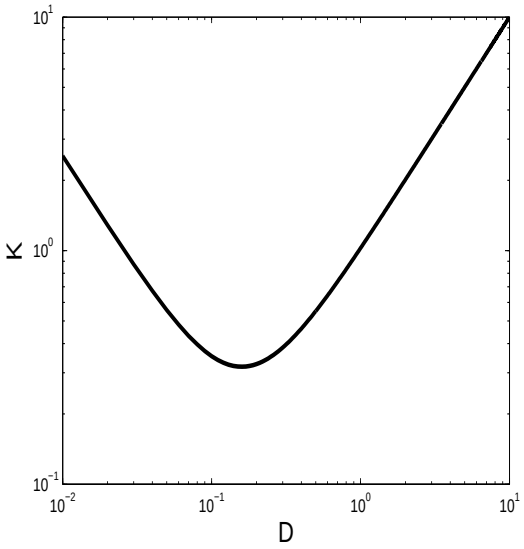


**Fig. 13.2.** Log-log plot of the effective diffusivity versus molecular diffusivity for the sine shear flow (13.6.35).

the molecular diffusivity $D$, on a log scale. We observe that $\mathcal{K}$ diverges like $\frac{1}{D}$ in the limit as $D \to 0$. □

## 13.7 The Connection to SDEs

Equation (13.2.1) is the backward Kolmogorov equation associated with the SDE

$$\frac{dx}{dt} = b(x) + \sqrt{2D}\,\frac{dW}{dt}, \tag{13.7.1}$$

where $W$ denotes standard Brownian motion on $\mathbb{R}^d$. Unsurprisingly, then, the homogenization results derived in this chapter have implications for the behavior of solutions to this SDE. To see this we first apply the rescaling used to derive (13.2.4) from (13.2.1) to the SDE (13.7.1). That is, we relabel according to

$$x \to x/\varepsilon, \quad t \to t/\varepsilon^2$$

giving the SDE

$$\frac{dx}{dt} = \frac{1}{\varepsilon} b\Big(\frac{x}{\varepsilon}\Big) + \sqrt{2D}\frac{dW}{dt}. \tag{13.7.2}$$

(Recall Remark 6.3 regarding the behavior of white noise under time rescaling).

If we introduce the variable $y = x/\varepsilon$ then we can write this SDE in the form

$$\begin{aligned}
\frac{dx}{dt} &= \frac{1}{\varepsilon} b(y) + \sqrt{2D}\frac{dW}{dt}, \\
\frac{dy}{dt} &= \frac{1}{\varepsilon^2} b(y) + \frac{1}{\varepsilon}\sqrt{2D}\frac{dW}{dt}.
\end{aligned}$$

Here we view $x$ as being an element of $\mathbb{R}^d$ whilst $y$ is on the torus $\mathbb{T}^d$. This is very similar to the form (11.2.1) which we analyzed in Chapter 11. The only difference is that the noises appearing in the $x$ and $y$ equations are *correlated* (in fact $U = V = W$). This has the effect of changing the operator $\mathcal{L}_1$ in that chapter, so that the results derived there do not apply directly. They can, however, be readily extended to the study of correlated noise – see Chapter 11, Exercises 5 and 1. Notice that the centering condition (13.2.8) is precisely the condition (11.2.5) since $\rho$ is the stationary solution of the same Fokker-Planck equation.

The calculations in this chapter show how the backward Kolmogorov equation for the coupled SDE in $(x, y)$ can be approximated by a diffusion equation in the $x$ variable alone. Indeed, the diffusion equation is the backward Kolmogorov equation for pure Brownian motion. Interpreted in terms of the SDE we obtain the following result.

**Result 13.21.** *Assume that the centering condition* (13.2.8) *holds. For* $\varepsilon \ll 1$ *and* $t = \mathcal{O}(1)$, $x$ *solving the SDE* (13.7.2) *can be approximated by* $X$ *solving*

$$\frac{dX}{dt} = \sqrt{(\mathcal{K} + \mathcal{K}^T)}\frac{dW}{dt}$$

*where the matrix* $\mathcal{K}$ *is given by* (13.3.2).

If the centering condition is not satisfied then the appropriate rescaling of (13.7.1) is an advective one, leading to the equations (14.6.1) considered in the next chapter.

## 13.8 Discussion and Bibliography

The problem of homogenization for second order parabolic PDEs and its connection to the study of the long time asymptotics of solutions of SDEs is studied in [33, Ch. 3]. References to the earlier literature can be found there. See also [238]. SDEs of the form (13.7.1), whose drift is the gradient of a periodic scalar function, describe Brownian motion in periodic potentials. This a very important problem in many applications, for example in solid state physics and biology. See [271, Ch. 11], [267] and the references therein. Multiscale techniques were applied to this problem in [257]. Periodic homogenization for gradient flows is also discussed in [238, 256, 323, 118]. Formula (13.6.13) for the effective diffusivity of a Brownian particle moving in a one dimensional periodic potential was derived in [191] without any appeal to homogenization theory. See also [138, Sec. VII]. Brownian motion in a two–scale periodic potential in one dimension is studied in [342]. The multidimensional problem is analyzed in [258].

On the other hand, the SDE (13.7.1) with divergence–free drift occurs naturally in the modeling of diffusion processes in fluids. Homogenization for periodic, incompressible flows is a part of the theory of *turbulent diffusion* [200, 99]. See also [221, 100, 101]. In this context an interesting question concerns the dependence of the effective diffusivity on the molecular diffusion $D$. It turns out that the small $D$–asymptotics of the effective diffusivity depends sensitively on the streamline topology of the fluid velocity field $b(y)$. See [63, 294, 295, 62, 140, 20, 22, 21]. Interesting experimental results concerning the dependence of the effective diffusivity on $D$ or, rather, on the *Peclet number* $Pe$ are reported in [293, 292]; rescaling enables these results to be interpreted in terms of molecular diffusivity. Homogenization for compressible flows with applications to atmospheric transport phenomena is studied in [223].

It is possible to derive a homogenized equation even when the centering condition (13.2.8) is not satisfied. In this case it is necessary to use a frame co–moving with the *mean flow*

$$\overline{b} = \int_{\mathbb{T}^d} b(y)\rho(y)\,dy. \tag{13.8.1}$$

Then, it is possible to derive a homogenized equation of the form (13.3.3) for the rescaled field

$$u^\varepsilon(x,t) = u\left(\frac{x}{\varepsilon} - \frac{\overline{b}t}{\varepsilon^2}, \frac{t}{\varepsilon^2}\right).$$

The effective diffusivity is given by the formula

$$\mathcal{K} = DI + 2D\int_{\mathbb{T}^d} \nabla_y \chi(y)^T \rho(y)\,dy + \int_{\mathbb{T}^d} \big(b(y) - \overline{b}\big) \otimes \chi(y)\rho(y)\,dy, \tag{13.8.2}$$

The cell problem (13.3.1) is also modified:

$$-\mathcal{L}_0\chi = b - \overline{b}. \tag{13.8.3}$$

See Exercise 5 in Chapter 14.

The mean flow $\overline{b}$ can have a dramatic effect in the small $D$ asymptotics of the effective diffusivity for periodic divergence–free flows; in particular, the scaling of $\mathcal{K}$ with $D$ for $D \ll 1$ depends on whether the mean flow is a rational or irrational vector. See [201, 222, 35, 295, 175]. A similar discontinuous dependence of the effective diffusivity on the wavelengths of the inhomogeneities was observed for gradient flows in [126].

It is proved in Section 13.6.1 that for gradient flows the diffusion is always depleted. In fact, much sharper results can be obtained: the effective diffusivity is "exponentially" smaller than $D$, for $D$ sufficiently small. That is, there exist positive constants $c_1$ and $c_2$ such that

$$\langle \xi, \mathcal{K}\xi \rangle = c_1 e^{-c_2/D}, \quad D \ll 1.$$

See [54] and the references therein. On the other hand, the effective diffusion coefficient can become arbitrarily large, when compared to the molecular diffusivity, when a constant external force is added to the gradient drift, see [268, 282].

The fact that the effective diffusivity along the direction of the shear is inversely proportional to the molecular diffusivity, formula (13.6.34), was discovered in [313], without any appeal to homogenization theory. This phenomenon is often refered to as *Taylor dispersion*. See also [11]. A similar result for time dependent periodic shear flows was obtained in [340] through a direct calculation with the advection–diffusion equation.

To derive the expression (13.6.34) for the effective diffusion coefficient (from Exercise 10) it is necessary to use formal calculations with Fourier series. Of course, we have to prove that we can differentiate the Fourier series and that the Fourier series that we get for the second derivative of $\chi(y)$ makes sense. For various properties of Fourier series we refer the reader to [132, Ch. 3].

We showed that the effective diffusion tensor is symmetric for gradient flows. The effective diffusivity, however, is not necessarily symmetric for general vector fields. Despite the fact that the antisymmetric part of the effective diffusivity does not affect the homogenized equation, it is of physical significance: it gives rise to a component of the flux which is perpendicular to the concentration gradient, [174]. Whereas the effective diffusivity is symmetric or not depends on the symmetry properties of the underlying vector field $b(y)$.[5] This issue is studied for divergence–free flows in [174, 253]; in those references the dependence of the antisymmetric part of the effective diffusivity on the Peclet number is also studied.

In addition to the Eulerian definition of the effective diffusivity giving rise to the effectivge diffusion tensor $\mathcal{K}$ we can also define a Lagrangian effective diffusivity through the long time average of the variance of the underlying stochastic process $x(t)$.

$$D^L_{eff} := \lim_{t\to\infty} \frac{\langle ((x(t) - \langle x(t)\rangle) \otimes (x(t) - \langle x(t)\rangle)\rangle}{2t}. \tag{13.8.4}$$

[5] For example, in the case of gradient flows the effective diffusivity is symmetric because of the reversibility (which, of course, is a symmetry property) of gradient flows.

Notice that $D^L_{eff}$ is a symmetric tensor. It is straightforward to show that the the Lagrangian effective diffusivity (13.8.4) agrees with the *symmetric part* of the Eulerian effective diffusivity.

The method of multiple scales can also be used to study the problem of homogenization for parabolic PDEs with time dependent coefficients which are periodic in both $x$ and $t$. See, e.g. [118, 228, 323, 257, 42].

Monte Carlo methods for advection diffusion and for transport PDEs are presented in [187]. Numerical methods for advection diffusion equations with a multiscale structure are developed in [1].

## 13.9 Exercises

1. Derive a formula for $u_2(x, x/\varepsilon, t)$, the third term in the expansion (13.4.1).
2. Consider the problem of homogenization for
$$\frac{\partial u^\varepsilon}{\partial t} = -\frac{1}{\varepsilon}\nabla V\left(\frac{x}{\varepsilon}\right)\cdot\nabla u^\varepsilon + D\Delta u^\varepsilon$$
in one dimension with the (1–periodic) potential
$$V(y) = \begin{cases} y & : \ y \in [0, \frac{1}{2}], \\ 1-y & : \ y \in (\frac{1}{2}, 1], \end{cases}$$
Calculate the effective diffusivity $\mathcal{K}$. Use Laplace's method to study the small $D$ asymptotics of $\mathcal{K}$.
3. Carry out the program from the previous exercise for the potential $V(y) = \sin(2\pi y)$. (Hint: use Bessel functions).
4. Calculate the effective diffusivity (13.3.2) for the 2–dimensional vector field $b(y) = (b_1(y_1), b_2(y_1))$.
5. Consider the problem of homogenization for the reaction–advection–diffusion equation
$$\frac{\partial u^\varepsilon}{\partial t} = \frac{1}{\varepsilon}b\left(\frac{x}{\varepsilon}\right)\cdot\nabla u^\varepsilon + \Delta u^\varepsilon + \frac{1}{\varepsilon}c\left(\frac{x}{\varepsilon}\right)u^\varepsilon, \tag{13.9.1}$$
where the vector field $b(y)$ and the scalar function $c(y)$ are smooth and periodic. Use the method of multiple scales to homogenize the above PDE. In particular:
   a) Derive the solvability condition.
   b) Obtain the conditions that $b(y)$ and $c(y)$ should satisfy so that you can derive the homogenized equation.
   c) Derive the homogenized equation, the cell problem(s) and the formula for the homogenized coefficients.
   d) Suppose that the reaction term is nonlinear: the zeroth order term in equation (13.9.1) is replaced by
$$c\left(\frac{x}{\varepsilon}, u^\varepsilon\right),$$
where the function $c(y, u)$ is 1–periodic in $y$ for every $u$. Can you homogenize equation. (13.9.1) in this case?

6. Consider the problem of homogenization for the PDE

$$\frac{\partial u^\varepsilon}{\partial t} = \left( b_1(x) + \frac{1}{\varepsilon} b_2\left(\frac{x}{\varepsilon}\right)\right) \cdot \nabla u^\varepsilon + \Delta u^\varepsilon, \tag{13.9.2}$$

where the vector field $b_2(y)$ is smooth and periodic and $b_1(x)$ is periodic. Use the method of multiple scales to homogenize the above PDE. In particular:

a) Derive the solvability condition.
b) Obtain the conditions that $b_2(y)$ should satisfy so that you can derive the homogenized equation.
c) Show that the homogenized equation is

$$\frac{\partial u}{\partial t} = b \cdot \nabla u + \mathcal{K} : \nabla\nabla u \tag{13.9.3}$$

and derive the cell problem(s) and the formulae for the homogenized coefficients $b$ and $\mathcal{K}$.

7. Consider the problem of homogenization for the PDE (13.9.2) in the case where

$$b_1(x) = -\nabla V(x) \quad \text{and} \quad b_2(y) = -\nabla p(y),$$

where $p(y)$ is periodic.

a) Show that in this case there exists a symmetric matrix $\widehat{\mathcal{K}}$ such that

$$\mathcal{K} = D\widehat{\mathcal{K}}, \quad B = -\widehat{\mathcal{K}}\nabla V.$$

b) Let

$$\mathcal{L} := b \cdot \nabla + \mathcal{K} : \nabla\nabla u.$$

1. Derive a formula for $\mathcal{L}^*$, the $L^2$–adjoint of $\mathcal{L}$.
2. Show that the function

$$\rho(y) := \frac{1}{Z} e^{-V(y)/D}, \quad Z = \int_{\mathbb{T}^d} e^{-V(y)/D}\, dy$$

solves the homogeneous adjoint equation

$$\mathcal{L}^*\rho = 0.$$

8. Consider the problem of homogenization for the following PDE

$$\frac{\partial u^\varepsilon}{\partial t} = b^\varepsilon \cdot \nabla u^\varepsilon + A^\varepsilon : \nabla_x \nabla_x u^\varepsilon$$

where $A^\varepsilon = A(x/\varepsilon)$, $b^\varepsilon = b(x/\varepsilon)$ and the vector field $b(y)$ and the matrix $A(y)$ are smooth and periodic, and $A(y)$ is positive definite. Use the method of multiple scales to derive the homogenized equation. In particular:

a) Derive the solvability condition.
b) Obtain conditions on $b(y)$ which ensure the existence of a homogenized equation.

c) Derive the homogenized equation, the cell problem and the formula for the homogenized coefficients.
d) Prove that the homogenized matrix is positive definite.

9. Consider the problem of homogenization for the following PDE

$$\frac{\partial u^\varepsilon}{\partial t} = \frac{1}{\varepsilon} b\left(\frac{x}{\varepsilon}, \frac{t}{\varepsilon^2}\right) \cdot \nabla u^\varepsilon + D\Delta u^\varepsilon,$$

where the vector field $b(y,\tau)$ is smooth, divergence free and 1–periodic in both $y$ and $\tau$. Use the method of multiple scales to derive the homogenized equation. In particular:
a) Derive the solvability condition.
b) Obtain conditions on $b(y,\tau)$ which ensure the existence of a homogenized equation.
c) Derive the homogenized equation, the cell problem and the formula for the homogenized coefficients.
d) Prove that the homogenized matrix is positive definite.

10. The $H^{-1}_{per}$ norm of a real–valued, periodic function with period 1 can be expressed in terms of Fourier series (see the discussion in Section 2.7) as follows:

$$\|f\|^2_{H^{-1}_{per}(0,1)} = \frac{1}{2\pi^2} \sum_{k=1}^{\infty} \frac{|f_k|^2}{|k|^2}.$$

Use this definition, and the Fourier series representation of the solution $\psi$ of the problem 13.6.31b, to establish formula (13.6.34) from the expression for $\mathcal{K}_{22}$ in terms of $\psi$.

11. Consider Exercise 9 in dimension $d = 2$ and with the velocity field

$$b(y_1, y_2, \tau) = (0, b_2(y_1, \tau)).$$

Derive a formula for the effective diffusivity $\mathcal{K}$. How does $\mathcal{K}$ depend on $D$?

12. Repeat the calculations of Section 13.6.2 and Exercise 11 for the 2d velocity fields

$$b(y_1, y_2) = (V, b(y_1))$$

and

$$b(y_1, y_2, \tau) = (V, b(y_1, \tau)),$$

respectively, where $V \in \mathbb{R}$. (Hint: you need to use equations (13.8.2) and (13.8.3) and their generalizations for time dependent velocity fields).

13. Let $b(y)$ be a smooth, real valued 1–periodic, mean zero function and let $\{b_k\}_{k=-\infty}^{+\infty}$ be its Fourier coefficients. Prove that $b_0 = 0$ and that $b_{-k} = \overline{b}_k$.

# 14

# Averaging for Linear Transport and Parabolic PDEs

## 14.1 Introduction

In this chapter we investigate the long time behavior of solutions to the linear transport (or advection) equation, and to the parabolic (advection–diffusion) equation from the previous chapter, when the centering condition is not satisfied. The techniques we employ are sometimes referred to as homogenization techniques in the literature. However in terms of the classification in Section 1.3 the methods are actually averaging methods. We use this terminology.

In Sections 14.2 and 14.3 we set up the problem of interest and then state the approximation result. Section 14.4 contains the derivation of the averaged equation, when the starting point is a parabolic equation. Section 14.5 is devoted to the case where the averaging is for a pure transport equation; the issues here are more subtle (no Fredholm alternative for the leading order linear operator) and this is why we devote a separate section to it. In Section 14.6 we make the connection to averaging for ODEs and SDEs. Section 14.7 contains bibliographical remarks.

## 14.2 Full Equations

We study the long time behavior of solutions to the linear parabolic equation corresponding to advection–diffusion in a steady periodic velocity field $b$:

$$\frac{\partial u}{\partial t} = b \cdot \nabla u + D\Delta u \quad \text{for } (x,t) \in \mathbb{R}^d \times \mathbb{R}^+, \tag{14.2.1a}$$

$$u = u_{in} \quad \text{for } (x,t) \in \mathbb{R}^d \times \{0\}. \tag{14.2.1b}$$

This is the parabolic equation (13.2.1) and, in the case $D = 0$, it reduces to a linear transport equation. As in Chapter 13 we study the case where

$$u_{in}(x) = g(\varepsilon x),$$

and rescale the equation in both space and time in order to understand the behavior of solutions to equation (14.2.1) at length and time scales which are long when compared to those of the velocity field $b(x)$. In this setting, the small parameter in the problem is the ratio between the characteristic length (time) scale of the velocity field – its period – and the largest length (time) scale of the problem – the one at which we are looking for an averaged description. In contrast to the analysis of the advection–diffusion equation in the previous chapter, we rescale time and space in the same fashion, namely

$$x \to \varepsilon^{-1} x, \quad t \to \varepsilon^{-1} t. \tag{14.2.2}$$

In the parabolic case $D > 0$ this is because we do not assume that the centering condition (13.2.8) holds; thus the advective effects do not average out. Such a transformation is also natural in the case $D = 0$ since the transport PDE (14.2.1a) is then of first order in both space and time.

The initial value problem that we wish to investigate becomes:

$$\frac{\partial u^\varepsilon}{\partial t} = b^\varepsilon \cdot \nabla u^\varepsilon + \varepsilon D \Delta u^\varepsilon \quad \text{for } (x,t) \in \mathbb{R}^d \times \mathbb{R}^+, \tag{14.2.3a}$$

$$u^\varepsilon = f \quad \text{for } x \in \mathbb{R}^d \times \{0\}. \tag{14.2.3b}$$

Here $b^\varepsilon(x) = b(x/\varepsilon)$, as in the previous chapter.

As in the previous chapter we define the operator

$$\mathcal{L}_0 = b(y) \cdot \nabla_y + D \Delta_y \tag{14.2.4}$$

with periodic boundary conditions. Note that constants in $y$ are in the null space of this operator; furthermore, for $D > 0$ the null space is one dimensional and comprises only constants. The $L^2$–adjoint of $\mathcal{L}_0$ is $\mathcal{L}_0^*$, also with periodic boundary conditions. Recall from Chapter 13 that, for $D > 0$, the **invariant distribution** $\rho(y)$ is the unique stationary solution of the adjoint equation

$$\mathcal{L}_0^* \rho = 0, \quad \int_{\mathbb{T}^d} \rho(y)\, dy = 1, \tag{14.2.5}$$

equipped with periodic boundary conditions. For $D > 0$ both operators $\mathcal{L}_0$ and $\mathcal{L}^*$ satisfy a Fredholm alternative.

## 14.3 Simplified Equations

In this and the following section we simply assume that the operator $\mathcal{L}_0$ has a one dimensional null space, comprising constants; and that the same holds for its adjoint $\mathcal{L}_0^*$, with null space spanned by $\rho$. This follows from the Fredholm alternative for $D > 0$. For $D = 0$ it requires some form of ergodicity of the underlying ODE for which $\mathcal{L}_0$ is the generator. We discuss this ergodicity issue in Sections 14.5 and 14.6.

Under the stated assumptions on $\mathcal{L}_0$ we have the following result:

**Result 14.1.** *Let $b$ be a smooth periodic vector field. Assume that the operator $\mathcal{L}_0$ defined in* (14.2.4) *satisfies*

$$\mathcal{N}(\mathcal{L}_0) = span(1), \quad \mathcal{N}(\mathcal{L}_0^*) = span(\rho).$$

*Then, for $\varepsilon \ll 1$ and times $t$ of $\mathcal{O}(1)$, the solution $u^\varepsilon(x,t)$ of* (13.2.4) *is approximated by $u(x,t)$, the solution of the averaged equation:*

$$\frac{\partial u}{\partial t} - \overline{b} \cdot \nabla_x u = 0, \quad \overline{b} := \int_{\mathbb{T}^d} \rho(y) b(y)\, dy,$$

*together with the same initial condition as for $u^\varepsilon$.*

The calculations leading to this approximation result take the rescaled parabolic equation (14.2.3a) as starting point and recover a transport equation by means of averaging. Naively it might appear that the diffusion term in (14.2.3a) simply disappears from the averaging calculation, since it is multiplied by $\varepsilon$. *This viewpoint is wrong*: the diffusion coefficient plays an essential role. In general the form of the stationary distribution, against which $b$ is averaged, depends crucially on $D > 0$, through $\rho$.[1]

Note that the centering condition (13.2.8) simply states that $\overline{b} = 0$. This is why a different scaling of space and time is used in Chapter 13 from that used here: specifically a longer timescale is used there, in order to see nonnegligible effects.

## 14.4 Derivation

We use the method of multiple scales as introduced in the two preceding chapters. We introduce the auxiliary variable $y = x/\varepsilon$. Let $\phi = \phi(x, x/\varepsilon)$ be scalar-valued. The chain rule gives

$$\nabla\phi = \nabla_x\phi + \frac{1}{\varepsilon}\nabla_y\phi \quad \text{and} \quad \Delta\phi = \Delta_x\phi + \frac{2}{\varepsilon}\nabla_x \cdot \nabla_y\phi + \frac{1}{\varepsilon^2}\Delta_y\phi.$$

The partial differential operator that appears on the right hand side of equation (14.2.3) has the form

$$\mathcal{L} = \frac{1}{\varepsilon}\mathcal{L}_0 + \mathcal{L}_1 + \varepsilon\mathcal{L}_2,$$

where

$$\begin{aligned}
\mathcal{L}_0 &= b(y) \cdot \nabla_y + D\Delta_y, \\
\mathcal{L}_1 &= b(y) \cdot \nabla_x + 2D\nabla_x \cdot \nabla_y, \\
\mathcal{L}_2 &= D\Delta_x.
\end{aligned}$$

In terms of $x$ and $y$ equation (14.2.3) becomes

[1] An exception is the case divergence–free flows: the invariant measure $\rho$ is the Lebesgue measure on the unit torus for all $D > 0$. See Proposition 13.15.

$$\frac{\partial u^\varepsilon}{\partial t} = \left(\frac{1}{\varepsilon}\mathcal{L}_0 + \mathcal{L}_1 + \varepsilon\mathcal{L}_2\right) u^\varepsilon.$$

We look for a solution in the form of a two-scale expansion:

$$u^\varepsilon(x,t) = u_0\left(x, \frac{x}{\varepsilon}, t\right) + \varepsilon u_1\left(x, \frac{x}{\varepsilon}, t\right) + \dots. \tag{14.4.1}$$

We assume that all terms in the expansion $u_j(x,y,t),\ j = 0,1,\dots$ are 1–periodic in $y$ and treat $x$ and $y := \frac{x}{\varepsilon}$ as independent variables. [2] We substitute (14.4.1) into equation (14.2.3a), use the assumed independence of $x$ and $y$ and collect equal powers of $\varepsilon$ to obtain the following set of equations:

$$\mathcal{O}(1/\varepsilon) \quad -\mathcal{L}_0 u_0 = 0, \tag{14.4.2a}$$

$$\mathcal{O}(1) \quad -\mathcal{L}_0 u_1 = \mathcal{L}_1 u_0 - \frac{\partial u_0}{\partial t}, \tag{14.4.2b}$$

where $u_j(x,y)$ is $1-$periodic in $y$.

We can now complete the averaging procedure. From the first equation in (14.4.2), and our assumptions on $\mathcal{L}_0$, we deduce that the first term in the expansion is independent of the oscillations which are expressed through the auxiliary variable $y$:

$$u_0 = u(x,t).$$

We use this to compute:

$$\mathcal{L}_1 u_0 = \frac{\partial u(x,t)}{\partial t} - b(y)\cdot\nabla_x u(x,t).$$

Since $\rho$ is in the null space of $\mathcal{L}_0^*$ the second equation in (14.4.2) implies that

$$0 = \frac{\partial u(x,t)}{\partial t} - \left(\int_{\mathbb{T}^d} \rho(y) b(y)\, dy\right)\cdot\nabla_x u(x,t). \tag{14.4.3}$$

We have thus obtained the desired averaged equation:

$$\frac{\partial u(x,t)}{\partial t} - \overline{b}\cdot\nabla_x u(x,t) = 0, \quad \overline{b} := \int_{\mathbb{T}^d} \rho(y) b(y)\, dy,$$

together with the same initial conditions as those for $u^\varepsilon$.

## 14.5 Transport Equations: $D = 0$

We have indicated that, in general, the averaged transport equation depends subtly on the diffusion coefficient $D$ through the invariant distribution $\rho$ against which $b$

[2] As in the elliptic and parabolic homogenization procedures in the previous two chapters, this is where we exploit scale separation: **we treat $x$ and $y$ as independent variables**. Justifying this assumption as $\varepsilon \to 0$ is one of the main issues in the rigorous theory of averaging. See Chapter 21.

is averaged. Existence and uniqueness of the stationary distribution $\rho$ is automatic when $D > 0$ but requires some form of ergodicity, which will depend upon the properties of $b$, when $D = 0$. It is therefore a nontrivial question to ask when, and to what extent, the preceding averaging calculations extend to the case $D = 0$. The calculations rely on the null spaces of $\mathcal{L}_0$ and $\mathcal{L}_0^*$ being one dimensional, something ensured by the Fredholm alternative in the case $D > 0$. We discuss the analogues of these results in the case $D = 0$.

Let

$$\mathcal{L}_0 = b(y) \cdot \nabla_y \tag{14.5.1}$$

with domain $C^1_{per}(\mathbb{T}^d)$. We can extend this operator to $\mathcal{D}(\mathcal{L}_0) \subset L^\infty_{per}(\mathbb{T}^d)$ as in (4.3.8). We assume for the moment that there are no nontrivial functions in the null space $\mathcal{N}$ of $\mathcal{L}_0$:

$$\mathcal{N}(\mathcal{L}_0) = \{\text{constants in } y\} \tag{14.5.2}$$

viewing the operator as acting on $\mathcal{D}(\mathcal{L}_0)$. From Chapter 4 we know that this is essentially an ergodicity assumption on the ODE with vector field $b$ – see Theorem 4.13(iii). In relation to this, the idea that $\mathcal{L}_0^*$ is nonempty with domain viewed as being $C^1_{per}(\mathbb{T}^d)$, implies the existence of an invariant measure which is absolutely continuous with respect to the Lebesgue measure – see Theorem 4.12(iii). Thus ergodicity with respect to absolutely continuous invariant measure $\mu$ provides us with the necessary tools to carry out the formal perturbation expansions of this chapter in the case $D = 0$. In particular, in the ergodic case, (14.4.2a) implies that $u_0$ is independent of $y$ and also that a necessary condition for a solution $u_1$ of (14.4.2b) to exist is the equation (14.4.3).

Note that if $b$ is divergence–free (the velocity field is incompressible) then $\mathcal{L}$ is skew–symmetric (Lemma 13.14) and so we deduce from (14.5.2) that

$$\mathcal{N}(\mathcal{L}_0^*) = \{\text{constants in } y\}. \tag{14.5.3}$$

However, in the general ergodic case, $\rho$ will not be a constant function.

### 14.5.1 The One–Dimensional Case

Consider the rescaled transport equation (14.2.3a) in one dimension:

$$\frac{\partial u^\varepsilon}{\partial t} - b^\varepsilon \frac{\partial u^\varepsilon}{\partial x} = 0 \quad \text{for } (x,t) \in \mathbb{R} \times \mathbb{R}^+, \tag{14.5.4a}$$

$$u = g \quad \text{for } (x,t) \in \mathbb{R} \times \{0\}, \tag{14.5.4b}$$

where $g = g(x)$ is independent of the oscillations.[3] We assume that $b(y)$ is a strictly positive, smooth, 1–periodic function. The stationary Liouville equation

$$\mathcal{L}_0^* \rho = 0, \quad \rho > 0, \ 1\text{–periodic}, \tag{14.5.5}$$

together with the normalization condition

[3] This is not necessary – see Exercise 3 from Chapter 21.

$$\int_0^1 \rho(y)\,dy = 1,$$

has unique normalized solution the probability density

$$\rho(y) = \frac{C}{b(y)}, \quad C = \langle b(y)^{-1}\rangle^{-1}; \tag{14.5.6}$$

here we have used the notation $\langle\cdot\rangle$ to denote averaging over $[0,1]$, as in Chapter 12. Positivity of $b$ is key to this existence and uniqueness result, and also to the ergodicity of the underlying flow. These issues are discussed in Example 4.14.

We obtain the averaged equation

$$\frac{\partial u}{\partial t} - \overline{b}\frac{\partial u}{\partial x} = 0, \tag{14.5.7}$$

with the same initial conditions as in (14.5.4b) and with

$$\overline{b} = \langle b(y)^{-1}\rangle^{-1}.$$

Notice that, in contrast to the ergodic divergence–free case presented in the next subsection, it is the harmonic average of the velocity field that appears in the averaged equation (14.5.7) rather than the standard average. (Note that the harmonic average also arises in the one–dimensional elliptic case – see Subsection 12.6.1).

### 14.5.2 Divergence–Free Velocity Fields

If $b$ is divergence–free (the velocity field is incompressible) then $\mathcal{L}$ given by (14.5.1) is skew–symmetric (Lemma 13.14) and so we deduce that, if (14.5.2) holds, then

$$\mathcal{N}(\mathcal{L}_0^*) = \big\{\text{constants in } y\big\}. \tag{14.5.8}$$

(See Example 4.15). Unfortunately, even for divergence–free fields, the ergodic hypothesis leading to (14.5.2) is often not satisfied. Consider an equation in the form (14.4.2a):

$$\mathcal{L}_0 u := b(y)\cdot\nabla_y u = 0 \tag{14.5.9}$$

with periodic boundary conditions. Although $u \equiv 1$ solves this equation, it is rare that this solution is unique: the null space of the operator $\mathcal{L}_0$ contains, in general, non-trivial functions of $y$. As an example, consider the smooth, 1–periodic, divergence–free field

$$b(y) = (\sin(2\pi y_2),\, \sin(2\pi y_1)).$$

It is easy to check that the function

$$u(y) = \cos(2\pi y_1) - \cos(2\pi y_2)$$

solves equation (14.5.9). Consequently, the null space of $\mathcal{L}_0$ depends on the velocity field $b(y)$ and it does not consist, in general, merely of constants in $y$. This implies that we cannot carry out the averaging procedure using the method of multiple scales.

It is natural to ask whether there is a way of deciding whether a given divergence–free velocity field on $\mathbb{T}^d$ is ergodic or not. This is indeed possible in two dimensions. A result along these lines is the following.

**Theorem 14.2.** *Let $b(y) : \mathbb{T}^2 \to \mathbb{R}^2$ be a smooth divergence–free velocity field satisfying*

$$b_1(y) \neq 0 \quad \forall y \in \mathbb{T}^2$$

*so that it has no stagnation points. Let $\overline{b}_i$, $i = 1, 2$ denote the average of the $i$th component of the velocity field over $\mathbb{T}^2$ and define the rotation number as*

$$\gamma = \frac{\overline{b}_1}{\overline{b}_2}.$$

*Then there exists a smooth change of variables $y \mapsto z$ under which the ODEs*

$$\frac{dy_1}{dt} = b_1(y), \quad \frac{dy_2}{dt} = b_2(y) \tag{14.5.10}$$

*transform into*

$$\frac{dz_1}{dt} = g(z), \quad \frac{dz_2}{dt} = \gamma g(z) \tag{14.5.11}$$

*where $g(z)$ is a nonvanishing smooth scalar function. Assume furthermore that $\gamma$ is irrational. Then the null space of the generator $\mathcal{L}_0$ is one–dimensional.*

*Proof.* The first part of the theorem can be proved by constructing explicitly the transformation that maps (14.5.10) into (14.5.11):[4]

$$z_1 = \frac{1}{\overline{b}_2} \int_0^{y_1} b_2(\xi, 0)\, d\xi, \quad z_2 = \frac{1}{\overline{b}_1} \int_0^{y_2} b_1(y_1, \xi)\, d\xi.$$

The second part of the theorem can be proved using Fourier analysis. See Exercise 7. □

Thus, under the conditions of this theorem, Theorem 4.13 holds and the formal perturbation expansions of this chapter may be applied.

## 14.6 The Connection to ODEs and SDEs

We consider first the case $D = 0$. Recall from Chapter 4 that the solution of (14.2.3) is given by

$$u(x, t) = g(\varphi^t(x)),$$

where $\varphi^t(x)$ solves the ODE

$$\frac{d}{dt}\varphi^t(x) = b\Big(\frac{\varphi^t(x)}{\varepsilon}\Big),$$
$$\varphi^t(x) = x.$$

Result 14.1 shows that, when the ergodicity assumption holds so that $\mathcal{L}_0$ has one-dimensional null space, this equation is well approximated by

[4] Under the additional assumption that $b_2 \neq 0$ which can be removed later.

$$\overline{\varphi}^t(x) = \overline{b}t + x,$$

the solution of

$$\begin{aligned} \frac{d}{dt}\overline{\varphi}^t(x) &= \overline{b}, \\ \overline{\varphi}^t(x) &= x. \end{aligned}$$

Here

$$\overline{b} = \int_{\mathbb{T}^d} \rho(y) b(y) dy = \langle b(y)^{-1} \rangle$$

by (14.5.6).

Another way to see this result is as follows. Let $x = \varphi^t(x_0)$ and $y = x/\varepsilon$. Then

$$\begin{aligned} \frac{dx}{dt} &= b(y), \\ \frac{dy}{dt} &= \frac{1}{\varepsilon} b(y). \end{aligned}$$

Under the ergodic hypothesis the fast process $y$ has invariant measure $\rho$ on the torus $\mathbb{T}^d$. Thus the averaging Result 10.1 gives that $x$ is well approximated by the solution of the equation

$$\frac{dX}{dt} = \overline{b}.$$

This is precisely the approximation derived above.

*Example 14.3.* In the one-dimensional case it is possible to derive the averaged equation (14.5.7) using the method of characteristics. To see this, consider the equation

$$\frac{dx}{dt} = b\Big(\frac{x}{\varepsilon}\Big)$$

in one dimension, and under the same assumptions as before. If we set $y = x/\varepsilon$ then it is straightforward to show that

$$\frac{dy}{dt} = \frac{1}{\varepsilon} b(y),$$

so that, if we define $T$ by

$$T = \int_0^1 \frac{1}{b(z)} dz = \frac{1}{\overline{b}},$$

then

$$y(n\varepsilon T) = \frac{x(0)}{\varepsilon} + n.$$

Hence

$$x(n\varepsilon T) = x(0) + n\varepsilon.$$

It follows from continuity that $x(t)$ converges to $X(t)$ where

$$X(t) = x(0) + \frac{t}{T}.$$

This limiting function $X(t)$ satisfies the homogenized equation

$$\frac{dX}{dt} = \frac{1}{T} = \overline{b}. \quad \square$$

If $D > 0$ then equation (14.2.3) is the backward Kolmogorov equation for the SDE

$$\frac{dx}{dt} = b\Big(\frac{x}{\varepsilon}\Big) + \sqrt{2\varepsilon D}\frac{dW}{dt}.$$

Another way to interpret the averaging result is thus as follows. Let $y = x/\varepsilon$ to obtain

$$\begin{aligned} \frac{dx}{dt} &= b(y) + \sqrt{2\varepsilon D}\frac{dW}{dt}, \\ \frac{dy}{dt} &= \frac{1}{\varepsilon}b(y) + \sqrt{\frac{2D}{\varepsilon}}\frac{dW}{dt}. \end{aligned} \tag{14.6.1}$$

Under the ergodic hypothesis the fast process $y$ has invariant measure $\rho$ on the torus $\mathbb{T}^d$. Thus a generalization of the averaging Result 10.1 gives that $x$ is well approximated by the ODE

$$\frac{dx}{dt} = \overline{b}.$$

## 14.7 Discussion and Bibliography

The perturbation expansion used here is analogous to that used in the method of averaging, for Markov chains, ODE and SDE, in Chapters 9 and 10. The problem of averaging for linear transport equations has been studied by many authors. See for example [80, 147, 312, 51]. Averaging for SDEs is studied in detail in [111].

When $D = 0$ the method of multiple scales enables us to obtain the averaged linear transport equation (14.2.3a) only in the case where the velocity field is ergodic. The method of multiple scales breaks down when the velocity field is not ergodic, since in this case we do not have a solvability condition which would enable us to average. In fact, when the velocity field is not ergodic, the $\varepsilon \to 0$ limit becomes much more complicated and the limiting process cannot be expressed through a simple PDE. In order to study the problem for general velocity fields, not necessarily ergodic, it is possible to use the method of two–scale convergence. This will be done in Chapter 21.

Theorem 14.2 is proved in [312], where the result is proved for two-dimensional flows that have a smooth invariant density, not only divergence–free flows (for which the invariant density is simply 1). It is not the sharpest result that one can prove, but the assumption $b_1 \neq 0$ leads to a particularly simple proof. The proof of the analogous theorem under the assumptions that there are no stagnation points can be

found in [290]. A similar theorem holds for velocity fields with an invariant measure other than the Lebesgue measure on $\mathbb{T}^2$; see [312].

The example studied in Section 14.5.1 can be found in [81, 312]. Monte Carlo methods for advection–diffusion and for transport PDEs are presented in [187].

## 14.8 Exercises

1. How does the dynamics of the ODE studied in Section 14.5.1 change if $b$ is allowed to change sign?
2. Consider the equation
$$\frac{dx}{dt} = a\Big(\frac{x}{\varepsilon}\Big)b\Big(\frac{t}{\varepsilon^\alpha}\Big)$$
in one dimension, and under the assumption that $a$ (resp. $b$) is smooth, $1-$periodic and $\inf_x a > 0$ (resp. $\inf_y b > 0$). Find the averaged equations.
3. Study the problem of averaging for (14.2.3) with a smooth periodic (shear) velocity field $b : \mathbb{T}^2 \mapsto \mathbb{R}^2$ of the form
$$b(y) = (0, b_2(y_1))^T.$$
4. Study the problem of averaging for (14.2.3) with a velocity field $b : \mathbb{T}^2 \mapsto \mathbb{R}^2$ of the form
$$b(y) = \widehat{b}(y)(0,\gamma)^T,$$
where $\widehat{b}(y)$ is a smooth, 1–periodic scalar function and $\gamma \in \mathbb{R}$.
5. Consider equation (13.2.4) in the case where the centering condition (13.2.8) does not hold. Show that it is possible to derive a homogenized equation of the form (13.3.3) for the rescaled field
$$u^\varepsilon(x,t) = u\left(\frac{x}{\varepsilon} - \frac{\overline{b}t}{\varepsilon^2}, \frac{t}{\varepsilon^2}\right)$$
where $u$ solves (14.2.1) and $\overline{b}$ is given in Result 14.1. Show that the cell-problem becomes
$$-\mathcal{L}_0\chi = b - \overline{b}. \tag{14.8.1}$$
and that the effective diffusivity is given by the formula
$$\mathcal{K} = DI + 2D\int_{\mathbb{T}^d} \nabla_y\chi(y)^T\rho(y)\,dy + \int_{\mathbb{T}^d}\Big(\big(b(y)-\overline{b}\big)\otimes\chi(y)\rho(y)\Big)\,dy. \tag{14.8.2}$$
6. Study the problem of homogenization for the ODE
$$\frac{dx}{dt} = -\nabla V\Big(\frac{x}{\varepsilon}\Big) + F$$
where $V(y)$ is a smooth periodic function and $F$ is a constant vector.
7. Complete the details in the proof of Theorem 14.2.

# Part III

## Theory