AN INTRODUCTION TO MULTISCALE METHODS

G.A. Pavliotis
Department of Mathematics
Imperial College London

and

A.M. Stuart
Mathematics Institute
University of Warwick

June 25, 2006
To my parents Ἀργυρὴ and Σουλτανα and to my brother Γιώργο.

Carry Home. Γρηγόρης.

For my children Natalie, Sebastian and Isobel.

AMS.
## Contents

<table>
<thead>
<tr>
<th>Preface</th>
<th>xi</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Motivating Examples</td>
<td>1</td>
</tr>
<tr>
<td>1.1.1 Example I: Steady Heat Conduction in a Composite Material</td>
<td>1</td>
</tr>
<tr>
<td>1.1.2 Example II: Homogenization for Advection–Diffusion Equations</td>
<td>3</td>
</tr>
<tr>
<td>1.1.3 Example III: Averaging, Homogenization and Dynamics</td>
<td>4</td>
</tr>
<tr>
<td>1.1.4 Example IV: Dimension Reduction in Dynamical Systems</td>
<td>5</td>
</tr>
<tr>
<td>1.2 Averaging Versus Homogenization</td>
<td>6</td>
</tr>
<tr>
<td>1.3 Discussion and Bibliography</td>
<td>6</td>
</tr>
</tbody>
</table>

## I Background | 9 |

<table>
<thead>
<tr>
<th>2 Analysis</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Set–Up</td>
<td>11</td>
</tr>
<tr>
<td>2.2 Notation</td>
<td>12</td>
</tr>
<tr>
<td>2.3 Banach and Hilbert Spaces</td>
<td>13</td>
</tr>
<tr>
<td>2.3.1 Banach Spaces</td>
<td>14</td>
</tr>
<tr>
<td>2.3.2 Hilbert Spaces</td>
<td>15</td>
</tr>
<tr>
<td>2.4 Function Spaces</td>
<td>16</td>
</tr>
<tr>
<td>2.4.1 $L^p$ Spaces</td>
<td>17</td>
</tr>
<tr>
<td>2.4.2 Sobolev Spaces</td>
<td>18</td>
</tr>
<tr>
<td>2.4.3 Banach Space–Valued Spaces</td>
<td>20</td>
</tr>
<tr>
<td>2.4.4 Sobolev spaces of Periodic Functions</td>
<td>21</td>
</tr>
<tr>
<td>2.5 Two–Scale Convergence</td>
<td>23</td>
</tr>
</tbody>
</table>
CONTENTS

2.5.1 Two–Scale convergence for steady problems . . . . . . . 23
2.5.2 Two–Scale convergence for time dependent problems . . . 26
2.6 Equations in Hilbert Spaces . . . . . . . . . . . . . . . . . . . . 27
  2.6.1 Lax-Milgram Theory . . . . . . . . . . . . . . . . . . . . 28
  2.6.2 Fredholm Alternative . . . . . . . . . . . . . . . . . . . . 29
2.7 Discussion and Bibliography . . . . . . . . . . . . . . . . . . . 30
2.8 Exercises . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 32

3 Probability Theory and Stochastic Processes 35
  3.1 Set–Up . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 35
  3.2 Probability and Expectation . . . . . . . . . . . . . . . . . . . 35
  3.3 Stochastic Processes . . . . . . . . . . . . . . . . . . . . . . . . 38
  3.4 Martingales and Stochastic Integrals . . . . . . . . . . . . . . . 42
    3.4.1 Martingales . . . . . . . . . . . . . . . . . . . . . . . . . 42
    3.4.2 The Itô Stochastic Integral . . . . . . . . . . . . . . . . . 44
    3.4.3 The Stratonovich Stochastic Integral . . . . . . . . . . . . 45
  3.5 Weak Convergence of Probability Measures . . . . . . . . . . . 46
  3.6 Discussion and Bibliography . . . . . . . . . . . . . . . . . . . 53
  3.7 Exercises . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 54

4 Ordinary Differential Equations 55
  4.1 Set-Up . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 55
  4.2 Existence and Uniqueness . . . . . . . . . . . . . . . . . . . . . 55
  4.3 The Generator . . . . . . . . . . . . . . . . . . . . . . . . . . . . 57
  4.4 Ergodicity . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 60
  4.5 Discussion and Bibliography . . . . . . . . . . . . . . . . . . . 61
  4.6 Exercises . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 62

5 Markov Chains 63
  5.1 Set-Up . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 63
  5.2 The Generator . . . . . . . . . . . . . . . . . . . . . . . . . . . . 65
  5.3 Existence and Uniqueness . . . . . . . . . . . . . . . . . . . . . 68
  5.4 Ergodicity . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 70
  5.5 Discussion and Bibliography . . . . . . . . . . . . . . . . . . . 71
  5.6 Exercises . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 72
CONTENTS

6 Stochastic Differential Equations 73
  6.1 Set-Up ......................................................... 73
  6.2 Existence and Uniqueness ................................. 74
  6.3 The Generator ................................................ 75
  6.4 Ergodicity ...................................................... 78
  6.5 Discussion and Bibliography ............................. 82
  6.6 Exercises ...................................................... 83

7 Partial Differential Equations 85
  7.1 Set-Up ........................................................... 85
  7.2 Elliptic PDE .................................................... 85
  7.3 The Dirichlet Problem for Elliptic PDE .................... 86
  7.4 The Periodic Problem for Elliptic PDE ..................... 88
  7.5 Fredholm alternative ........................................ 89
  7.6 Parabolic PDE .................................................. 95
  7.7 Hyperbolic PDE ............................................... 96
  7.8 Discussion and Bibliography ............................. 97
  7.9 Exercises ...................................................... 98

II Perturbation Expansions 99

8 Invariant Manifolds for ODE 101
  8.1 Introduction ................................................... 101
  8.2 Full Equations ............................................... 101
  8.3 Simplified Equations ...................................... 102
  8.4 Derivation ..................................................... 103
  8.5 Applications .................................................. 104
     8.5.1 Linear Fast Dynamics .................................. 104
     8.5.2 Large Time Dynamics ................................ 104
     8.5.3 Centre Manifold ....................................... 105
  8.6 Discussion and Bibliography ............................. 106
  8.7 Exercises ...................................................... 106

9 Averaging for Markov Chains 109
  9.1 Introduction ................................................... 109
  9.2 Full Equations ............................................... 109
CONTENTS

9.3 Simplified Equations .......................... 111
9.4 Derivation .................................... 112
9.5 Application .................................. 113
9.6 Discussion and Bibliography ................. 114

10 Averaging for ODE and SDE .................. 115
10.1 Introduction ................................. 115
10.2 Full Equations ............................... 115
10.3 Simplified Equations ......................... 117
10.4 Derivation .................................. 118
10.5 Applications ................................. 119
    10.5.1 A Skew-Product SDE .................... 119
    10.5.2 Hamiltonian Mechanics ................. 120
10.6 Discussion and Bibliography ................. 123
10.7 Exercises .................................. 125

11 Homogenization for ODE and SDE ............ 127
11.1 Introduction ................................ 127
11.2 Full Equations ............................... 127
11.3 Simplified equations ......................... 129
11.4 Derivation .................................. 132
11.5 Properties of the Effective Equation ...... 134
11.6 Applications ................................. 135
    11.6.1 Fast Ornstein-Uhlenbeck Noise ......... 135
    11.6.2 Fast Chaotic Noise ..................... 137
    11.6.3 Stratonovich Corrections ............... 138
    11.6.4 Stokes’ Law ............................ 140
    11.6.5 Kubo Formula ........................... 141
    11.6.6 Neither Itô nor Stratonovich .......... 143
    11.6.7 The Levy Area Correction ............. 145
11.7 Discussion and Bibliography ................. 146
11.8 Exercises .................................. 147

12 Homogenization for Elliptic Equations ...... 149
12.1 Introduction ................................. 149
12.2 Full Equations ............................... 149
12.3 Simplified Equations ......................... 150
CONTENTS

12.4 Derivation ........................................... 151
12.5 Properties of the Homogenized Coefficients ........ 154
12.6 Applications ........................................ 157
  12.6.1 The One–Dimensional Case ...................... 158
  12.6.2 Layered Materials ................................ 159
12.7 Discussion and Bibliography ......................... 162
12.8 Exercises .......................................... 166

13 Homogenization for Parabolic Equations ............. 171
  13.1 Introduction ..................................... 171
  13.2 Full Equations ................................... 171
  13.3 Simplified Equations ................................ 173
  13.4 Derivation of the Homogenized equation .......... 173
  13.5 Effective Diffusivity .............................. 175
  13.6 Applications ....................................... 176
    13.6.1 Gradient Vector Fields ......................... 176
    13.6.2 The One Dimensional Case ..................... 179
    13.6.3 Divergence–Free Fields ......................... 181
    13.6.4 Shear Flow in 2D .............................. 184
  13.7 The Connection to SDE ............................. 188
  13.8 Discussion and Bibliography ........................ 189
  13.9 Exercises .......................................... 191

14 Homogenization for Linear Transport PDE .......... 195
  14.1 Introduction ...................................... 195
  14.2 Full Equations ................................... 195
  14.3 Simplified Equations ................................ 196
  14.4 Derivation ........................................ 197
  14.5 The Connection to ODE ............................. 198
  14.6 The One–Dimensional Case ......................... 199
  14.7 The Two–Dimensional Case. Kolmogorov’s Theorem ... 201
  14.8 Discussion and Bibliography ........................ 202
  14.9 Exercises .......................................... 203
### III Theory

#### 15 Invariant Manifolds for ODE: The Convergence Theorem
- 15.1 The Theorem ........................................ 207
- 15.2 The Proof ........................................ 209
- 15.3 Discussion and Bibliography ..................... 210
- 15.4 Exercises ........................................ 211

#### 16 Averaging for Markov Chains: The Convergence Theorem
- 16.1 The Theorem ........................................ 213
- 16.2 The Proof ........................................ 214
- 16.3 Discussion and Bibliography ..................... 215
- 16.4 Exercises ........................................ 216

#### 17 Averaging for SDE: The Convergence Theorem
- 17.1 The Theorem ........................................ 217
- 17.2 The Proof ........................................ 218
- 17.3 Discussion and Bibliography ..................... 220
- 17.4 Exercises ........................................ 221

#### 18 Homogenization for SDE: The Convergence Theorem
- 18.1 Theorem ........................................... 223
- 18.2 The Proof ........................................ 225
- 18.3 Discussion and Bibliography ..................... 229
- 18.4 Exercises ........................................ 229

#### 19 Homogenization for Elliptic PDE: The Convergence Theorem
- 19.1 Theorems .......................................... 231
- 19.2 Proof of the Homogenization Theorem ........... 232
- 19.3 Proof of the Corrector Result .................... 236
- 19.4 Discussion and Bibliography ..................... 237
- 19.5 Exercises ........................................ 238

#### 20 Homogenization for Parabolic PDE: The Convergence Theorem
- 20.1 Introduction ....................................... 239
- 20.2 The Theorem ....................................... 239
- 20.3 The Proof ........................................ 240
CONTENTS

20.4 Discussion and Bibliography ........................................... 243
20.5 Exercises ................................................................. 244

21 Homogenization for Transport PDE: The Convergence Theorem 247
21.1 Introduction ............................................................. 247
21.2 The Theorem ............................................................ 247
21.3 The Proof ................................................................. 248
21.4 Discussion and Bibliography ......................................... 250
21.5 Exercises ................................................................. 250
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. 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 (PDE), stochastic differential equations (SDE), ordinary differential equations (ODE) 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 theoretical side of the subject, and we devote a significant fraction of the book to this background material.

The first section 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 cover this material in order. A natural way to get an overview of the subject is to read through the section on Perturbation Expansions, referring back to the Background material as needed. ¹ The Theory can then be

¹This material is necessarily terse and the Discussion and Bibliography section at the end of each chapter provide pointers to more detailed literature.
studied, after the form of the approximations is understood, on a case by case basis.

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 PDE, SDE, ODE 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.

These notes are meant to be an introduction, aimed primarily towards graduate students in mathematics, applied mathematics and physics. This is particularly true for the third part of the book, where we analyze rigorously only simplified versions of the models that were studied in the second part of the book. Extensions and generalizations of the results presented these notes as well as references to the literature are presented in the Discussion and Bibliography section, at the end of each chapter. All chapters are supplemented with exercises. We hope that the exercises will make these notes more appropriate for use as a textbook or for self study.

Warning
This is a draft set of notes. It has not been checked and refined to the standard required of a published book. The authors would welcome comments, remarks and suggestions by all readers, from typos through to major errors and structural deficiencies. All comments can be sent to the authors via e-mail (at g.pavliotis@imperial.ac.uk and stuart@maths.warwick.ac.uk). Updated versions of the notes will be available from the authors’ web pages: http://www.ma.ic.ac.uk/~pavl and http://www.maths.warwick.ac.uk/~stuart.

Acknowledgements
We are grateful to Dror Givon, Martin Hairer, Raz Kupferman, David White and Konstantinos Zygalakis for many helpful comments which substantially im-
proved these lecture notes. The book grew out of courses on multiscale methods and homogenization that were taught by the authors at Warwick University and Imperial College London. We thank the students who attended these courses.
Chapter 1

Introduction

1.1 Motivating Examples

In this section we describe four examples which illustrate the range of problems that we will study in the notes. The first illustrates homogenization in the context of linear elliptic PDE. The second illustrates related ideas in the context of time-dependent PDE of advection-diffusion type. Through the connection between hyperbolic (transport) PDE and ODE via the method of characteristics, and parabolic PDE and SDE via Itô’s formula, we show that the methods of homogenization developed for the study of linear PDE can also be applied to study dimension reduction in ODE and SDE. We then finish with an example concerning variable elimination in dynamical systems.

1.1.1 Example I: Steady Heat Conduction in a Composite Material

To introduce ideas related to homogenization we consider the problem of steady heat conduction in a composite material whose material properties vary rapidly on the macroscopic scale. For simplicity we assume that the heterogenieties are periodic in space. If $\Omega \subset \mathbb{R}^d$ denotes the domain occupied by the material, then the size of the domain defines the macroscopic length scale $L$. On the other hand, the period of the heterogeneities defines the microscopic length scale $\varepsilon$ of the problem. Assuming that $L = O(1)$ and that the size of heterogeneities is small $\varepsilon \ll 1$, then the phenomenon of steady heat conduction can be described by the following
elliptic boundary value problem\(^1\):
\[
- \nabla \cdot \left(A \left(\frac{x}{\varepsilon}\right) \nabla u^\varepsilon\right) = f, \text{ for } x \in \Omega, \quad (1.1.1a)
\]
\[
u^\varepsilon(x) = 0, \text{ for } x \in \partial \Omega. \quad (1.1.1b)
\]

The matrix \(A(y)\) is the thermal conductivity tensor and \(u^\varepsilon(x)\) denotes the temperature field. The purpose of homogenization theory is to study the limit of this equation as \(\varepsilon \to 0\). From a physical point of view, this limit corresponds to the case where the heterogeneities become vanishingly small. In other words, we aim to replace the original, highly heterogeneous material characterized by the coefficients \(A \left(\frac{x}{\varepsilon}\right)\) with an effective, homogeneous material which is characterized by constant coefficients \(\overline{A}\). Hence the name homogenization.

In later chapters we will show that, under appropriate assumptions on the matrix \(A(y)\), the function \(f(x)\) and the domain \(\Omega\) the homogenized equation is
\[
- \overline{\mathbf{A}} : \nabla \nabla u = f, \text{ for } x \in \Omega, \quad (1.1.2a)
\]
\[
u(x) = 0, \text{ for } x \in \partial \Omega. \quad (1.1.2b)
\]
The constant homogenized coefficients \(\overline{\mathbf{A}}\) are given by the formula:
\[
\overline{\mathbf{A}} = \int_Y A(y) \left(I + \nabla \chi_T(y)\right) \, dy. \quad (1.1.3)
\]
The (first order) corrector \(\chi(y)\) solves the cell problem
\[
- \nabla_y \cdot \left(A(y) \nabla \chi_T(y)\right) = \nabla A^T(y), \quad \chi(y) \text{ is 1-periodic.} \quad (1.1.4)
\]
Thus, the calculation of the effective coefficients involves the solution of a partial differential equation posed on the unit cell (i.e. with periodic boundary conditions), together with the computation of the integrals in (1.1.3). Notice that finding the homogenized solution \(u\) requires the solution of two elliptic PDE: the periodic cell problem (1.1.4), which allows construction of \(\overline{\mathbf{A}}\) given by (1.1.3), and the Dirichlet problem (1.1.2). The important point is that these elliptic equations do not depend upon the small scale \(\varepsilon\) and hence their numerical solution is far less time-consuming than is direct numerical solution of (1.1.1).

As well as deriving the homogenized equations by use of perturbation expansions, we will also prove that the solution \(u^\varepsilon(x)\) of (1.1.1) converges to the solution \(u(x)\) of the homogenized equation (1.1.2) as \(\varepsilon \to 0\), in some appropriate sense. The homogenized equation will be derived using asymptotic expansions in Chapter 12. The rigorous homogenization theorem will be proved in Chapter 19.

\(^1\)The notation that we will be using in these notes is explained in Section 2.2
1.1. MOTIVATING EXAMPLES

1.1.2 Example II: Homogenization for Advection–Diffusion Equations

We now show how the ideas of homogenization can be used to study time-dependent parabolic PDE. Consider some physical quantity which is immersed in a fluid, for example a chemical pollutant in the atmosphere. Under the assumption that this quantity does not affect the fluid velocity field \( v(x, t) \), its concentration field \( T(x, t) \) satisfies the advection-diffusion equation

\[
\frac{\partial T}{\partial t} + v(x, t) \cdot \nabla T = D \Delta T, \tag{1.1.5a}
\]

\[
T(x, t = 0) = T_0(x). \tag{1.1.5b}
\]

Let us assume that the fluid velocity is smooth, steady, \( v(x, t) = -b(x) \), incompressible \( \nabla \cdot b(x) = 0 \), and periodic with period 1 in all directions. Imagine that \( T_0(x) = f(\varepsilon x) \) so that initially the concentration is slowly varying in space. It is then reasonable to expect that the concentration will only vary significantly on large length and time scales. Homogenization techniques enable us to show that the re–scaled concentration field \( T(x/\varepsilon, t/\varepsilon^2) \) – scaled so as to bring out large length and timescales – converges, as \( \varepsilon \) tends to 0, to the solution of the heat equation

\[
\frac{\partial T}{\partial t} = \mathcal{K} : \nabla \nabla T. \tag{1.1.6}
\]

Here \( \mathcal{K} \) denotes effective diffusion tensor

\[
\mathcal{K} = DI + \int_{\mathbb{T}^d} b(y) \otimes \chi(y) \, dy.
\]

The corrector field \( \chi(y) \) solves the cell problem

\[
- D \Delta_y \chi(y) - b(y) \cdot \nabla_y \chi(y) = b(y),
\]

with periodic boundary conditions.

The comments made in Example I apply equally well here: finding the homogenized field requires solution of two PDE (one periodic elliptic, the other parabolic) which are independent of \( \varepsilon \), and hence amenable to numerical solution. Furthermore these ideas can be made rigorous and error estimates found. The homogenized equation will be derived in Chapter 13. The rigorous homogenization theorem will be proved in Chapter 20.
1.1.3 Example III: Averaging, Homogenization and Dynamics

Under the rescaling described above, namely $x \to x/\varepsilon$ and $t \to t/\varepsilon^2$, the equation (1.1.5) becomes, for steady velocity fields,

$$\frac{\partial T}{\partial t} - \frac{1}{\varepsilon} b\left(\frac{x}{\varepsilon}\right) \cdot \nabla T = D \Delta T,$$

(1.1.6a)

$$T(x, t = 0) = f(x).$$

(1.1.6b)

In the case $D = 0$ this equation can be solved by the method of characteristics; the characteristics are obtained by solving the ODE

$$\frac{dx^\varepsilon}{dt} = b\left(\frac{x^\varepsilon}{\varepsilon}\right).$$

Since $b(x)$ is periodic, and $x/\varepsilon$ varies rapidly on the scale of the period, it is natural to try and average the equation to eliminate these rapid oscillations. Thus we see that eliminating fast scales in a time-dependent transport PDE is intimately related to averaging in ODE. See Chapters 14, 21.

In the case where $D \neq 0$ the equation (1.1.6) is the backward Kolmogorov equation for $x$ solving the SDE

$$\frac{dx^\varepsilon}{dt} = v\left(\frac{x^\varepsilon}{\varepsilon}\right) + \sqrt{2D}dW,$$

(1.1.7)

where $W$ is Brownian motion. This means that

$$T(x(0), t) = \mathbb{E}(f(x(t))|x(0)),$$

where $\mathbb{E}$ denotes averaging with respect to the measure on Brownian motion (Wiener measure). Again we can try and eliminate the rapidly varying quantity $x/\varepsilon$. If $v$ is periodic, divergence free and mean zero then the result described in the previous example shows that, $x^\varepsilon(t)$, the solution of (1.1.7), converges, in the limit as $\varepsilon \to 0$, to $X(t)$ where $X(t)$ is a Brownian motion with diffusion coefficient $\sqrt{2K}$:

$$\frac{dX}{dt} = \sqrt{2K}\frac{dW}{dt}.$$

The connection between homogenization in parabolic PDE and in SDE is discussed in Chapter 11. Rigorous homogenization theorems for SDE are proved 18.
1.1. MOTIVATING EXAMPLES

1.1.4 Example IV: Dimension Reduction in Dynamical Systems

The methods applied to derive homogenized elliptic and parabolic PDE can also be used to average out, or homogenize, the fast scales in systems of ODE and SDE. Doing so leads to effective equations which do not contain the small parameter \(\varepsilon\) and are hence more amenable to numerical solution. The prototypical example is a dynamical systems of the form

\[
\frac{dx}{dt} = f(x, y), \\
\frac{dy}{dt} = \frac{1}{\varepsilon}g(x, y).
\]

In situations of this type, where there is a scale separation, it is often the case that \(y\) can be eliminated and an approximate equation for the evolution of \(x\) can be found. We write the approximate equation in the form

\[
\frac{dX}{dt} = F(X).
\]

In the simplest situation \(y\) is eliminated through an invariant manifold; in more complex systems it is eliminated through averaging. Such results may be viewed as functional versions of the law of large numbers.

In some situations \(F \equiv 0\) and it is then necessary to scale the equations to a longer time \(s = t/\varepsilon\) to see interesting effects. The starting point is then

\[
\frac{dx}{ds} = \frac{1}{\varepsilon}f(x, y), \\
\frac{dy}{dt} = \frac{1}{\varepsilon^2}g(x, y).
\]

In this situation \(f\) essentially averages to zero when \(y\) is eliminated and the effective equation sees the fluctuations in \(f\); the approximate equation takes the form

\[
\frac{dX}{dt} = F(X) + A(X) \frac{dW}{dt}
\]

where \(W\) is a standard unit Brownian motion; this may be viewed as a central limit theorem. Derivation of this equation we refer to as homogenization.

As for linear PDE, techniques from analysis enable us to estimate errors between the original and simplified equations. The formal asymptotics underlying these ideas are fleshed out in Chapters 8, 10 and 11; related rigorous results are given in Chapters 15, 17 and 18.
CHAPTER 1. INTRODUCTION

1.2 Averaging Versus Homogenization

The unifying principle underlying the derivation of most of the effective equations in these notes concerns formal perturbation expansions for linear operator equations of the form

\[ \lambda \frac{\partial u}{\partial t} = \mathcal{L}^\varepsilon u. \]

In particular we will be interested in cases where the operator \( \mathcal{L}^\varepsilon \) has the form

\[ \mathcal{L}^\varepsilon = \frac{1}{\varepsilon} \mathcal{L}_0 + \mathcal{L}_1, \]  

(1.2.1)

or

\[ \mathcal{L}^\varepsilon = \frac{1}{\varepsilon^2} \mathcal{L}_0 + \frac{1}{\varepsilon} \mathcal{L}_1 + \mathcal{L}_2. \]  

(1.2.2)

In both cases \( \mathcal{L}_0 \) will have a nontrivial null space and interest focusses on capturing the dynamics within this subspace. The first case (1.2.1) we will refer to as \textit{averaging}, or \textit{first order perturbation theory}. It can be thought of as a form (or consequence) of the law of large numbers. The second, (1.2.2), will be referred to as \textit{homogenization} or \textit{second order perturbation theory}. It can be thought of as a form (or consequence) of the central limit theorem; indeed (1.2.2) enables us to study fluctuations around the average.

Our main interest will be in the case \( \lambda \neq 0 \). This will apply to the study of homogenization for parabolic PDE and transport PDE, and also to averaging and homogenization for ODE, Markov chains and SDE, via the Kolmogorov equations and variants (the forward equation, the method of characteristics). We will also study the case \( \lambda = 0 \) which arises when studying homogenization for elliptic PDE. The case where a second order time-derivative appears (wave equations) will not be covered explicitly in the notes, but the techniques developed do apply; references to the literature will be given.

1.3 Discussion and Bibliography

The use of formal multiscale expansions in the study of homogenization was developed systematically and applied to many different problems in [17]. See also [116, 115, 120]. Mode reduction for either stochastic or deterministic systems has become a very important area in applied mathematics. Applications include atmosphere/ocean science [96], molecular dynamics, [134], mathematical finance [48].
Various aspects of stochastic and deterministic mode reduction, including numerical algorithms for these problems, are discussed in the review articles [58, 71, 72]. Homogenization for advection–diffusion equations is a part of the theory of turbulent diffusion. See [94] for an excellent review of the theory of turbulent diffusion.

One of the main technical tools that will be used in these notes is the connection between parabolic and elliptic PDE and SDE through the Feynmann–Kac formula, [111]. The relation between homogenization and averaging for PDE and limit theorems for (Markov) stochastic processes has been recognized, and studied extensively, since the early 1960s [49]. A rather general theory of homogenization and averaging for stochastic processes was developed in the 1970s [85, 115, 116, 120]. The mathematical theory is presented in the monograph [41].

There are many important topics related to homogenization and averaging for ODE, SDE and PDE that will not be discussed in these notes. Let us comment on some of these topics are provide some references to the literature. Non–periodic, deterministic, homogenization will not be covered here. The interested reader may consult [30, Ch. 13], [34] and the references therein. Bounds on the homogenized coefficients and their dependence on the non–dimensional parameters of the problem will be discussed very briefly. Information about these topics can be found in [106]. We will not study homogenization problems for non–linear PDE; see [42] and the references therein. Variational methods–which apply to elliptic PDE–and the theory of $\Gamma$–convergence will not be discussed in these notes; see [102, 33, 23]. Related concepts such as $H$ and $G$ convergence will not be developed in these notes. See, e.g. [138, 139, 144]. We will not discuss about homogenization for random PDE (i.e. PDE whose coefficients are random fields). See [114]. More detailed pointers to the literature will be provided in the Discussion and Bibliography section which concludes each chapter.
Part I

Background
Chapter 2

Analysis

2.1 Set–Up

In this chapter we collect a variety of definitions and theorems (mainly without proofs) from analysis. The topics are selected because they will be needed in the sequel. Proofs and additional material can be found in the books cited in Section 2.7. The presentation is necessarily terse, and it is not recommended that this chapter is read in its entirety on a first read through the book; rather that it be referred to as needed when reading later chapters. The rigorous setting developed in this chapter is required primarily in Chapters 15–21 where we prove a number of results concerning dimension reduction for dynamical problems (averaging and homogenization) and homogenization for PDE. The results themselves are derived in Chapters 8–14 by means of perturbation expansions and in that context the rigorous setting is not required.

When proving rigorous results about our reduced systems, the natural setting to answer these questions in the context of PDE is the theory of weak convergence in Hilbert and Banach spaces; in this context we will also develop an appropriate kind of weak convergence, that of two–scale convergence, which is very useful for problems related to periodic homogenization. A complementary chapter on probability, Chapter 3, provides the appropriate convergence tools for the study of SDE and Markov chains; in particular we develop background material on weak convergence of probability measures.
2.2 Notation

In this book we will encounter scalar, vector and matrix fields – all examples of tensor fields. We will adopt an index free notation throughout the book and the purpose of this section is to explain this notation, using indices to do so. We adopt the Einstein summation convention, whereby repeated indexes imply a summation, single indexes range freely over \{1, 2, \ldots, d\}, \(d\) being the dimension, and no index may appear more than twice in any product term. We start with tensor algebra and then move to tensor calculus. We use \(\mathbb{R}^d\) to denote \(d\)-dimensional Euclidean space, and \(\mathbb{T}^d\) the \(d\)-dimensional torus \(\mathbb{R}^d/\mathbb{Z}\).

We will denote by \(\{e_i\}_{i=1}^d\) the standard basis in \(\mathbb{R}^d\). Thus arbitrary \(\xi \in \mathbb{R}^d\) can be written as \(\xi = \xi_i e_i\), where \(\xi_i = \langle \xi, e_i \rangle\). We use \(\cdot\) to denote the inner-product between two vectors, so that

\[ a \cdot b = a_i b_i. \]

We will also use \(\langle \cdot, \cdot \rangle\) to denote this standard inner-product on \(\mathbb{R}^d\). The norm induced by this inner-product is the Euclidean norm

\[ |a| = \sqrt{a \cdot a} \]

and it follows that

\[ |a|^2 = \sum_{j=1}^{d} a_i^2, \quad \xi \in \mathbb{R}^d. \]

We will also use the usual \(L^p\) norms on \(\mathbb{R}^d\), \(1 \leq p \leq \infty\) and denote them by \(|\cdot|_p\). Note that \(|\cdot| = |\cdot|_2\). In addition we use \(|\cdot|_p\) to denote the associated operator norm defined by

\[ |A|_p = \sup_{x \neq 0} \frac{|Ax|_p}{|x|_p}. \]

In particular we use the notation \(|A| = |A|_2\), for operator norms.

The inner-product between matrices is denoted by

\[ A : B = \text{tr}(A^T B) = a_{ij} b_{ij}. \]

The norm induced by this inner product is the Frobenius norm

\[ |A|_F = \sqrt{\text{tr}(A^T A)}. \]

The outer-product between two vectors \(a\) and \(b\) is the matrix \(a \otimes b\) defined by

\[ (a \otimes b)c = (b \cdot c)a \quad \forall c \in \mathbb{R}^d. \]
Let $\nabla$ and $\nabla \cdot$ denote gradient and divergence in $\mathbb{R}^d$. The gradient lifts a scalar (resp. vector) to a vector (resp. matrix) whilst the divergence contracts a vector (resp. matrix) to a scalar (resp. vector). The gradient acts on scalar valued functions $\phi$, or vector valued functions $v$, via

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

The divergence acts on vector valued functions $v$, or matrix valued functions $A$ via

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

Given vector fields $a, v$ we use the notation

$$a \cdot \nabla v := (\nabla v)a.$$ 

Since the gradient is defined for scalars and vectors we readily make sense of the expression

$$\nabla \nabla \phi$$

for any scalar $\phi$; it is the Hessian matrix $\frac{\partial^2 \phi}{\partial x_i \partial x_j}$. Similarly, we can also make sense of the expression

$$\nabla \nabla v$$

by applying $\nabla \nabla$ to each scalar component of the vector $v$.

In many instances in what follows we will take the gradient or divergence with respect to a variety of different independent variables. We use $\nabla_x$ to denote gradient or divergence with respect to $x$ coordinates alone, and similarly for other independent variables.

### 2.3 Banach and Hilbert Spaces

We assume that the reader is already familiar with the definitions of a norm, an inner product and of vector, metric, normed and inner product spaces. In the sequel $X$ will denote a normed space and $\| \cdot \|$ will denote its norm. We say that a sequence $\{x_j\}_{j=1}^{\infty} \subset X$ converges strongly to $x \in X$, written as

$$x_j \to x,$$
provided that
\[ \lim_{j \to \infty} \|x_j - x\| = 0. \]

Furthermore, we say that \( \{x_j\}_{j=1}^\infty \subset X \) is a Cauchy sequence provided that for each \( \varepsilon > 0 \) there exists an \( N \in \mathbb{N} \) such that
\[ \|x_j - x_k\| < \varepsilon \quad \forall j, k \geq N. \]

Every convergent sequence in a normed space \( X \) is a Cauchy sequence. The converse in not always true. If, however, every Cauchy sequence in \( X \) is convergent, then the space \( X \) is called complete.

### 2.3.1 Banach Spaces

**Definition 2.1.** A Banach space \( X \) is a complete normed vector space.

**Definition 2.2.** Let \( X \) be a Banach space. We say that a map \( \ell : X \to \mathbb{R} \) is a bounded linear functional on \( X \) provided that
\[ i) \quad \ell(\alpha x + \beta y) = \alpha \ell(x) + \beta \ell(y) \quad \forall x, y \in X, \alpha, \beta \in \mathbb{R}. \]
\[ ii) \quad \exists C > 0 : |\ell(x)| \leq C \|x\| \quad \forall x \in X. \]

**Definition 2.3.** The collection of all bounded linear functionals on \( X \) is called the dual space and denoted by \( X^* \).

**Theorem 2.4.** \( X^* \) equipped with the norm
\[ \|\ell\| = \sup_{x \neq 0} \frac{|\ell(x)|}{\|x\|}, \]

is a Banach space.

**Definition 2.5.** A Banach space \( X \) is called reflexive if the dual of its dual is \( X \):
\[ (X^*)^* = X. \]

The concept of the dual space enables us to introduce another topology on \( X \), the so called weak topology.

**Definition 2.6.** A sequence \( \{x_n\}_{n=1}^\infty \) is said to converge weakly to \( x \in X \), written
\[ x_n \rightharpoonup x, \]
if
\[ \ell(x_n) \to \ell(x) \quad \forall \ell \in X^*. \]
Every strongly convergent sequence is also weakly convergent. However, the converse is not true. The importance of weak convergence stems from the following theorem, in particular part (ii).

**Theorem 2.7.** Let $X$ be a Banach space.

(i) Every weakly convergent sequence in $X$ is bounded.

(ii) (Eberlein–Smuljan) Assume that $X$ is reflexive. Then from every bounded sequence in $X$ we can extract a weakly convergent subsequence.

In addition to the weak topology on $X$, the duality between $X$ and $X^*$ enables us to define a topology on $X^*$.

**Definition 2.8.** Let $X$ be a Banach space. A sequence $\{\ell_n\}_{n=1}^{\infty} \in X^*$ is said to converge weakly–$\star$ to $\ell \in X^*$ written

$$\ell_n \xrightarrow{\text{weak–$\star$}} \ell,$$

if

$$\lim_{n \to \infty} \ell_n(x) = \ell(x) \quad \forall x \in X.$$

We remark that if $X$ is reflexive then weak–$\star$ convergence coincides with weak convergence on $X$.

A compactness result similar to Theorem 2.7 holds for bounded sequences in $X^*$, provided that $X$ is separable. We remind the reader that a subset $X_0$ of $X$ is called dense if for every $x \in X$ there exists a sequence $\{x_j\}_{j=1}^{\infty} \subset X_0$ which converges to $x$. In other words, $X_0$ is dense in $X$ if its closure is $X$: $\overline{X_0} = X$.

**Definition 2.9.** A Banach space $X$ is called separable if it contains a countable dense subset.

The compactness theorem for sequences in $X^*$ can be stated as follows.

**Theorem 2.10.** Let $X$ be a separable Banach space. Then from any bounded sequence in $X^*$ we can extract a weakly convergent subsequence.

### 2.3.2 Hilbert Spaces

**Definition 2.11.** A Hilbert space is a complete inner product space.
We denote the inner-product by \((\cdot, \cdot)\). Clearly, every Hilbert space is a Banach space. Naturally, the inner product defines a norm on \(H\):

\[
\|x\| = (x, x)^{\frac{1}{2}}.
\]

Furthermore, all elements of \(H\) satisfy the Cauchy–Schwarz inequality:

\[
|(u, v)| \leq \|u\| \|v\|.
\]

A very important property of a Hilbert space \(H\) is that we can identify the dual of \(H\) with itself through the Riesz representation theorem.

**Theorem 2.12. (Riesz Representation.)** For every \(\ell \in H^*\) there exists a unique \(y \in H\) such that

\[
\ell(x) = (x, y) \quad \forall x \in H.
\]

We will usually denote the action of \(\ell\) on \(x \in H\) by \(^1\langle \cdot, \cdot \rangle\):

\[
\langle \ell, x \rangle := \ell(x).
\]

The Riesz representation theorem implies that every Hilbert space is reflexive and consequently Theorem 2.7 applies. Furthermore, the definition of weak convergence simplifies to

\[
x_n \rightharpoonup x \iff (x_n, y) \to 0 \quad \forall y \in H.
\]

### 2.4 Function Spaces

Let \(\Omega\) be an open subset of \(\mathbb{R}^d\). We will denote by \(C(\Omega)\) the space of continuous functions \(f : \Omega \to \mathbb{R}\). This space, when equipped with the supremum norm

\[
\|f\|_{C(\Omega)} = \sup_{x \in \Omega} |f(x)|,
\]

is a Banach space. Similarly, we can define the space \(C^p(\Omega)\) of \(p\)-times continuously differentiable functions. We will denote by \(C^\infty(\Omega)\) the space of smooth functions. The notation \(C^\infty_0(\Omega)\) will be used to denote the space of smooth functions over \(\Omega\) with compact support. We will use the notation \(C^k_b(\mathbb{R}^d)\) to denote

---

\(^1\)In \(\mathbb{R}^d\) we use \((\cdot, \cdot)\) to denote the inner-product and it is useful to retain this convention; in the infinite dimensional setting we will always use \((\cdot, \cdot)\) for the inner-product and \(\langle \cdot, \cdot \rangle\) for the dual pairing between \(H\) and \(H^*\).
2.4. FUNCTION SPACES

the set of \( k \)-times continuously differentiable functions from \( \mathbb{R}^d \) to \( \mathbb{R} \) which are bounded, together with all of their derivatives. \( C^\infty_k(\mathbb{R}^d) \) will denote the space of bounded continuous functions having bounded continuous derivatives of all orders. More generally, let \((E, \rho)\) be a metric space. Then \( C_b(E) \) will denote the space of bounded continuous functions on \((E, \rho)\).

2.4.1 \( L^p \) Spaces

Let \( 1 \leq p \leq \infty \) and let \( f : \Omega \to \mathbb{R} \) be a measurable function. Define

\[
\|f\|_{L^p(\Omega)} := \begin{cases} 
\left( \int_\Omega |f|^p \, dx \right)^{1/p}, & \text{for } 1 \leq p < \infty \\
\text{ess sup}_\Omega |f|, & \text{for } p = \infty.
\end{cases}
\]

In the above definition we have used the notation

\[
\text{ess sup} = \inf \{C, |f| \leq C \text{ a.e. on } \Omega\}.
\]

**Definition 2.13.** \( L^p(\Omega) \) is the vector space of all measurable functions \( f : \Omega \to \mathbb{R} \)

for which \( \|f\|_{L^p(\Omega)} < \infty \).

**Theorem 2.14.** (Basic properties of \( L^p \) spaces.)

i) The vector space \( L^p(\Omega) \), equipped with the \( L^p \)-norm defined above, is a Banach space for every \( p \in [1, +\infty] \).

ii) \( L^2(\Omega) \) is a Hilbert space equipped with the inner product

\[
(u, v)_{L^2(\Omega)} = \int_\Omega u(x)v(x) \, dx \quad \forall u, v \in L^2(\Omega).
\]

iii) \( L^p(\Omega) \) is separable for \( p \in [1, +\infty) \) and reflexive for \( p \in (1, +\infty) \). In particular, \( L^1(\Omega) \) is not reflexive and \( L^\infty(\Omega) \) is neither separable nor reflexive.

Let \( p \in [1, +\infty] \) and define \( q \in [1, +\infty] \) through

\[
\frac{1}{p} + \frac{1}{q} = 1.
\]  

Then the Hölder inequality states that

\[
\left| \int_\Omega u(x)v(x) \, dx \right| \leq \|u\|_{L^p(\Omega)}\|v\|_{L^q(\Omega)} \quad \forall u \in L^p(\Omega), \ v \in L^q(\Omega).
\]
Let $p \in [1, +\infty)$ and let $q$ be defined through (2.4.1). Then

$$(L^p(\Omega))^* = L^q(\Omega).$$

The last part of the above theorem, together with the fact that $L^p(\Omega)$ is a Banach space and Definition 2.6, implies the following definition of weak convergence in $L^p(\Omega)$, $p \in [1, +\infty)$.

**Definition 2.15.** A sequence $\{u_n(x)\}_{n=1}^{\infty} \in L^p(\Omega)$, $p \in [1, +\infty)$ is said to converge weakly to $u(x) \in L^p(\Omega)$, written

$$u_n(x) \rightharpoonup u(x) \quad \text{$w$–$L^p(\Omega)$},$$

provided that

$$\int_{\Omega} u_n(x)v(x) \, dx \to \int_{\Omega} u(x)v(x) \, dx \quad \forall v \in L^q(\Omega),$$

where $q$ is defined in (2.4.1).

Whereas weak convergence in $L^\infty$ is very rarely used, the notion of weak–* convergence in that space is very useful.

**Definition 2.16.** A sequence $\{u_n\}_{n=1}^{\infty} \in L^\infty(\Omega)$ converges weak–* in $L^\infty$, written

$$u_n(x) \rightharpoonup u(x) \quad \text{$w^*$–$L^\infty(\Omega)$},$$

provided that

$$\int_{\Omega} u_n(x)\phi(x) \, dx \to \int_{\Omega} u(x)\phi(x) \, dx, \quad \forall \phi \in L^1(\Omega).$$

### 2.4.2 Sobolev Spaces

We start with the definition of the weak derivative.

**Definition 2.17.** Let $u, v \in L^2(\Omega)$. We say that $v$ is the first weak derivative of $u$ with respect to $x_i$ if

$$\int_{\Omega} u \frac{\partial \phi}{\partial x_i} \, dx = - \int_{\Omega} v \phi \, dx \quad \forall \phi \in C_0^\infty(\Omega).$$

In the context of a function $u \in L^2(\Omega)$ we will use the notation $\frac{\partial u}{\partial x_i}$ to denote the weak derivative with respect to $x_i$. We will also use the notation $\nabla u = \frac{\partial u}{\partial x_i} e_i$. 
2.4. FUNCTION SPACES

**Definition 2.18.** The Sobolev space $H^1(\Omega)$ consists of all square integrable functions from $\Omega$ to $\mathbb{R}$ whose first order weak derivatives exist and are square integrable:

$$H^1(\Omega) = \left\{ u \mid u \in L^2(\Omega), \nabla u \in L^2((\Omega))^d \right\}.$$  

The space $H^1(\Omega)$ is a separable Hilbert space with norm

$$\|u\|_{H^1(\Omega)} = \left( \|u\|^2_{L^2(\Omega)} + \|\nabla u\|^2_{L^2(\Omega)} \right)^{\frac{1}{2}}$$

and inner product

$$(u, v)_{H^1(\Omega)} = (u, v)_{L^2(\Omega)} + (\nabla u, \nabla v)_{L^2(\Omega)}.$$  

A very useful property of $H^1(\Omega)$ is that its embedding into $L^2(\Omega)$ is compact. We will state this result in the following way.

**Theorem 2.19. (Rellich Compactness Theorem).** From every bounded sequence in $H^1(\Omega)$ we can extract a subsequence which is strongly convergent in $L^2(\Omega)$.

In many applications one is interested in elements of $H^1(\Omega)$ which vanish on the boundary $\partial\Omega$ of the domain. Functions with this property belong to the following subset of $H^1_0(\Omega)$.

**Definition 2.20.** The Sobolev space $H^1_0(\Omega)$ is defined as the completion of $C_0^\infty(\Omega)$ with respect to the $H^1$–norm.

A very important property of $H^1_0(\Omega)$ is the fact that we can control the $L^2$ norm of its elements in terms of the $L^2$–norm of their gradient. Below we present this result, together with its analogue for elements of $H^1(\Omega)$.

**Theorem 2.21. (Poincaré Inequality) Let $\Omega$ be a bounded open set in $\mathbb{R}^d$. Then there is a constant $C_\Omega$, which depends only on $\Omega$, such that**

$$\|u\|_{L^2(\Omega)} \leq C_\Omega \|\nabla u\|_{L^2(\Omega)} \quad (2.4.2)$$

**for every $u \in H^1_0(\Omega)$.**

An immediate corollary of the first part of the above theorem is that $\|\nabla \cdot\|_{L^2(\Omega)}$ can be used as the norm in $H^1_0(\Omega)$:

$$\|u\|_{H^1_0(\Omega)} = \|\nabla u\|_{L^2(\Omega)}. \quad (2.4.3)$$
We will denote by $H^{-1}(\Omega)$ the dual space of $H^1_0(\Omega)$. Further, we will denote by $\langle \cdot, \cdot \rangle_{H^1_0, H^{-1}}$ the pairing between $H^{-1}(\Omega)$ and $H^1_0(\Omega)$. In other words, the action of $f \in H^{-1}(\Omega)$ on $v \in H^1_0(\Omega)$ will be denoted by $\langle f, v \rangle_{H^1_0, H^{-1}}$. Then $H^{-1}(\Omega)$ is a Banach space equipped with the norm

$$\|f\|_{H^{-1}(\Omega)} = \sup_{v \in H^1_0(\Omega), \|v\|_{H^1_0(\Omega)} \leq 1} |\langle f, v \rangle_{H^1_0, H^{-1}}|.$$ 

The following version of the Cauchy–Schwarz inequality holds

$$|\langle f, v \rangle_{H^1_0, H^{-1}}| \leq \|f\|_{H^{-1}(\Omega)} \|v\|_{H^1_0(\Omega)} \quad \forall f \in H^{-1}(\Omega), \forall v \in H^1_0(\Omega). \quad (2.4.4)$$

### 2.4.3 Banach Space–Valued Spaces

It is possible to define $L^p$ spaces over sets and spaces more general than subsets of $\mathbb{R}^d$; in particular, one can define spaces $L^p(\Omega)$ when $\Omega$ is a Banach space itself. The spaces defined below appear quite often in applications. They are particularly relevant for the rigorous analysis of periodic homogenization since we have to deal with functions of two arguments ($x$ and $y$, say.)

**Definition 2.22.** Let $X$ be a Banach space with norm $\| \cdot \|_X$ and let $\Omega$ denote a subset of $\mathbb{R}^d$, not necessarily bounded. The space $Y := L^p(\Omega; X)$ with $p \in [1, +\infty]$ consists of all measurable functions $u : x \in \Omega \mapsto u(x) \in X$ such that $\|u(x)\|_X \in L^p(\Omega)$.

The space $Y$ defined above has various nice properties which we list below.

**Theorem 2.23.** Let $Y = L^p(\Omega; X)$ with $X$ and $\Omega$ as in Definition 2.22. Then

(i) $Y$ equipped with the norm

$$\|u\|_Y = \left( \int_{\Omega} \|u(x)\|_X^p \, dx \right)^{\frac{1}{p}},$$

is a Banach space.

(ii) If $X$ is reflexive and $p \in (1, +\infty)$ then $Y$ is also reflexive.

(iii) If $X$ is separable and $p \in [1, +\infty)$ then $Y$ is also separable.
One can define spaces of the form $H^1(\Omega; X)$, where $X$ is a Banach space, in a similar fashion.

We will use $C^{k,p}([0, T] \times \mathbb{R}^d)$ to denote the space of functions which are $k$ and $p$ times continuously differentiable in $t \in [0, T]$ and $x \in \Omega$, respectively.

**Example 2.24.** Let $T > 0$; then function spaces of the form $L^p((0, T); X)$ appear in the analysis of evolution PDE of parabolic and hyperbolic type – see Chapter 7.

**Example 2.25.** If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space then $L^p(\Omega; X)$ will denote a Banach space found by defining the $L^p$ integration with respect to the measure $\mathbb{P}$ on $\Omega$, rather than with respect to Lebesgue measure as above – see Chapter 3.

### 2.4.4 Sobolev spaces of Periodic Functions

We will use the notation $\mathcal{Y} = [0, 1]^d$. We will refer to $\mathcal{Y}$ as the unit cell. In this section we will study some properties of periodic functions $f : \mathcal{Y} \rightarrow \mathbb{R}$. We thus identify $\mathcal{Y}$ with the torus $\mathbb{T}^d$. Functions which are periodic with period 1 will be called 1–periodic functions.

We will denote by $C^\infty_{\text{per}}(\mathcal{Y})$ the space of smooth functions $C^\infty(\mathbb{R}^d)$ which are 1–periodic. $L^p_{\text{per}}(\mathcal{Y})$ is defined to be the completion of $C^\infty_{\text{per}}(\mathcal{Y})$ with respect to the $L^p$–norm. A similar definition holds for $H^1_{\text{per}}(\mathcal{Y})$.

The Poincaré inequality does not hold in the space $H^1_{\text{per}}$. It does hold, however, if we remove the constants from this space. We define the space

$$H = \left\{ u \in H^1_{\text{per}}(\mathcal{Y}) \left| \int_{\mathcal{Y}} u \, dy = 0 \right. \right\}. \quad (2.4.5)$$

There exists a constant $C_{\Omega} > 0$, depending only on $\Omega$, such that

$$\|u\|_{L^2(\mathcal{Y})} \leq C_{\Omega} \|\nabla u\|_{L^2(\mathcal{Y})} \quad \forall u \in H.$$

Hence, we can use

$$\|u\|_H = \|\nabla u\|_{L^2(\mathcal{Y})}, \quad \forall u \in H, \quad (2.4.6)$$

as the norm in $H$. The dual of $H$ may be shown to comprise all elements of $(H^1_{\text{per}}(\mathcal{Y}))^*$ which are orthogonal to constants:

$$H^* = \left\{ u \in (H^1_{\text{per}}(\mathcal{Y}))^* ; \langle f, 1 \rangle = 0 \right\},$$

where $\langle \cdot, \cdot \rangle$ denotes the pairing between $H^1_{\text{per}}(\mathcal{Y})$ and its dual.
The space

\[ L^2(\Omega; L^2(\mathcal{Y})) := L^2(\Omega \times \mathcal{Y}) \]

will be used in our study of periodic homogenization. This is a Hilbert space with inner-product

\[ (u, v)_{L^2(\Omega \times \mathcal{Y})} = \int_\Omega \int_\mathcal{Y} u(x, y)v(x, y) \, dy \, dx, \]

together with the corresponding norm. We will on occasion use the space \( L^2(\Omega; C_{per}(\mathcal{Y})) \), which is the set of all measurable functions \( u : x \in \Omega \rightarrow u(x) \in C_{per}(\mathcal{Y}) \) such that \( \|u(x)\| \in L^2(\Omega) \). By Theorem 2.23 the norm of this space is

\[ \|u\|^2_{L^2(\Omega; C_{per}(\mathcal{Y}))} = \int_\Omega \left| \sup_{y \in \mathcal{Y}} u(x, y) \right|^2 \, dy. \]

This is a separable Banach space which is dense in \( L^2(\Omega \times \mathcal{Y}) \). It enjoys various properties which we will need.

**Theorem 2.26.** Let \( u \in L^2(\Omega; C_{per}(\mathcal{Y})) \) and \( \varepsilon > 0 \). Then

(i) \( u \left( x, \frac{x}{\varepsilon} \right) \in L^2(\Omega) \) and

\[ \left\| u \left( x, \frac{x}{\varepsilon} \right) \right\|_{L^2(\Omega)} \leq \|u(x, y)\|_{L^2(\Omega; C_{per}(\mathcal{Y}))}. \]

(ii) \( u \left( x, \frac{x}{\varepsilon} \right) \) converges to \( \int_\mathcal{Y} u(x, y) \, dy \), weakly in \( L^2(\Omega) \) as \( \varepsilon \to 0 \).

(iii) We have

\[ \left\| u \left( x, \frac{x}{\varepsilon} \right) \right\|_{L^2(\Omega)} \to \|u(x, y)\|_{L^2(\Omega \times \mathcal{Y})} \]

as \( \varepsilon \to 0 \).

The following property of periodic functions will be used in the sequel.

**Theorem 2.27.** Let \( p \in [1, +\infty] \) and \( f \) be a \( \mathcal{Y} \)-periodic function in \( L^p(\mathcal{Y}) \). Set

\[ f_\varepsilon(x) = f \left( \frac{x}{\varepsilon} \right) \text{ a.e. on } \mathbb{R}^d. \]

Then, if \( p < +\infty \), as \( \varepsilon \to 0 \)

\[ f_\varepsilon \rightharpoonup \int_\mathcal{Y} f(y) \, dy \text{ weakly in } L^p(\Omega), \]

for any bounded open subset \( \Omega \) of \( \mathbb{R}^d \).

We also have

\[ f_\varepsilon \rightharpoonup \star \text{ in } L^\infty(\mathbb{R}^d). \]
2.5 Two–Scale Convergence

A form of weak convergence that is particularly well suited for problems in periodic homogenization is that of two–scale convergence. In this section we define this concept and we study some of its basic properties. Once again we use $\mathcal{Y}$ to denote the unit cell and, since we consider periodic functions on $\mathcal{Y}$, identify it with the torus $T^d$. As before $\Omega$ denotes a subset of $\mathbb{R}^d$. We start by discussing two-scale convergence for steady (time-independent) problems. We then discuss related issues for time-dependent problems.

2.5.1 Two–Scale convergence for steady problems

In order to define the concept of two–scale convergence we first need to consider appropriate test functions, the admissible test functions.

**Definition 2.28.** A function $\phi(x, y) \in L^2(\Omega \times \mathcal{Y})$ is called an admissible test function if it satisfies

$$
\lim_{\varepsilon \to 0} \int_{\Omega} \left| \phi(x, \frac{x}{\varepsilon}) \right|^2 dx = \int_{\Omega} \int_{\mathcal{Y}} |\phi(x, y)|^2 dy dx. \quad (2.5.1)
$$

Note that an admissible test function is one for which, asymptotically, the dependence on $x$ and $x/\varepsilon$ decouples, in the calculation of $L^2$–norms. By Theorem 2.26 any $\phi \in L^2(\Omega; C_{\text{per}}(\mathcal{Y}))$ is an admissible test function. Now we are ready to give the definition of two–scale convergence.

**Definition 2.29.** Let $u^\varepsilon$ be a sequence in $L^2(\Omega)$. We will say that $u^\varepsilon$ two–scale converges to $u_0(x, y) \in L^2(\Omega \times \mathcal{Y})$ and write $u^\varepsilon \rightharpoonup u_0$ if for every admissible test function $\phi$ we have

$$
\lim_{\varepsilon \to 0} \int_{\Omega} u^\varepsilon(x) \phi \left( x, \frac{x}{\varepsilon} \right) dx = \int_{\Omega} \int_{\mathcal{Y}} u_0(x, y) \phi(x, y) dy dx. \quad (2.5.2)
$$

Two–scale convergence implies weak convergence in $L^2(\Omega)$. In particular, we have the following lemma.

**Lemma 2.30.** Let $u^\varepsilon$ be a sequence in $L^2(\Omega)$ which two–scale converges to $u_0(x, y) \in L^2(\Omega \times \mathcal{Y})$. Then

$$
u^\varepsilon \rightharpoonup \nu_0(x) := \int_{\mathcal{Y}} u_0(x, y) dy, \quad \text{weakly in } L^2(\Omega).$$
CHAPTER 2. ANALYSIS

Proof. Choose a test function $\phi(x) \in L^2(\Omega)$, independent of $\varepsilon$. This is an admissible test function since $\int_Y dy = 1$ and we can use it in (2.5.2) to deduce:

$$
\lim_{\varepsilon \to 0} \int_{\Omega} u^\varepsilon(x) \phi(x) \, dx = \int_{\Omega} \int_Y u_0(x, y) \phi(x) \, dy \, dx
= \int_{\Omega} \left( \int_Y u_0(x, y) \, dy \right) \phi(x) \, dx
= (\mathfrak{m}_0, \phi)_{L^2(\Omega)}.
$$

The above holds for every $\phi(x) \in L^2(\Omega)$ and, hence, $u^\varepsilon$ converges to $u_0$ weakly in $L^2(\Omega)$.

An immediate consequence of the above lemma is the following.

**Corollary 2.31.** Let $u^\varepsilon$ be a sequence in $L^2(\Omega)$ which two–scale converges to $u_0(x) \in L^2(\Omega)$, i.e. the two–scale limit is independent of $y$. Then the weak $L^2$–limit and the two–scale limit coincide.

Two–scale convergence is a useful tool for studying multiscale expansions, with a periodic dependence on the "fast variable", as the next result indicates.

**Lemma 2.32.** Consider a function $u^\varepsilon(x) \in L^2(\Omega)$ of the form

$$
u^\varepsilon(x) = u_0\left(x, \frac{x}{\varepsilon}\right) + \varepsilon u_1\left(x, \frac{x}{\varepsilon}\right),
$$

where $u_j(x, y) \in L^2(\Omega; C_{\text{per}}(Y))$, $j = 0, 1$, $\Omega$ being a bounded domain in $\mathbb{R}^d$. Then $u^\varepsilon \overset{w}{\to} u_0$.

**Proof.** Let $\phi(x, y) \in L^2(\Omega; C_{\text{per}}(Y))$ and define $f_j(x, y) = u_j(x, y)\phi(x, y)$, $j = 0, 1$. We will use the notation $f^\varepsilon(x) = f(x, \frac{x}{\varepsilon})$. We clearly have

$$
\int_{\Omega} \int_Y f_0^\varepsilon(x) \phi\left(x, \frac{x}{\varepsilon}\right) \, dx = \int_{\Omega} f_0^\varepsilon(x) \, dx + \varepsilon \int_{\Omega} f_1^\varepsilon(x) \, dx.
$$

Now, $f_j^\varepsilon(x) \in L^2(\Omega; C_{\text{per}}(Y))$ for $j = 0, 1$. This implies, by Theorem 2.26, that $f_0^\varepsilon$ converges to its average over $Y$ $\int_Y f_0(x, y) \, dy$, weakly in $L^2(\Omega)$. Now choose $\psi = 1$ (notice that $1 \in L^2(\Omega)$, since $\Omega$ is a bounded subset of $\mathbb{R}^d$) to obtain

$$
\int_{\Omega} f_0^\varepsilon(x) \, dx \to \int_{\Omega} \int_Y f_0(x, y) \, dy \, dx
= \int_{\Omega} \int_Y u_0(x, y) \phi(x, y) \, dy \, dx.
$$
2.5. **TWO–SCALE CONVERGENCE**

Now consider the second integral on the right hand side of (2.5.3). Since the sequence \( f_1^\varepsilon \) is weakly convergent in \( L^2(\Omega) \), it is bounded by Theorem 2.7. Thus, using again the boundedness of \( \Omega \), together with Cauchy–Schwartz inequality, we obtain:

\[
\varepsilon \int_\Omega f_1^\varepsilon(x) \, dx \leq \varepsilon C \| f_1^\varepsilon \|_{L^2(\Omega)} \leq \varepsilon C \to 0.
\]

We use the above two calculations in (2.5.3) to obtain:

\[
\int_\Omega u^\varepsilon(x) \phi \left( x, \frac{x}{\varepsilon} \right) \, dx \to \int_\Omega \int_Y u_0(x, y) \phi(x, y) \, dy \, dx.
\]

Hence, \( u^\varepsilon \) two–scale converges to \( u_0 \).

Now, we would like to find criteria which enable to conclude that a given sequence in \( L^2(\Omega) \) is two–scale convergent. The following compactness result provides us with such a criterion.

**Theorem 2.33.** Let \( u^\varepsilon \) be a bounded sequence in \( L^2(\Omega) \). Then there exists a subsequence, still denoted by \( u^\varepsilon \), and function \( u_0(x, y) \in L^2(\Omega \times Y) \) such that \( u^\varepsilon \) two–scale converges to \( u_0(x, y) \).

The two–scale convergence defined in Definition 2.29 is still a weak type of convergence, since it is defined in terms of the product of a sequence \( u^\varepsilon \) with an appropriate test function. We can also define a notion of strong two–scale convergence.

**Definition 2.34.** Let \( u^\varepsilon \) be a sequence in \( L^2(\Omega) \). We will say that \( u^\varepsilon \) two–scale converges strongly to \( u_0(x, y) \in L^2(\Omega \times Y) \) and write \( u^\varepsilon \overset{2}{\rightharpoonup} u_0 \) if

\[
\lim_{\varepsilon \to 0} \int_\Omega |u^\varepsilon(x)|^2 \, dx = \int_\Omega \int_Y |u_0(x, y)|^2 \, dy \, dx.
\]  

(2.5.4)

Although every strongly two–scale convergent sequence is also two–scale convergent, the converse is not true. Notice that, in view of the above definition, we can define the class of admissible test functions to be the subset of elements \( \phi(x, y) \) of \( L^2(\Omega \times Y) \) which are periodic in \( y \) and for which \( \phi^\varepsilon(x) := \phi(x, x/\varepsilon) \) is strongly two–scale convergent.

As is always the case with weak convergence, the limit of the product of two two–scale convergent sequences is not in general the product of the limits. However, we can pass to the limit when we one of the two sequences is strongly two–scale convergent.
Theorem 2.35. (i) Let \( u^\varepsilon, v^\varepsilon \) be sequences in \( L^2(\Omega) \) such that \( u^\varepsilon \overset{2}{\rightharpoonup} u_0 \) and \( v^\varepsilon \overset{2}{\rightharpoonup} v_0 \). Then \( u^\varepsilon v^\varepsilon \overset{2}{\rightharpoonup} u_0 v_0 \).

(ii) Assume further that \( u_0(x, y) \in L^2(\Omega; C_{\text{per}}(Y)) \). Then

\[
\| u^\varepsilon(x) - u_0 \left( x, \frac{x}{\varepsilon} \right) \|_{L^2(\Omega)} \to 0.
\]

So far we have only considered bounded sequences in \( L^2(\Omega) \) whose two–scale limit is an element of \( L^2(\Omega \times Y) \) and depends explicitly on \( y \). It is now natural to ask whether more information on the two–scale limit can be obtained when our sequence is bounded in a stronger norm.

Theorem 2.36. (i) Let \( u^\varepsilon \) be a bounded sequence in \( H^1(\Omega) \). Then \( u^\varepsilon \) two–scale converges to its weak–\( H^1 \) limit \( u(x) \). Further, there exists a function \( u_1(x, y) \in L^2(\Omega; H_{\text{per}}^1(Y)/\mathbb{R}) \) such that, up to a subsequence, \( \nabla u^\varepsilon \) two–scale converges to \( \nabla x u(x) + \nabla_y u_1(x, y) \).

(ii) Let \( u^\varepsilon \) and \( \varepsilon \nabla u^\varepsilon \) be uniformly bounded sequences in \( L^2(\Omega) \). Then there exists a function \( u_0(x, y) \in L^2(\Omega; H_{\text{per}}^1(Y)) \) such that, up to a subsequence, \( u^\varepsilon \) and \( \varepsilon \nabla u^\varepsilon \) two–scale converge to \( u_0(x, y) \) and to \( \nabla_y u_0(x, y) \), respectively.

(iii) Let \( v^\varepsilon \) be a divergence–free field which is bounded in \( [L^2(\Omega)]^d \). Then the two–scale limit satisfies \( \nabla_y \cdot v_0(x, y) = 0 \) and \( \int_Y \nabla_x \cdot v_0(x, y) \, dy = 0 \).

Since we only know that \( v_0(x, y) \in [L^2(\Omega \times Y)]^d \), we have to interpret the divergence of \( v_0(x, y) \) with respect to \( x \) and \( y \) in the sense of weak derivatives.

2.5.2 Two–Scale convergence for time dependent problems

When studying homogenization problems for evolution PDE it is necessary to modify the concept of two–scale convergence to take into account the fact the time dependence of the sequences of functions we consider. Below we present the relevant definitions and theorems. As in the previous subsection, we let \( \Omega \) be a subset of \( \mathbb{R}^d \), not necessarily bounded and \( Y = [0, 1]^d \). We use \( (x, y, t) \) to denote a point in \( Y \times \Omega \times [0, T] \).

Definition 2.37. A function \( \phi(x, y, t) \in L^2((0, T) \times \Omega; L^2_{\text{per}}(Y)) \) is called an admissible test function if it satisfies

\[
\lim_{\varepsilon \to 0} \int_0^T \int_\Omega \left| \phi \left( x, \frac{x}{\varepsilon}, t \right) \right|^2 \, dxdt = \int_0^T \int_\Omega \int_Y |\phi(x, y, t)|^2 \, dydxdt.
\]
Definition 2.38. A sequence $u^\varepsilon \in L^2((0, T) \times \Omega)$ two-scale converges to $u_0(x, y, t) \in L^2((0, T) \times \Omega \times \mathcal{Y})$ and write $u^\varepsilon \rightharpoonup u_0$ if for every admissible test function $\phi(x, y, t)$ we have

$$\lim_{\varepsilon \to 0} \int_0^T \int_{\Omega} u^\varepsilon(x, t) \phi\left(x, \frac{x}{\varepsilon}, t\right) \, dxdt = \int_0^T \int_{\Omega} \int_{\mathcal{Y}} u_0(x, y, t) \phi(x, y, t) \, dydxdt.$$

(2.5.6)

The basic compactness theorem of two-scale convergence, Theorem 2.33, is still valid:

Theorem 2.39. Let $u^\varepsilon$ be a bounded sequence in $L^2((0, T) \times \Omega)$. Then there exists a subsequence, still denoted by $u^\varepsilon$, and function $u_0(x, y, t) \in L^2((0, T) \times \Omega \times \mathcal{Y})$ such that $u^\varepsilon(x, t)$ two-scale converges to $u_0(x, y, t)$. Moreover, $u^\varepsilon$ converges weakly in $L^2((0, T) \times \Omega)$ to the average of the two-scale limit over the unit cell:

$$u^\varepsilon \rightharpoonup \int_{\mathcal{Y}} u_0(x, y, t), \quad \text{weakly in } L^2((0, T) \times \Omega).$$

We already know that bounds on better spaces give us more information on the two-scale limit. The theorem below is the analogue of the first two parts of Theorem 2.36.

Theorem 2.40. (i) Let $u^\varepsilon$ be a bounded sequence in $L^2((0, T); H^1(\Omega))$. Then

$u^\varepsilon$ two-scale converges to its weak--$L^2((0, T); H^1(\Omega))$ limit $u(x, t)$. Furthermore, there exists a function $u_1(x, y, t) \in L^2((0, T) \times \Omega; H^1_{\text{per}}(\mathcal{Y})/\mathbb{R})$ such that, up to a subsequence, $\nabla u^\varepsilon$ two-scale converges to $\nabla_x u(x, t) + \nabla_y u_1(x, y, t)$.

(ii) Let $u^\varepsilon$ and $\varepsilon \nabla u^\varepsilon$ be uniformly bounded sequences in $L^2((0, T) \times \Omega)$ and $(L^2((0, T) \times \Omega))^d$, respectively. Then there exists a function $u_0(x, y, t) \in L^2((0, T) \times \Omega; H^1_{\text{per}}(\mathcal{Y})/\mathbb{R})$ such that, up to a subsequence, $u^\varepsilon$ and $\varepsilon \nabla u^\varepsilon$ two-scale converge to $u_0(x, y, t)$ and to $\nabla_y u_0(x, y, t)$, respectively.

The proofs of the these two theorems are almost identical to the proofs of the corresponding results from Section 2.5.

2.6 Equations in Hilbert Spaces

Throughout this book we will frequently encounter linear PDE in a Hilbert space setting. It is consequently useful to develop an abstract formulation for such prob-
lems. We summarize this theory here. There are two main components: the Lax-Milgram existence theory and the Fredholm alternative.

### 2.6.1 Lax-Milgram Theory

Let $H$ be a Hilbert space with inner product $(\cdot, \cdot)$ and let $A : H \to H^*$ be a linear operator. Let $f \in H^*$ and let $(\cdot, \cdot)$ denote the pairing between $H$ and $H^*$. We are interested in studying the equation

$$Au = f. \quad (2.6.1)$$

An equivalent formulation of this equation is

$$(Au, v) = \langle f, v \rangle \quad \forall v \in H. \quad (2.6.2)$$

The linearity of $A$ implies that the left hand side of the above equation defines a bilinear form $a : H \times H \to \mathbb{R}$ given by

$$a[u, v] = (Au, v).$$

Existence and uniqueness of solutions for equations of the form (2.6.2) can be proved by means of the following theorem.

**Theorem 2.41. (Lax–Milgram).** Let $H$ be a Hilbert space with norm $\| \cdot \|$ and inner product $(\cdot, \cdot)$. Let $\langle \cdot, \cdot \rangle$ denote the pairing between $H^*$ and $H$. Let $a : H \times H \to \mathbb{R}$ be a bilinear mapping which satisfies the following properties:

(i) **(Coercivity)** There exists a constant $\alpha > 0$ such that

$$a[u, u] \geq \alpha \|u\|^2 \quad \forall u \in H.$$

(ii) **(Continuity)** There exists a constant $\beta > 0$ such that

$$a[u, v] \leq \beta \|u\| \|v\| \quad \forall u, v \in H.$$

Let now $f : H \to \mathbb{R}$ be a bounded linear functional on $H$. Then there exists a unique element $u \in H$ such that

$$a[u, v] = \langle f, v \rangle$$

for all $v \in H$. 
2.6. EQUATIONS IN HILBERT SPACES

2.6.2 Fredholm Alternative

Assume now that $A : H \to H$ is a linear operator and consider equation (2.6.1) with $f \in H$. Let $A^* : H \to H$ denote the adjoint of $A$ which is defined through

$$(Au,v) = (u,A^*v) \quad \forall u, v \in H.$$  

Let now $v \in H$ belong to the null space $\mathcal{N}(A^*)$ of $A^*$ where

$$\mathcal{N}(A^*) = \{ v \in H : A^*v = 0 \}.$$  

Equation (2.6.2) implies that

$$(f,v) = 0 \quad \forall v \in \mathcal{N}(A^*).$$

Consequently, a necessary condition for the existence of a solution for (2.6.1) is that the right hand side of this equation is orthogonal to the null space of the adjoint operator of $A$.

The above formal argument can be made rigorous in the case where $A$ is a compact perturbation of the identity: $A = I - K$, with $K$ compact. A bounded operator $K : H \to H$ is compact if it maps bounded sets into precompact ones. Equivalently, $K$ is compact if and only if for every bounded sequence $\{u_n\}_{n=1}^{\infty} \subseteq H$, the sequence $\{Ku_n\}_{n=1}^{\infty}$ has a strongly convergent subsequence in $H$. We now study (2.6.1) in the case where $A = I - K$. The following result will be used repeatedly in these notes.

**Theorem 2.42. (Fredholm Alternative).** Let $H$ be a Hilbert space and let $K : H \to H$ be a compact operator. Then the following alternative holds.

(i) Either the equations

$$(I - K)u = f \quad \text{(2.6.3a)}$$

$$(I - K^*)U = F \quad \text{(2.6.3b)}$$

have unique solutions for every $f, F \in H$ or

(ii) the homogeneous equations

$$(I - K)u_0 = 0, \quad (I - K^*)U_0 = 0$$

have the same finite number of non-trivial solutions:

$$\dim(\mathcal{N}(I - K)) = \dim(\mathcal{N}(I - K^*)) < \infty.$$
In this case equations (2.6.3a) and (2.6.3b) have a solution if and only if

\[(f, U_0) = 0 \quad \forall \ U_0 \in \mathcal{N}(I - K^*)\]

and

\[(F, u_0) = 0 \quad \forall \ u_0 \in \mathcal{N}(I - K),\]

respectively.

### 2.7 Discussion and Bibliography

Most of the material presented in this chapter is very standard and can be found in many books on functional analysis and PDE. For more information on Banach, Hilbert and \(L^p\) spaces see e.g. [25, 84, 89, 129, 157].

Since \(L^1(\Omega)\) is not reflexive, a bounded sequence in \(L^1(\Omega)\) does not necessarily have any weakly convergent subsequences. A natural question that arises is under what conditions can we extract a weakly convergent sequence from a sequence in \(L^1(\Omega)\). The answer to this question is given through the Dunford–Pettis theorem: apart from boundendness we also need \textit{equi–integrability}. We refer to e.g. [40] for details. The non-reflexivity of \(L^1(\Omega)\) implies that this space cannot be characterized as the dual of a Banach space, and hence weak-* convergence is not a useful concept for this space. Weak-* convergence becomes, however, an extremely important concept for \((C_0(\Omega))^* =: M(\Omega)\), the space of \textbf{Radon measures} on \(\Omega^2\). In fact, a bounded sequence in \(L^1(\Omega)\) is weakly-* compact in \(M(\Omega)\): we can extract a weakly convergent subsequence that converges to an element of \(M(\Omega)\). Probabilists prefer to call weak-* convergence in \(C_0(\Omega)\) \textbf{weak convergence of probability measures}. This the most useful (and natural) concept for limit theorems in probability theory, and will be the topic of Section 3.5. The interested reader may also consult [75, 19].

Sobolev spaces of periodic functions can be defined using Fourier series. For example, the space \(H^1_{\text{per}}(Y)\) can be defined as

\[
H^1_{\text{per}}(Y) = \left\{ u : u = \sum_{k \in \mathbb{Z}^d} u_k e^{2\pi i k \cdot x}, \ u_k = u_{-k}, \sum_{k \in \mathbb{Z}^d} |k|^2 |u_k|^2 < \infty \right\}.
\]

Similarly we have

\[
H = \left\{ u \in H^1_{\text{per}}(Y) : u_k = 0 \right\}.
\]

\footnote{This space is (much) larger than \(L^1(\Omega)\), which is a proper subset of \(M(\Omega)\).}
Sobolev spaces of periodic functions are discussed in [129, 147]. An exhaustive treatment of Sobolev spaces is can be found in [2]. For many applications to the theory of PDEs, the material presented in [45, Ch. 5] is sufficient.

The concept of two–scale convergence was introduced by Nguetseng [108, 109] and later popularized and developed further by Allaire [3, 4]. Most of the results presented in Section 2.5 are take from [4]. We remark that the set of admissible test functions in Definition 2.28 is a proper subset of $L^2(\Omega \times \mathcal{Y})$; there are elements of $L^2(\Omega \times \mathcal{Y})$ which do not satisfy (2.5.1); see [4] for an example. Moreover, a strongly two–scale convergent sequence converges to its two–scale limit strongly in $L^2(\Omega)$, provided that the limit is regular enough. Notice however that the two–scale limit will not in general possess any further regularity. In fact, every function $u_0(x, y)$ in $L^2(\Omega \times \mathcal{Y})$ is attained as a two–scale limit of some sequence in $L^2(\Omega)$ [4, Lem. 1.13].

In Section 2.5.2 we considered sequences of functions which do not oscillate in time. The concept of two–scale convergence has been extended to cover the case of sequences of the form

$$u^\varepsilon = u \left( x, \frac{x}{\varepsilon}, t, \frac{t}{\varepsilon^p} \right),$$

where $p > 0$ and $u(x, y, t, \tau)$ is periodic in both $y$ and $\tau$. See [67]. The concept of two–scale convergence has also been extended to cover the case of functions which depend on more than two scales, i.e.

$$u^\varepsilon = u \left( x, \frac{x}{\varepsilon}, \frac{x}{\varepsilon^2}, \ldots \right).$$

See [5]. A concept similar to that of two–scale convergence has also been developed for non–periodic oscillations. See [101].

The use of appropriately chosen test functions in order to study asymptotic problems for PDE is a standard technique. See e.g. [42]. A very similar technique to that of two–scale convergence was introduced many years ago by Kurtz [85]. The approach is taken further in the perturbed test function method of Evans [44, 43].

The Fredholm alternative holds in normed spaces. That is, neither completeness, nor the inner product structure are necessary, see for instance [84, sec. 8.7].
2.8 Exercises

1. Let \( \{X, \| \cdot \| \} \) be a Banach space. Show that every strongly convergent sequence is weakly convergent.

2. Let \( \{X, \| \cdot \| \} \) be a finite dimensional Banach space. Show that every weakly convergent sequence is strongly convergent.

3. Let \( \Omega = (0, 1) \in \mathbb{R} \). Define
\[
    u(x) = \begin{cases} 
    x & \text{for } 0 \leq x \leq \frac{1}{2}, \\
    1 - x & \text{for } \frac{1}{2} \leq x \leq 1.
\end{cases}
\]
Show that the weak derivative of \( u(x) \) is
\[
    \frac{du}{dx} = \begin{cases} 
    1 & \text{for } 0 \leq x \leq \frac{1}{2}, \\
    -1 & \text{for } \frac{1}{2} \leq x \leq 1.
\end{cases}
\]
Is this function differentiable in the classical sense?

4. Consider the function \( u(x) \) from the previous exercise. Show that \( u(x) \in H^1_0(\Omega) \).

5. Recall \( H \) defined in (2.4.5). Use Fourier series to prove that the Poincaré inequality holds in \( H \):
\[
    \|u\|_{L^2} \leq \frac{1}{2\pi} \|\nabla u\|_{L^2} \quad \forall u \in H.
\]

6. Let \( \Omega = (0, 1) \in \mathbb{R}^1 \) and define \( u_\alpha(x) = |x|^\alpha \). For what values of \( \alpha \) does \( u_\alpha(x) \in H^1(\Omega) \)?

7. Prove the Poincaré inequality for a function \( f \in C^\infty(0, L), L > 0 \). Estimate the optimal value of the Poincaré constant \( C_L \). Show that
\[
    \lim_{L \to \infty} C_L = \infty.
\]
Interpret this result.

8. Consider a function \( u^\varepsilon(x) \in L^2(\Omega) \) which admits the following two–scale expansion
\[
    u^\varepsilon(x) = \sum_{j=0}^{N} u_j \left( x, \frac{x}{\varepsilon} \right),
\]
where $u_j(x, y) \in L^2(\Omega; C_{per}(Y))$, $j = 0, 1, \ldots, N$, and $\Omega$ is bounded domain in $\mathbb{R}^d$. Show that $u^\varepsilon \rightharpoonup u_0$ (this is a generalization of Lemma 2.32).
CHAPTER 2. ANALYSIS
Chapter 3

Probability Theory and Stochastic Processes

3.1 Set–Up

In this chapter we present some basic definitions and results from probability theory and from the theory of stochastic processes. We define the Wiener process (Brownian motion) and develop the Itô theory of stochastic integration. We summarize the basic properties of Martingales, and apply these to Itô integrals. When studying dimension reduction for Markovian problems we will mainly work in the context of weak convergence of probability measures. In particular, when applied on path-space, this will require us to study tightness (relative compactness) of probability measures. Hence in this chapter we will also summarize some basic limit theorems for stochastic processes, in particular weak convergence results in paths pace. We will see in later chapters that averaging and homogenization for SDE are essentially limit theorems for stochastic processes, and that they are intimately related to the theory of weak convergence of probability measures in metric spaces.

3.2 Probability and Expectation

A measurable space is a pair \((\Omega, \mathcal{F})\) where \(\Omega\) is a set and \(\mathcal{F}\) is a \(\sigma\)-algebra of subsets of \(\Omega\). Let \((\Omega, \mathcal{F})\) and \((E, \mathcal{G})\) be two measurable spaces. A function \(X : \Omega \to E\) is a random variable if \(X^{-1}(G) \in \mathcal{F}\) for all \(G \in \mathcal{G}\). The expected value of a random variable \(X\) is the Lebesgue integral 

\[
E[X] = \int_{\Omega} X(\omega) \lambda(\mathrm{d}\omega)
\]

where \(\lambda\) is a probability measure on \(\Omega\).

\footnote{A collection of subsets of a set \(\Omega\) is called a \(\sigma\)-algebra if it contains \(\Omega\) and is closed under the operations of taking complements and countable unions of its elements.}
36 CHAPTER 3. PROBABILITY THEORY AND STOCHASTIC PROCESSES

\( \Omega \mapsto E \) such that the event

\[ \{ \omega \in \Omega : X(\omega) \in A \} =: \{ X \in A \} \]

belongs to \( \mathcal{F} \) for arbitrary \( A \in \mathcal{G} \) is called a measurable function or random variable.

Let \( (\Omega, \mathcal{F}) \) be a measurable space. A function \( \mu : \mathcal{F} \mapsto [0, 1] \) is called a probability measure if \( \mu(\emptyset) = 1, \mu(\Omega) = 1 \) and \( \mu(\cup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} A_k \) for all sequences of pairwise disjoint sets \( \{A_k\}_{k=1}^{\infty} \in \mathcal{F} \). The triplet \( (\Omega, \mathcal{F}, \mu) \) is called a probability space. Let \( X \) be a measurable function (random variable) from \( (\Omega, \mathcal{F}, \mu) \) to \( (E, \mathcal{G}) \). The expectation of \( X \) is defined as

\[ \mathbb{E}[X] = \int_{\Omega} X(\omega) \, d\mu(\omega). \]

More generally, let \( f : E \mapsto \mathbb{R} \) be \( \mathcal{G} \)-measurable. Then,

\[ \mathbb{E}[f(X)] = \int_{\Omega} f(X(\omega)) \, d\mu(\omega), \]

provided that the integral exists. Every random variable from a probability space \( (\Omega, \mathcal{F}, \mu) \) to a measurable space \( (S, \mathcal{B}(S)) \)\(^2\) induces a probability measure on \( S \):

\[ \mu_X = \mathbb{P}X^{-1}(B) = \mu(\omega \in \Omega; X(\omega) \in B), \quad B \in \mathcal{B}(S). \]

The measure \( \mu_X \) is called the distribution of \( X \). We can use the distribution of a random variable to compute expectations and probabilities:

\[ \mathbb{E}[f(X)] = \int_{S} f(x) \, d\mu_X(x) \]

and

\[ \mathbb{P}[X \in G] = \int_{G} d\mu_X(x), \quad G \in \mathcal{B}(S). \]

When \( S = \mathbb{R}^d \) and we can write

\[ d\mu_X(x) = f(x) dx, \]

then we refer to \( f(x) \) as the probability density function (pdf) for \( X \).

When \( E = \mathbb{R}^d \) then by \( L^p(\Omega; \mathbb{R}^d) \), or simply \( L^p(\mu) \), we mean the Banach space of functions on \( \Omega \) with norm

\[ \|X\| = \left( \mathbb{E}|X|^p \right)^{1/p}. \]

\(^2\)Let \( U \) be a topological space. We will use the notation \( \mathcal{B}(U) \) to denote the Borel \( \sigma \)-algebra of \( U \), i.e. the smallest \( \sigma \)-algebra containing all open sets of \( U \).
3.2. PROBABILITY AND EXPECTATION

Example 3.1.  

i) Consider the random variable \( X : \Omega \mapsto \mathbb{R} \) with pdf

\[
\gamma_{\sigma,m}(x) := (2\pi\sigma)^{-\frac{1}{2}} e^{-\frac{(x-m)^2}{2\sigma}} dx.
\]

Such a random variable is a Gaussian or normal random variable. The mean is

\[
\mathbb{E}(X) = \int_{\mathbb{R}} x\gamma_{\sigma,m}(x) \, dx = m
\]

and the variance is

\[
\mathbb{E}[(X - m)^2] = \int_{\mathbb{R}} (x - m)^2 \gamma_{\sigma,m}(x) \, dx = \sigma.
\]

Since the mean and variance completely specify a Gaussian random variable on \( \mathbb{R} \), the Gaussian is commonly denoted by \( \mathcal{N}(m, \sigma) \).

ii) Let \( m \in \mathbb{R}^d \) and \( \Sigma \in \mathbb{R}^{d\times d} \) symmetric and positive definite. The random variable \( X : \Omega \mapsto \mathbb{R}^d \) with pdf

\[
\gamma_{\Sigma,m}(x) = ((2\pi)^n |\Sigma|)^{-\frac{1}{2}} e^{-\frac{1}{2} (x-m)^T \Sigma^{-1} (x-m)},
\]

with \( |\Sigma| = \det|\Sigma| \), is called the multivariate normal (Gaussian) distribution. The mean is

\[
\mathbb{E}(X) = m \quad (3.2.1)
\]

and the covariance matrix is

\[
\mathbb{E}((X - m) \otimes (X - m)) = \Sigma. \quad (3.2.2)
\]

Since the mean and covariance matrix completely specify a Gaussian random variable on \( \mathbb{R}^d \), the Gaussian is commonly denoted by \( \mathcal{N}(m, \Sigma) \).

Example 3.2. An exponential random variable \( T : \Omega \mapsto \mathbb{R}^+ \) with rate \( \lambda > 0 \) satisfies

\[
\mathbb{P}(T > t) = e^{-\lambda t}.
\]

We write \( T \sim \exp(\lambda) \). The related pdf is

\[
f_T(t) = \begin{cases} 
\lambda e^{-\lambda t}, & t \geq 0, \\
0, & t < 0. 
\end{cases} \quad (3.2.3)
\]
CHAPTER 3. PROBABILITY THEORY AND STOCHASTIC PROCESSES

Notice that
\[ E(T) = \int_{-\infty}^{\infty} tf_T(t)\,dt = \frac{1}{\lambda} \int_{0}^{\infty} (\lambda t)e^{-\lambda t}d(\lambda t) = \frac{1}{\lambda}. \]

If the times \( \tau_n = t_{n+1} - t_n \) are i.i.d random variables with \( \tau_0 \sim \text{exp}(\lambda) \) then, if \( t_0 = 0 \),
\[ t_n = \sum_{k=0}^{n-1} \tau_k \]
and it is straightforward to show that
\[ P(0 \leq t_k \leq t < t_{k+1}) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}. \quad (3.2.4) \]

Assume that \( E[|X|] < \infty \) and let \( G \) be a sub-\( \sigma \)-algebra of \( \mathcal{F} \). The conditional expectation of \( X \) with respect to \( G \) is defined to be the function \( \mathbb{E}[X|G] : \Omega \rightarrow E \) which is \( G \)-measurable and satisfies
\[ \int_G \mathbb{E}[X|G] \,d\mu = \int_G X \,d\mu \quad \forall G \in \mathcal{G}. \]

We can define the conditional probability \( P[x \in F|G] \) in a similar manner.

### 3.3 Stochastic Processes

Let \( T \) be an ordered set. A stochastic process is a collection of random variables \( X = \{X_t; t \in T\} \) from \( (\Omega, \mathcal{F}) \) to \( (E, \mathcal{G}) \). The measurable space \( \Omega, \mathcal{F} \) is called the sample space. The space \( (E, \mathcal{G}) \) is called the state space. In this book we will take the set \( T \) to be either \([0, +\infty)\) or \( \mathbb{N} \). Thus we will assume that all elements of \( T \) are nonnegative. The state space \( E \) will usually be \( \mathbb{R}^d \) equipped with the \( \sigma \)-algebra of Borel sets or, on some occasions, \( T^d \). When the ordered set \( T \) is clear from the context we will sometimes write \( \{X_t\} \) rather than \( \{X_t; t \in T\} \). Notice that \( X \) may be viewed as a function of both \( t \in T \) and \( \omega \in \Omega \). It is sometimes convenient to write \( X(t), x(t, \omega) \) or \( X_t(\omega) \) instead of \( X_t \).

The finite dimensional-distributions of a stochastic process are the \( \mathbb{R}^k \) valued random variables \( (X(t_1), X(t_2), \ldots, X(t_k)) \) for arbitrary positive integer \( k \) and arbitrary times \( t_i \in T \). A process is called stationary if all such collections of random variables are equal in distribution when translated in time: the distribution of \( (X(t_1), X(t_2), \ldots, X(t_k)) \) is equal to that of \( (X(s + t_1), X(s + t_2), \ldots, X(s + t_k)) \) for any \( s \) such that \( s + t_i \in T \).
The most important continuous–time stochastic process is the Wiener process – Brownian motion.

**Definition 3.3.**

i) A one dimensional Brownian motion \( W(t) : \mathbb{R}^+ \to \mathbb{R} \) is a real valued stochastic process with the following properties:

(a) \( W(0) = 0 \).

(b) \( W(t) \) is continuous.

(c) \( W(t) \) has independent increments. Furthermore, for every \( t > s \geq 0 \) \( W(t) - W(s) \) has a Gaussian distribution with mean 0 and variance \( t - s \). That is, the pdf of the random variable \( W(t) - W(s) \) is

\[
g(t, s; x) = \left( \frac{2\pi(t - s)}{2(t - s)} \right)^{-\frac{1}{2}} \exp \left( -\frac{x^2}{2(t - s)} \right). \tag{3.3.1} \]

ii) A \( d \)-dimensional Brownian motion \( W(t) : \mathbb{R}^+ \to \mathbb{R}^d \) is a collection of \( d \) independent one dimensional Brownian motions:

\[
W(t) = (W_1(t), \ldots, W_d(t)),
\]

where \( W_i(t), i = 1, \ldots, d \) are independent 1–dimensional Brownian motions. The density of the Gaussian random vector \( W(t) - W(s) \) is

\[
g(t, s; x) = \left( \frac{2\pi(t - s)}{2(t - s)} \right)^{-d/2} \exp \left( \frac{-\langle x, x \rangle}{2(t - s)} \right).
\]

Notice that, for the \( d \)-dimensional Brownian motion, and for \( I \) the \( d \times d \) dimensional identity, we have (see (3.2.1) and (3.2.2))

\[
\mathbb{E}(W(t)) = 0
\]

and

\[
\mathbb{E}[(W(t) - W(s)) \otimes (W(t) - W(s))] = (t - s)I.
\]

Moreover,

\[
\mathbb{E}[W(t)W(s)] = \min(t, s)I.
\]

Another important continuous time process is the **Poisson process**:

**Definition 3.4.** The Poisson process with intensity \( \lambda \), denoted by \( N(t) \), is a continuous–time stochastic process with independent increments satisfying

\[
\mathbb{P}[(N(t) - N(s)) = k] = \frac{e^{-\lambda(t-s)}(\lambda(t-s))^k}{k!}, \quad t > s \geq 0, \ k \in \mathbb{N}.
\]
Notice the connection to exponential random variables via (3.2.4).

Both Brownian motion and the Poisson process are homogeneous (or time-homogeneous): the increments between successive times $s$ and $t$ depend only on $t - s$. In these notes we will consider only homogeneous stochastic processes.

Let $(\Omega, \mathcal{F}, \mu)$ be a probability space, $(E, \rho)$ a metric space and let $T = [0, \infty)$. Let $\{X_t\}$ be a stochastic process from $(\Omega, \mathcal{F}, \mu)$ to $(E, \rho)$ with continuous sample paths. In other words, for all $\omega \in \Omega$ we have that $X_t(\omega)\in C(T, E)$. In this case $\{X_t\}$ can be thought of as a random variable on $(\Omega, \mathcal{F}, \mu)$ with state space $(C(T, E), \mathcal{B}(C(T, E)))$. The probability measure $\mathbb{P}X_t^{-1}$ on $(C(T, E), \mathcal{B}(C(T, E)))$ is called the law of $\{X_t\}$.

Although Brownian motion is continuous, many stochastic processes do not have continuous sample paths. We denote by $D(T, E)$ the space of right continuous processes $X_t : T \mapsto E$ with left limits\(^3\) (càdlàg processes). The space $D(T, E)$ will play an important role in the sequel.

Let $(\Omega, \mathcal{F})$ be a measurable space and $T$ an ordered set. Let $X = X_t(\omega)$ be a stochastic process from the sample space $(\Omega, \mathcal{F})$ to the state space $(E, \mathcal{G})$. It is a function of two variables, $t \in T$ and $\omega \in \Omega$. For a fixed $\omega \in \Omega$ the function $X_t(\omega), t \in T$ is the sample path of the process $X$ associated with $\omega$. Let $\mathcal{K}$ be a collection of subsets of $\Omega$. The smallest $\sigma$–algebra on $\Omega$ which contains $\mathcal{K}$ is denoted by $\sigma(\mathcal{K})$ and is called the $\sigma$–algebra generated by $\mathcal{K}$. Let $X_t : \Omega \mapsto E, t \in T$. The smallest $\sigma$–algebra $\sigma(X_t, t \in T)$ such that the family of mappings $\{X_t, t \in T\}$ is a stochastic process with sample space $(\Omega, \sigma(X_t, t \in T))$ and state space $(E, \mathcal{G})$ (that is, $\{X_t, t \in T\}$ is measurable), is called the $\sigma$–algebra generated by $\{X_t, t \in T\}$.

A filtration on $(\Omega, \mathcal{F})$ is a nondecreasing family $\{\mathcal{F}_t, t \in T\}$ of sub-$\sigma$–algebras of $\mathcal{F}$: $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$ for $s \leq t$. We set $\mathcal{F}_\infty = \sigma(\cup_{t \in T} \mathcal{F}_t)$. The filtration generated by a stochastic process $X_t$ is

$$\mathcal{F}_t^X := \sigma \left( X_s; s \leq t \right).$$

We say that a stochastic process $\{X_t; t \in T\}$ is adapted to the filtration $\{\mathcal{F}_t\} := \{\mathcal{F}_t, t \in T\}$ if for all $t \in T$, $X_t$ is an $\mathcal{F}_t$–measurable random variable.

**Definition 3.5.** Let $\{X_t\}$ be a stochastic process defined on a probability space $(\Omega, \mathcal{F}, \mu)$ with values in $E$ and let $\mathcal{F}_t^X$ be the filtration generated by $\{X_t\}$. Then

---

\(^3\)That is, $\lim_{s \to t^+} X_s = X_t$ and $\lim_{s \to t^-} X_s := X_{t-}$ exists for all $t \in [0, \infty)$.\]
{X_t} is a Markov process if

$$\mathbb{P}(X_{t+s} \in \Gamma | \mathcal{F}_t^X) = \mathbb{P}(X_{t+s} \in \Gamma | X_t) \quad (3.3.2)$$

for all \( s, t \in T \) and \( \Gamma \in \mathcal{B}(E) \). If \( \{\mathcal{F}_t\} \) is a filtration with \( \mathcal{F}_t^X \subset \mathcal{F}_t \), \( t \in T \), then \( \{X_t\} \) is a Markov process with respect to \( \{\mathcal{F}_t\} \) if (3.3.2) holds with \( \{\mathcal{F}_t^X\} \) replaced by \( \{\mathcal{F}_t\} \).

Roughly speaking the statistics of \( X_{t+s} \) for \( s \geq 0 \) are completely determined once \( X_t \) is known; information about \( X_{t-s} \) for \( s > 0 \) is superfluous. With a Markov process \( \{X_t\} \) we can associate a function \( P : T \times T \times E \times \mathcal{B}(E) \to \mathbb{R}^+ \) defined through the relation

$$\mathbb{P} \left[ X_{t+s} \in \Gamma | \mathcal{F}_t^X \right] = P(s, t, X_t, \Gamma),$$

for all \( t, s \in T \), \( \Gamma \in \mathcal{B}(E) \). The transition function \( P(s, t, x, \Gamma) \) satisfies the Chapman–Kolmogorov equation

$$\int_E P(s, t, x, dy)P(t, y, \Gamma) = P(s, v, x, \Gamma), \quad (3.3.3)$$

for all \( x \in E \), \( \Gamma \in \mathcal{B}(E) \), \( s < t < v \). A Markov process is homogeneous if

$$P(s, t, x, \Gamma) = P(0, t-s, x, \Gamma).$$

In this case we simplify the notation by setting \( P(0, t, \cdot, \cdot) = P(t, \cdot, \cdot) \) and the Chapman–Kolmogorov equation becomes

$$\int_E P(s, x, dy)P(t, y, \Gamma) = P(t+s, x, \Gamma).$$

Let \((E, \rho)\) be a metric space and let \( \{X_t\} \) be an \( E \)-valued homogenous stochastic process. Then we can define a semigroup of operators through

$$T(t)f(x) = \int f(y)P(t, x, dy)$$

for all \( f(x) \in C_b(E) \). The generator \( \mathcal{L} \) of the semigroup \( T(t) \) is called the generator of the Markov process \( \{X_t\} \). Conversely, let \( T(t) \) be a contraction semigroup with \( \mathcal{D}(T(t)) \subset C_b(E) \), closed. Then, under mild technical hypotheses, there is an \( E \)-valued Markov process \( \{X_t\} \) associated with \( T(t) \) defined through

$$\mathbb{E}[f(X(t+s)|\mathcal{F}_t^X)] = T(t)f(X(t))$$
for all $s, t \geq 0$ and $f \in D(T(t))$. A Markov process is **ergodic** if the semigroup $T(t)$ is such that $T(t) - I$ has only constants in its null-space. Under some additional compactness assumptions, the Markov process then has an invariant measure $\mu$ with the property that, if $X_0$ is distributed according to $\mu$, then so is $X_t$ for all $t > 0$. The resulting stochastic process, with $X_0$ distributed in this way, is stationary.

**Example 3.6.** Brownian motion is a time-homogeneous Markov process. The transition function is the Gaussian defined in Example 3.1:

$$P(t, x, dy) = \gamma_{t,x}(y)dy.$$  

The semigroup associated to the standard Brownian motion is the heat semigroup $T(t) = e^{\frac{t}{2}\Delta}$. The generator of this Markovian process is $\frac{1}{2}\Delta$.

**Example 3.7.** The Poisson process is a homogeneous Markov process.

As mentioned above, it is sometimes useful to view a stochastic process $X_t$ as a function of two variables $t \in T$ and $\omega \in \Omega : X(t, \omega)$. In this context it is then of interest to look at Banach space-valued spaces, as in the previous chapter (see Definition 2.22 and the discussion following it).

**Example 3.8.** Consider a stochastic process $X_t$ taking values in the space of real-valued continuous functions $C([0, T], \mathbb{R})$. We define $L^p(\Omega, C([0, T], \mathbb{R}))$ to be the Banach space equipped with norm $\|X\|$ given by

$$\|X\|^p = \mathbb{E}\left(\sup_{t \in [0,T]} \|X(t)\|^p\right).$$  

Notice that this is equal to

$$\|X\|^p = \mathbb{E}\left(\sup_{t \in [0,T]} \|X(t)\|^p\right)$$  

and the norm is usually written this way.

### 3.4 Martingales and Stochastic Integrals

#### 3.4.1 Martingales

**Definition 3.9.** Let $\{F_t\}$ be a filtration of the probability space $(\Omega, \mathcal{F}, \mu)$ and let $\{X_t\}$ be a stochastic process adapted to $\{F_t\}$. We will say that $\{X_t\}$ is an $\{F_t\}$--martingale if

$$\mathbb{E}[X_{t+s}|F_t] = X_t, \quad t, s \in T.$$
A martingale $M_t$ is square–integrable if $\mathbb{E}(M_t^2) < \infty$ for all $t \in T$.

**Example 3.10.** By Definition 3.3 Brownian motion is a martingale with respect to the filtration generated by itself.

**Definition 3.11.** Let $M_t$ be a square integrable $\mathcal{F}_t$–martingale with sample paths in $D([0, \infty), \mathbb{R}^d)$. Then there exists an increasing process with sample paths in $D([0, \infty), \mathbb{R}^d)$, called the quadratic variation and denoted by $\langle M, M \rangle_t$, such that for each $t \geq 0$ and each sequence of partitions $\{u_k^{(n)}\}$ of $[0, t]$ with $\max_k(u_{k+1}^{(n)} - u_k^{(n)}) \to 0$,

$$
\sum_k \left( M(u_{k+1}^{(n)}) - M(u_k^{(n)}) \otimes M(u_{k+1}^{(n)}) - M(u_k^{(n)}) \right) \to \langle M, M \rangle_t, \text{ in } L^1(\mu).
$$

**Example 3.12.**

i) The quadratic variation of the one dimensional standard Brownian motion is $t$. Let $W(t)$ be the standard Brownian motion in $\mathbb{R}^d$. Then

$$
\langle W, W \rangle_t = t.
$$

ii) Let $X_t$ be a Markov process with generator $L$ and let $f \in \mathcal{D}(L)$, the domain of definition of $L$. Then

$$
M_t = f(X_t) - f(X_0) - \int_0^t (Lf)(X_s) \, ds
$$

is a martingale. Assume that $f^2 \in \mathcal{D}(L)$. Then the quadratic variation of $M_t$ is given by the formula

$$
\langle M, M \rangle_t = \int_0^t \left( (Lf^2)(X_s) - 2f(X_s)(Lf)(X_s) \right) \, ds.
$$

Every continuous martingale satisfies Doob’s inequality

$$
\mathbb{E} \left[ \sup_{0 \leq t \leq T} |M_t| \geq \lambda \right] \leq \frac{1}{\lambda^p} \mathbb{E}|M_T|^p, \quad (3.4.2)
$$

for all $\lambda > 0$, $p \geq 1$, $T \geq 0$. Furthermore

$$
\mathbb{E} \sup_{0 \leq t \leq T} |M_t|^p \leq \left( \frac{p-1}{p} \right)^p \mathbb{E}|M_T|^p, \quad (3.4.3)
$$

for all $p > 1$, $T \geq 0$. 

3.4.2 The Itô Stochastic Integral

For the rigorous analysis of stochastic differential equations it is necessary to define stochastic integrals of the form

\[ I(t) = \int_0^t f(s) \, dW(s), \]  
(3.4.4)

where \( W(t) \) is a standard one dimensional Brownian motion and \( f(t) \) is a random process, adapted to the filtration \( \mathcal{F}_t \) generated by the \( W(t) \), and such that

\[ \mathbb{E} \left( \int_0^T f(s)^2 \, ds \right) < \infty. \]

Throughout we use the Itô interpretation of the stochastic integral. The Itô stochastic integral is defined as the \( L^2 \)-limit of the Riemann sum approximation of (3.4.4)

\[ I(t) := \lim_{N \to \infty} \sum_{k=1}^{N-1} f(t_{k-1}) \left( W(t_k) - W(t_{k-1}) \right), \]
(3.4.5)

where \( t_n = n \Delta t \) and \( N \Delta t = t \). Notice that the function \( f(t) \) is evaluated at the left end of each interval \([t_{n-1}, t_n]\) in (3.4.5). The multidimensional Itô integral is defined in a similar way. The resulting Itô stochastic integral \( I(t) \) is a.s. continuous in \( t \). These ideas are readily generalized to the case where \( W(s) \) is a standard \( d \) dimensional Brownian motion and \( f(s) \in \mathbb{R}^{m \times d} \) for each \( s \). The integral then satisfies the Itô isometry

\[ \mathbb{E}[I(t)]^2 = \int_0^t \mathbb{E}[f(s)]^2 \, ds. \]

Furthermore, it is a martingale:

\[ \mathbb{E}I(t) = 0 \]

and

\[ \mathbb{E}[I(t)|\mathcal{F}_s] = I(s) \quad \forall \, t \geq s, \]

where \( \mathcal{F}_s \) denotes the filtration generated by \( W(s) \).

Example 3.13. Consider the Itô stochastic integral

\[ I(t) = \int_0^t f(s) \, dW(s). \]
This is a martingale with quadratic variation

\[ \langle I, I \rangle_t = \int_0^t (f(s))^2 \, ds, \]

where \( f, W \) are scalar-valued.

We now state a converse result: every square integrable martingale can be expressed in terms of an Itô integral.

**Theorem 3.14.** Let \((\Omega, \mathcal{F}, \mu)\) be a probability space and let \(B(t)\) be a standard, \(d\)-dimensional Brownian motion on it. Let \(\mathcal{F}_t\) be the filtration generated by the Brownian motion and let \(M_t\) be a square integrable \(\mathcal{F}_t\)-martingale with respect to \(\mu\). Then there exists a \(B([0, \infty)) \times \Omega\)-measurable function \(f \in L^2([0, \infty) \times \Omega)\) such that

\[ M_t = \mathbb{E}[M_0] + \int_0^t f(s, \omega) \, dB_s. \]

One of the most useful features of martingales is that they satisfy various inequalities. These inequalities are particularly important when proving limit theorems. Perhaps the most important inequality satisfied by martingales is the following.

**Theorem 3.15. Burkholder-Davis-Gundy Inequality** Consider the Itô stochastic integral (3.4.4) with quadratic variation \(\langle I, I \rangle_t\). For every \(p > 0\) there are constants \(C^\pm\) such that

\[ C^- \left( \mathbb{E} \langle I, I \rangle_t \right)^{p/2} \leq \mathbb{E} \left( \sup_{0 \leq s \leq t} |I(s)|^p \right) \leq C^+ \left( \mathbb{E} \langle I, I \rangle_t \right)^{p/2}. \]

The proof of this theorem is based on Itô’s formula and the Doob Martingale inequality 3.4.3.

**3.4.3 The Stratonovich Stochastic Integral**

In addition to the Itô stochastic integral, the following Stratonovich integral is also useful. It is defined as the \(L^2\)-limit of a different Riemann sum approximation of (3.4.4) namely

\[ I(t) := \lim_{N \to \infty} \sum_{k=1}^{N-1} \frac{1}{2} \left( f(t_{k-1}) + f(t_k) \right) (W(t_k) - W(t_{k-1})), \quad (3.4.6) \]
where \( t_n = n \Delta t \) and \( N \Delta t = t \). Notice that the function \( f(t) \) is evaluated at both endpoints of each interval \([t_{n-1}, t_n]\) in (3.4.6). The multidimensional Stratonovich integral is defined in a similar way. The resulting integral is written as

\[
I_{\text{strat}}(t) = \int_0^t f(s) \circ dW(s).
\]

The resulting limit differs from the Itô integral. The situation is different from that arising in the theory of Riemann integration for functions of bounded variation; in that case the points in \([t_{k-1}, t_k]\) where the integrand is evaluated do not effect the definition of the integral. In the case of integration against Brownian motion, which does not have bounded variation, the definitions differ. However, when \( f \) and \( W \) are correlated through an SDE, then a formula exists to convert between them – see Chapter 6.

### 3.5 Weak Convergence of Probability Measures

A type of convergence that is very often used in probability theory is that of weak convergence of probability measures.

**Definition 3.16.** Let \((S, \rho)\) be a metric space with Borel \(\sigma\)-field \(\mathcal{B}(S)\). Let \(\{\mu_n\}_{n=1}^{\infty}\) be a sequence of probability measures on \((S, \mathcal{B}(S))\), and let \(\mu\) be another measure on this space. We say that \(\{\mu_n\}_{n=1}^{\infty}\) converges weakly to \(\mu\), and write \(\mu_n \Rightarrow \mu\) if

\[
\lim_{n \to \infty} \int_S f(s) \, d\mu_n(s) = \int_S f(s) \, d\mu(s),
\]

for every \(f \in C_b(S)\).

**Definition 3.17.** Let \((\Omega_n, \mathcal{F}_n, \mu_n)_{n=1}^{\infty}\) be a sequence of probability spaces and let \((S, \rho)\) be a metric space. Let \(X_n : \Omega_n \mapsto S, n = 1, \ldots, \infty\) be a sequence of random variables. Assume that \((\Omega, \mathcal{F}, \mu)\) is another probability space and let \(X : \Omega \mapsto S\), be another random variable. We will say that \(\{X_n\}_{n=1}^{\infty}\) converges to \(X\) in distribution and write \(X_n \Rightarrow X\) if the sequence of measures \(\{\mathbb{P}^{-1}X_n\}_{n=1}^{\infty}\) converges weakly to the measure \(\mathbb{P}^{-1}\).

In other words, \(X_n \Rightarrow X\) if and only if

\[
\lim_{n \to \infty} \mathbb{E}_n f(X_n) = \mathbb{E} f(X)
\]

for all \(f \in C_b(S)\).
Example 3.18. (The central limit theorem). Let \( \{\xi_n\}_{n=1}^\infty \) be a sequence of i.i.d. random variables with mean zero and variance 1, and define

\[
S_n := \sum_{k=1}^n \xi_k.
\]

Then the sequence

\[
X_n := \frac{1}{\sqrt{n}} S_n
\]

converges in distribution to a standard normal variable.

There are various types of convergence that are useful in the study of limit theorems for random variables. Let \( \{X_n\}_{n=1}^\infty \) be a sequence of random variables.

We will say that the sequence \( \{X_n\}_{n=1}^\infty \) converges in probability to a random variable \( X \) if

\[
\lim_{n \to \infty} P(|X_n - X| > \varepsilon) \to 0.
\]

We will say that \( \{X_n\}_{n=1}^\infty \) converges almost surely (or with probability 1) if

\[
P(\lim_{n \to \infty} X_n = X) = 1.
\]

Finally, we say that \( \{X_n\}_{n=1}^\infty \) converges in \( p \)-th mean (or in \( L^p \)) provided that

\[
\lim_{n \to \infty} \mathbb{E}|X_n - X|^p = 0
\]

Example 3.19. (The strong law of large numbers). Let \( \{\xi_n\}_{n=1}^\infty \) be a sequence of i.i.d. random variables with mean 1. Define

\[
X_n = \frac{1}{n} \sum_{k=1}^n \xi_k.
\]

Then the sequence \( X_n \) converges to 1 almost surely. Assume furthermore that \( \xi_n \in L^2 \). Then \( X_n \) converges to 1 also in \( L^2 \).

Remark 3.20. There are relations between the different notions of convergence for random variables: for example almost sure convergence implies weak convergence; convergence in probability implies weak convergence.

Weak convergence is preserved under continuous mappings.
CHAPTER 3. PROBABILITY THEORY AND STOCHASTIC PROCESSES

Theorem 3.21. Let \((\Omega_n, \mathcal{F}_n, \mu_n)_{n=1}^\infty\) be a sequence of probability spaces and let \((S_i, \rho_i)\) be metric spaces \(i = 1, 2\). Let \(X_n : \Omega_n \mapsto S, n = 1, \ldots, \infty\) be a sequence of random variables. Assume that \((\Omega, \mathcal{F}, \mu)\) is another probability space and let \(X : \Omega \mapsto S\) be another random variable. If \(f : S_1 \mapsto S_2\) is continuous then \(\{f(X_n)\}_{n=1}^\infty\) converges to \(f(X)\) in distribution if \(\{X_n\}_{n=1}^\infty\) converges to \(X\) in distribution.

Example 3.22. Let \(S_1 = C[0,1]\) and \(S_2 = \mathbb{R}\). The function \(f : S_1 \mapsto S_2\) continuous. Hence, theorem 3.21 applies and we have that \(X_n \Rightarrow X\) in \(S_1\) implies that \(\sup_{t \in [0,1]} X_n(t) \Rightarrow \sup_{t \in [0,1]} X(t)\) in \(S_2\).

Let \((\Omega, \mathcal{F}, \mu)\) be a probability space, \(T = [0, \infty)\) and \((E, \rho)\) a metric space. Let \(\{X^n_t\}_{n=1}^\infty\), by a family of stochastic processes, and \(X_t\) another stochastic process, all with sample paths in \(C(T, E)\). We will say that the sequence \(\{X^n_t\}_{n=1}^\infty\) converges weakly to \(X_t\) and write \(X^n_t \Rightarrow X_t\) if the sequence of probability measures \(\mathbb{P}(X^n_t)^{-1}\) converges weakly to the probability measure \(\mathbb{P}X^{-1}_t\) on \((C(T, E), \mathcal{B}(C(T, E)))\). Sometimes we will also say that \(\{X^n_t\}_{n=1}^\infty\) converges weakly in \(C(T, E)\).

Example 3.23. Consider the situation of Example 3.18 and let \(S_n\) be given by (3.5.1). Let \(\lfloor t \rfloor\) denote the integer part of a real number \(t\) and define the continuous–
time process \(Y_t = S_{\lfloor t \rfloor} + (t - \lfloor t \rfloor)\xi_{\lfloor t \rfloor + 1}\). (3.5.2)

The process \(Y_t\) has continuous paths. Define the sequence of stochastic processes

\[
X^n_t = \frac{1}{\sqrt{n}} Y_{nt}, \quad t \geq 0.
\]

Then \(\{X^n_t\}_{n=1}^\infty\) converges weakly in \(C([0,\infty), \mathbb{R})\) to a standard one–dimensional Brownian motion.

Let \((\Omega, \mathcal{F})\) be a measurable space. We will denote by \(M_1(\Omega, \mathcal{F})\) the space of all probability measures on \((\Omega, \mathcal{F})\). Assume that \(\Omega\) is a topological space and let \(\mathcal{B}(\Omega)\) denote the Borel \(\sigma\)-algebra. We will write \(M_1(\Omega) := M_1(\Omega, \mathcal{B}(\Omega))\).

Definition 3.24. Let \((E, \rho)\) be a metric space and let \(M \subseteq M_1(E)\). We will say that \(M\) is relatively compact if every sequence of elements of \(M\) contains a weakly
3.5. WEAK CONVERGENCE OF PROBABILITY MEASURES

We will say that $M$ is tight if for every $\varepsilon > 0$ there exists a compact subset $K \subseteq E$ such that $\mu(K) \geq 1 - \varepsilon \forall \mu \in M$.

**Theorem 3.25.** Let $(E, \rho)$ be a complete, separable metric space. Then a set $M \subseteq M_1((E, \mathcal{B}(E)))$ is relatively compact if and only if it is tight.

**Theorem 3.26.** Let $\{X^n_t\}$ be a sequence of continuous stochastic processes on $(\Omega, \mathcal{F}, \mu)$ satisfying the following conditions:

i. $\exists \nu > 0 : \sup_{1 \leq n \leq \infty} \mathbb{E}|X^n_0|^\nu < \infty$,

ii. $\exists \alpha, \beta > 0$ and $C = C(T) > 0$:

\[
\sup_{1 \leq n \leq \infty} \mathbb{E}|X^n_t - X^n_s|^\alpha \leq C_T|t - s|^{1+\beta}, \quad \forall T > 0, \ t, s \in [0, T]
\]

Let $\mu^n$ denote the law of the process $X^n_t$ for $n = 1, \ldots, \infty$. Then the sequence $\{\mu^n\}_{n=1}^\infty$ (which forms a sequence of probability measures on $(C([0, \infty), \mathcal{B}(C([0, \infty))))$ is tight.

Tightness and convergence of finite dimensional distributions imply convergence in law.

**Theorem 3.27.** Let $\{X^n_t\}_{n=1}^\infty, \{X_t\}$ be continuous stochastic processes on a probability space $(\Omega, \mathcal{F}, \mu)$ with state space $(E, \rho)$ and assume that the following conditions are satisfied.

i. $\{X^n_t\}_{n=1}^\infty$ is tight.

ii. The finite dimensional distributions of $\{X^n_t\}_{n=1}^\infty$ converge to those of $\{X_t\}$.

Then the sequence $\{X^n_t\}_{n=1}^\infty$ converges weakly in $C([0, \infty), E)$ to $\{X_t\}$.

Let $(E, \rho)$ be a metric space. In Section 3.3 we introduced the space $D([0, \infty), E)$, the space of right continuous processes $X_t : [0, \infty) \to E$. The space of continuous functions $C([0, \infty), E)$ is a closed subset of $D([0, \infty), E)$. If $E$ is separable then so is $D([0, \infty), E)$. If $(E, \rho)$ is complete then there exists a metric $d$ on $D([0, \infty), E)$ such that $(D([0, \infty), E), d)$ is a complete metric space.
Example 3.28. Consider the situation of Examples 3.18 and 3.23 and let \( S_n \) be given by (3.5.1). Define the continuous–time process

\[
Z_t = S_{\lfloor t \rfloor}.
\]

Notice that \( Z_t \in D([0, \infty), \mathbb{R}) \). Define the sequence of stochastic processes

\[
X_n^t = \frac{1}{\sqrt{n}} Z_{nt}, \quad t \geq 0.
\]

Then \( \{X_n^t\}_{n=1}^\infty \) converges weakly in \( D([0, \infty), \mathbb{R}) \) to a standard one–dimensional Brownian motion.

Remark 3.29. Let \( C_E := C([0, \infty), E) \) and \( D_E := D([0, \infty), E) \). Assume that \( \{\mu_n\}_{n=1}^\infty, \mu \in M_1(D_E) \), and assume that \( \mu_n(C_E) = \mu(C_E) = 1 \) \( \forall n = 1, \ldots, \infty \). Define \( \{Q_n\}_{n=1}^\infty, Q \in M_1(C_E) \) by \( Q_n = \mu_n|_{\mathcal{B}(C_E)}, \ n = 1, \ldots, \infty \) and \( Q = \mu|_{\mathcal{B}(C_E)} \). Then \( Q_n \) converges to \( Q \) weakly in \( C_E \) if and only if \( \mu_n \) converges to \( \mu \) weakly in \( D_E \). In particular, for processes with continuous sample paths weak convergence in \( D_E \) is a necessary and sufficient condition for weak convergence in \( C_E \).

In Examples 3.18 and 3.23 we studied the classical central limit theorem and invariance principle (functional central limit theorem), respectively. These results were stated for sums of independent random variables. Similar results hold even for dependent random variables, provided that the dependence between the random variables is not too strong. In many instances the martingale property is sufficient.

Theorem 3.30. Let \( (\Omega, \mathcal{F}, \mu) \) be a probability space and let \( \{\mathcal{F}_j, j \geq 1\} \) be a filtration. Let \( \{Z_j, j \geq 1\} \) be a sequence of stationary, ergodic random variables such that

\[
\mathbb{E}[Z_1]^2 = \sigma^2
\]

and

\[
\mathbb{E}[Z_{k+1}|\mathcal{F}_k] = 0. \quad (3.5.3)
\]

Define

\[
S_n = \sum_{k=1}^{n} Z_k. \quad (3.5.4)
\]
Then
\[ X^n = \frac{1}{\sqrt{n}} S_n \]
converges in distribution to a Gaussian variable with mean 0 and variance \( \sigma^2 \).
Furthermore, the process
\[ X^n_t = \frac{1}{\sqrt{n}} S_{[nt]} + \frac{1}{\sqrt{n}} (nt - [nt]) Z_{[nt]+1} \]
converges weakly in \( C([0, \infty), \mathbb{R}) \) to a Brownian motion with variance \( \sigma^2 \). Finally, the process
\[ X^n_t = \frac{1}{\sqrt{n}} S_{[nt]} \]
converges weakly in \( D([0, \infty), \mathbb{R}) \) to a Brownian motion with variance \( \sigma^2 \).

Notice that the condition (3.5.3) implies that \( S_N \) defined in (3.5.4) is an \( \mathcal{F}_j \)-martingale.

The above theorem is also valid in arbitrary dimensions, that is, in the case where \( \{Z_j, j \geq 1\} \) is a vector–valued sequence of stationary, ergodic random variables. In this case the covariance matrix of the limiting Brownian motion is the covariance matrix of \( Z_1 \).

A result similar to that of Theorem 3.30 can be proved for continuous–time martingales.

**Theorem 3.31. (The Martingale Central Limit Theorem).** Let \( \{M(t) : R^+ \mapsto R\} \) be a right–continuous square integrable martingale on a probability space \( (\Omega, \mathcal{F}, \mu) \) with respect to a given increasing filtration \( \{\mathcal{F}_t : t \geq 0\} \) and let \( \langle M \rangle_t \) denote its quadratic variation. Assume that

i) \( M(0) = 0 \);

ii) The increments of \( M(t) \) are stationary;

iii) The quadratic variation of \( M(t) \) converges in \( L^1(\mu) \) to some positive constant \( \sigma \):

\[ \lim_{t \to \infty} \mathbb{E} \left[ \left| \frac{\langle M \rangle_t}{t} - \sigma \right| \right] = 0. \]  
(3.5.5)

Then the process \( \frac{1}{\sqrt{t}} M_t \) converges in distribution to a \( \mathcal{N}(0, \sigma) \) random variable. Furthermore, for every fixed \( t \in [0, T], \ T < \infty \) the rescaled martingale

\[ M^{\varepsilon}(t) := \frac{1}{\varepsilon} M \left( \frac{t}{\varepsilon^2} \right) \]
converges weakly in \(D([0, T], \mathbb{R})\) to a Brownian motion with variance \(\sigma\).

A similar result is valid in arbitrary dimensions. Notice that, on account of Remark 3.29, if the martingale has continuous sample paths then the above theorem provides us with weak convergence in \(C([0, T], \mathbb{R})\). The assumptions of the above theorem can be weakened considerably as we now show.

**Theorem 3.32.** Let \((\Omega, \mathcal{F}, \mu)\) be a probability space and let \(\{\mathcal{F}_n, n \geq 1\}\) be a filtration. Let \(\{M_n\}\) be an \(\{\mathcal{F}_n\}\)-local martingale with sample paths in \(C([0, \infty), \mathbb{R}^d)\) and \(M_n(0) = 0\). Let \(A_n := \langle A_{ij}^n \rangle\) be the quadratic variation of the martingale.

\[ A_n(t) = \langle M_n, M_n \rangle_t. \]

Assume that there exists a nonnegative definite symmetric matrix \(C = (C_{ij})\) such that \(A_n(t)\) converges to \(Ct\) in probability. Then \(M_n\) converges weakly in \(C([0, \infty), \mathbb{R}^d)\) to a \(d\)-dimensional Brownian motion with covariance matrix \(C\).

Upon combining the fact that stochastic integrals are martingales, together with an ergodicity assumption, we obtain the following corollary.

**Corollary 3.33.** Let \(x(t) : \mathbb{R}^+ \rightarrow \mathbb{Z}\) be a stationary, ergodic Markov process with invariant measure \(\mu(dx)\), and let \(f(\cdot) : \mathbb{R}^d \rightarrow \mathbb{R}^m\) be a smooth bounded function. Define

\[ I(t) = \int_0^t f(x(s)) dW(s) \]

Then, for every finite \(T > 0\), the rescaled stochastic integral

\[ I^\varepsilon(t) := \varepsilon I \left( \frac{t}{\varepsilon^2} \right) \]

converges in distribution to an \(m\) dimensional Brownian motion with covariance matrix

\[ \Sigma = \int_{\mathbb{Z}} f(x) \otimes f(x) \mu(dx). \]

**Proof.** We have to check that the martingale \(I(t)\) satisfies the assumptions of Theorem 3.31. We clearly have that \(I(0) = 0\). Furthermore, the stationarity of the process \(x(t)\) implies that \(I(t)\) has independent increments. By ergodicity

\[ \lim_{t \to \infty} \frac{1}{t} \langle I \rangle_t = \lim_{t \to \infty} \frac{1}{t} \int_0^t f(x(s)) f(x(s))^T ds = \int_{\mathbb{Z}} f(x) f(x)^T \mu(dx) \text{ in } L^1(\mu). \]

Hence, Theorem 3.31 applies. \(\square\)
3.6 Discussion and Bibliography

An excellent reference on convergence of probability measures is [19]. Various convergence results for sequences of Markov processes can be found in [41]. The text [55] has good background on stochastic processes and weak convergence for probability measures on path space. For discussion about relationships between different modes of convergence see [60].

Continuous local martingales can be expressed as time changed Brownian motions. This is the Dambis–Dubins–Schwarz theorem [75, Thm 3.4.6]. \( M = \{ M_t, \mathcal{F}_t ; 0 \leq t < \infty \} \) be a Martingale satisfying \( \lim_{t \to \infty} \langle M_t, M_t \rangle_t = \infty \), \( \mathbb{P} \) a.s. Define, for each \( 0 \leq s < \infty \), the stopping time

\[
T(s) = \inf\{ t \geq 0 ; \langle M, M \rangle_t > s \}.
\]

Then the time–changed process

\[
B_s = M_{T(s)}, \quad \mathbb{G}_s = \mathcal{F}_{T(s)}, \quad 0 \leq s < \infty,
\]

is a standard one–dimensional Brownian motion. In particular \( \mathbb{P} \) a.s:

\[
M_t = B_{\langle M, M \rangle_t}.
\]

Notice that the quantities \( B_t \) and \( M_t \) are in general highly correlated and that they are not adapted to the same filtration (because of the time change).

Weak convergence results of the form of Theorem 3.32 are valid even when the limiting process is not a Brownian motion but a more general diffusion process. See [41, Sec.7.4]. The proof of the martingale central limit theorem can be found in [41, ch. 7]. See also [87, 86].

The martingale central limit theorem leads to a general central limit theorem for additive functionals of Markov Processes: let \( y(t) \) be an ergodic Markov process with generator \( L \) and invariant measure \( \mu(dy) \). Consider the integral (i.e. additive) functional of \( y(t) \)

\[
x(t) = x_0 + \int_0^t f(y(s)) \, ds.
\]

Then, the re–scaled process

\[
x^\varepsilon(t) = \varepsilon x(t/\varepsilon^2)
\]
converges weakly to a Brownian motion with variance $\Sigma$, provided that the Poisson equation

$$-L\phi = f$$

is well posed, in some appropriate (weak) sense. The variance of the limiting Brownian motion is given by the Dirichlet form of the generator $L$, evaluated at the solution of (3.6.1):

$$\Sigma = \int (-L\phi) \otimes \phi \, \mu(dy).$$

To see this, notice that (3.4.1) implies that the rescaled process $x^\varepsilon(t)$ satisfies

$$x^\varepsilon(t) = \varepsilon \left( x_0 - \phi(X(t/\varepsilon^2)) + \phi(X(0)) \right) + \varepsilon M(t/\varepsilon^2),$$

where $M(t)$ is a martingale. The first term on the right hand side of the above equation tends to 0, provided that $\phi$ satisfies appropriate estimates. The second term converges to a Brownian motion with variance $\Sigma$, by the martingale central limit theorem. This theorem was proved for reversible Markov processes in [82]. See also [118]. Various extensions of this result have been proved. See [87] and the references therein.

### 3.7 Exercises

1. Let $W(t) : \mathbb{R}^+ \to \mathbb{R}$ be a standard Brownian motion. Calculate all moments of $W(t) - W(s)$, $t > s \geq 0$.

2. Let $W(t)$ be a standard Brownian motion in one dimension. Show that

$$\mathbb{E} \left( \left| \frac{\Delta W(t)}{\Delta t} \right| \right) = \sqrt{\frac{2}{\mu}} \frac{1}{\sqrt{\Delta t}}.$$  

   Deduce that the Brownian motion is nowhere differentiable a.s.

3. Let $W_t$ be a standard Brownian motion in $\mathbb{R}^d$ and let $\mathcal{F}_t$ denote the filtration generated by $W_t$. Prove that $W_t$ is an $\mathcal{F}_t$-martingale. Prove that $W_t - t$ is also an $\mathcal{F}_t$-martingale.

4. State the analogue of Theorem 3.30 in arbitrary dimensions.
Chapter 4

Ordinary Differential Equations

4.1 Set-Up

We study the ordinary differential equations

\[
\frac{dz}{dt} = h(z), \quad z(0) = z_0
\]  

(4.1.1)

where \( h : Z \rightarrow \mathbb{R}^d \). Typically \( Z = \mathbb{T}^d, \mathbb{R}^d \) or \( \mathbb{R}^l \oplus \mathbb{T}^d - l \). In later chapters we will often consider \( z = (x^T, y^T)^T \), with \( x \in \mathcal{X}, y \in \mathcal{Y} \). If \( Z = \mathbb{T}^d \) (resp. \( \mathbb{R}^d \)) then \( \mathcal{X} = \mathbb{T}^l \) (resp. \( \mathbb{R}^l \)) and \( \mathcal{Y} = \mathbb{T}^d - l \) (resp. \( \mathbb{R}^d - l \)). If \( Z = \mathbb{R}^l \oplus \mathbb{T}^d - l \) then \( \mathcal{X} = \mathbb{R}^l \) and \( \mathcal{Y} = \mathbb{T}^d - l \).

4.2 Existence and Uniqueness

When there is a unique solution to the initial value problem (4.1.1) for all initial data in \( Z \) and \( t \in \mathbb{R} \) we write

\[
z(t) = \varphi^t(z_0).
\]

Thus \( \varphi^t : Z \rightarrow Z \) is the solution operator and forms a one parameter group, that is

\[
\varphi^{t+s} = \varphi^t \circ \varphi^s \quad \forall t, s \in \mathbb{R} \quad \text{and} \quad \varphi^0 = \mathcal{I},
\]

where \( \mathcal{I} : Z \rightarrow Z \) denotes the identity operator.

In practice existence and uniqueness of solutions of (4.1.1) can be verified in a wide range of different scenarios. For ease of exposition we work within to the simplest scenario, namely when \( h \) is Lipschitz on \( Z \). This condition can be weakened when \textit{a priori} bounds on the solution prevent blow-up.
**Definition 4.1.** A function \( f : \mathbb{Z} \rightarrow \mathbb{R}^d \) is Lipschitz on \( \mathbb{Z} \) with Lipschitz constant \( L \) if
\[
|f(x) - f(y)| \leq L|x - y|, \quad \forall x, y \in \mathbb{Z}.
\]

We observe that a Lipschitz function is also linearly bounded:
\[
|f(x)| \leq |f(0)| + L|x|.
\]

**Theorem 4.2.** If \( h \) is Lipschitz on \( \mathbb{Z} \) then the ODE (4.1.1) has a unique solution \( z(t) \in C^1(\mathbb{R}, \mathbb{Z}) \) for any \( x_0 \in \mathbb{Z} \). Moreover
\[
|x(t) - y(t)| \leq e^{Lt}|x_0 - y_0|.
\]

**Proof.** The existence and uniqueness follows from a standard fixed point argument (Picard iteration). The bound (4.2.1) follows from
\[
\frac{1}{2} \frac{d}{dt}|x - y|^2 = \langle h(x) - h(y), x - y \rangle,
\]
and
\[
\langle h(x) - h(y), x - y \rangle \leq L|x - y|^2,
\]
using the Gronwall Lemma 4.3 below.\( \square \)

A useful tool for studying evolution equations is Gronwall’s lemma; it is also required to prove the preceding theorem.

**Lemma 4.3.** 1. (Differential form). Let \( \eta(t) \in C^1([0, T]; \mathbb{R}^+ \) satisfy the differential inequality
\[
\frac{d\eta(t)}{dt} \leq \phi(t)\eta(t) + \psi(t), \quad \eta(0) = \eta,
\]
where \( \phi(t), \psi(t) \in L^1([0, T]; \mathbb{R}^+) \). Then
\[
\eta(t) \leq \exp\left( \int_0^t \phi(s) \, ds \right) \left[ \eta + \int_0^t \psi(s) \, ds \right]
\]
for all \( t \in [0, T] \).

2. (Integral form). Assume that \( \xi(t) \in L^1([0, T]; \mathbb{R}^+) \) satisfies the integral inequality
\[
\xi(t) \leq \phi \int_0^t \xi(s) \, ds + \psi,
\]
for some positive constants \( \phi, \psi \). Then
\[
\xi(t) \leq \phi \left( 1 + \psi t e^{\phi t} \right).
\]
4.3. THE GENERATOR

Proof. 1. We multiply equation (4.2.2) by \( \exp \left( -\int_0^t \phi(s) \, ds \right) \) to obtain

\[
\frac{d}{dt} \eta(t) \exp \left( -\int_0^t \phi(s) \, ds \right) \leq \left( \psi(t) + \phi(t) \eta(t) \right) \exp \left( -\int_0^t \phi(s) \, ds \right).
\]

Consequently

\[
\frac{d}{dt} \left( \eta(t) \exp \left( -\int_0^t \phi(s) \, ds \right) \right) \leq \psi(t) \exp \left( -\int_0^t \phi(s) \, ds \right) \leq \psi(t),
\]

since \( \phi(t) \) is nonnegative. Integrating this inequality from 0 to \( t \) and multiplying through by \( \exp \left( \int_0^t \phi(s) \, ds \right) \) gives (4.2.3).

2. Define \( \eta(t) = \int_0^t \xi(s) \, ds \). Then \( \eta(t) \) satisfies the inequality

\[
\frac{d\eta}{dt} \leq \phi\eta + \psi.
\]

We apply the first part of the lemma with \( \eta(0) = 0 \) to deduce that

\[
\eta(t) \leq \psi t e^{\phi t}.
\]

Consequently

\[
\xi(t) \leq \phi\eta(t) + \psi \leq \psi \left( 1 + \phi t e^{\phi t} \right).
\]

\[\square\]

4.3 The Generator

It is often of importance to understand how functions of \( z(t) \) change with time. We may achieve this by using the generator \( \mathcal{L} \):

\[
(\mathcal{L}V)(z) = \langle h(z), \nabla V(z) \rangle. \tag{4.3.1}
\]

If \( z(t) \) solves (4.1.1) and \( V \in C^1(\mathbb{Z}, \mathbb{R}) \) then

\[
\frac{d}{dt} \left\{ V(z(t)) \right\} = \langle \nabla V(z(t)), \frac{dz}{dt} \rangle = \langle \nabla V(z(t)), h(z(t)) \rangle = \mathcal{L}V(z(t)). \tag{4.3.2}
\]
If \( L^* V \) can be bounded above in terms of \( V \) then it is possible to use differential inequalities, such as the Gronwall Lemma 4.3, to obtain bounds on \( V(z(t)) \), and hence on \( z(t) \). Then \( V(z(t)) \) is a Lyapunov function.

Also important is the formal \( L^2 \)-adjoint operator \( L^* \), given by

\[
L^* V = -\nabla \cdot (hV). \tag{4.3.3}
\]

We now show the crucial role played by \( L \) and \( L^* \) in understanding how families of solutions of (4.1.1), parameterized by the initial data, and possibly carrying a probability measure, behave.

Let \( v \) satisfy the linear PDE

\[
\frac{\partial v}{\partial t} = L v, \quad (z, t) \in Z \times (0, \infty),
\]

\[
v(z, 0) = \phi(z), \quad z \in Z. \tag{4.3.4}
\]

We will denote the solution of (4.3.4) by \( v(z, t) = (e^{Lt})\phi(z) \). This is often referred to as the semigroup notation for the solution of a time-dependent linear operator equation.

**Theorem 4.4.** Assume that the solution of (4.1.1) generates a one-parameter group on \( Z \) so that \( \phi^{-t}(z) = Z \) for all \( t \in \mathbb{R} \). Assume also that \( \phi \) is sufficiently smooth so that (4.3.4) has a classical solution. Then the classical solution satisfies

\[
v(z, t) = \phi(\phi^{-t}(z)), \quad \forall t \in \mathbb{R}^+, z \in Z. \tag{4.3.5}
\]

**Proof.** Note that (4.3.5) satisfies the initial condition \( v(z, 0) = \phi(\phi^0(z)) = \phi(z) \).

Using the group property of \( \phi^t \) we deduce that \( v(z, t) \) given by (4.3.5) satisfies \( v(\phi^{-t}(z), t) = \phi(z), \forall t \in \mathbb{R}^+, z \in Z \). By differentiating with respect to \( t \) we obtain from (4.3.5) that

\[
\frac{d}{dt} \{ v(\phi^{-t}(z), t) \} = 0
\]

and so

\[
\frac{\partial v(\phi^{-t}(z), t)}{\partial t} + \langle \nabla v(\phi^{-t}(z), t), \frac{d}{dt} \{ \phi^{-t}(z) \} \rangle = 0.
\]

But \( \phi^{-t}(z) \) is the backwards time solution for (4.1.1) and hence satisfies

\[
\frac{d}{dt} (\phi^{-t}(z)) = -h(\phi^{-t}(z)).
\]
Thus
\[ \frac{\partial v(\varphi^{-t}(z), t)}{\partial t} + \langle \nabla v(\varphi^{-t}(z), t), -h(\varphi^{-t}(z)) \rangle = 0, \quad \forall t \in \mathbb{R}^+, z \in \mathbb{Z}. \]

This is equivalent to
\[ \frac{\partial v(z, t)}{\partial t} + \langle \nabla v(z, t), -h(z) \rangle = 0, \quad \forall t \in \mathbb{R}^+, z \in \mathbb{Z} \]
showing that (4.3.5) solves the linear PDE (4.3.4).

Formula (4.3.5) represents solution of the PDE (4.3.4) by the method of characteristics. Conversely it shows that the family of all solutions of (4.1.1) for \( z_0 \in \mathbb{Z} \) can be found by solving a linear PDE. We now study what happens when we place a probability measure on \( z_0 \), so that \( z(t) \) solving (4.1.1) is a random variable.

To this end consider the adjoint of (4.3.4), namely the Liouville equation
\[ \frac{\partial \rho}{\partial t} = \mathcal{L}^* \rho, \quad (z, t) \in \mathbb{Z} \times (0, \infty), \]
\[ \rho(z, 0) = \rho_0(z), \quad z \in \mathbb{Z}. \]

Using the semigroup notation the solution can be denoted by \( \rho(z, t) = (e^{t\mathcal{L}^*}) \rho_0(z) \).

Now let \( \mathbb{E} \) denote expectation with respect to initial data distributed according to a random variable on \( \mathbb{Z} \) with density \( \rho_0(z) \), i.e.
\[ \mathbb{E} f := \int_{\mathbb{Z}} f(z) \rho_0(z) \, dz. \]

**Theorem 4.5.** Assume that the solution of (4.1.1) generates a one-parameter group on \( \mathbb{Z} \) so that \( \varphi^{-t}(\mathbb{Z}) = \mathbb{Z} \). Assume also that \( \phi \) is sufficiently smooth so that (4.3.4) has a classical solution. Finally, assume that the initial data for (4.1.1) is distributed according to a random variable on \( \mathbb{Z} \) with density \( \rho_0(z) \), smooth enough so that (4.3.7) has a classical solution. Then, \( z(t) \) is a random variable on \( \mathbb{Z} \) with density \( \rho(z, t) \) satisfying (4.3.7).
Proof. Note that, by the previous result,

\[
\mathbb{E}(\phi(z(t))) = \int_Z \phi(\varphi^t(z)) \rho_0(z) dz
\]
\[
= \int_Z v(z, t) \rho_0(z) dz
\]
\[
= \int_Z (e^{Lt} \phi)(z) \rho_0(z) dz
\]
\[
= \int_Z (e^{L^* t} \rho_0)(z) \phi(z) dz.
\]

Also, if \( \rho(z, t) \) is the density of \( z(t) \), then

\[
\mathbb{E}(\phi(z(t))) = \int_Z \rho(z, t) \phi(z) dz.
\]

Equating these two expressions for the expectation at time \( t \) and using the arbitrariness of \( \phi \), together with an approximation argument to extend the equality to all \( \phi \) in \( L^2 \), shows that

\[
\rho(z, t) = (e^{L^* t} \rho_0)(z)
\]

in \( L^2 \). Hence, by the assumed smoothness, the density \( \rho(z, t) \) satisfies the adjoint equation (4.3.7).

4.4 Ergodicity

In this section we will assume for simplicity that the phase space \( Z \) is compact with no boundary. The natural example is \( Z = \mathbb{T}^d \), in which case \( L \), and its formal adjoint, are equipped with periodic boundary conditions. We will consider the measure space \( (Z, A, \mu) \), where \( A \) denotes a \( \sigma \)-algebra of measurable subsets of \( Z \) and \( \mu \) denotes a measure on \( Z \). Let \( \varphi^t \) denote the solution operator for (4.1.1). We will say that a set \( A \in A \) is invariant under \( \varphi^t \) provided that for all \( t \in \mathbb{R} \)

\[
\varphi^t(A) = A.
\]

Definition 4.6. The ODE (4.1.1) is called ergodic if every invariant set \( A \) of \( \varphi^t \) is such that either \( \mu(A) = 0 \) or \( \mu(A) = 1 \).

Note that the definition of ergodicity is relative to the measure space in question.
Theorem 4.7. The following properties are equivalent.

(i) The ODE (4.1.1) is ergodic.

(ii) 1 is a simple eigenvalue of the operator $e^{Lt}$.

(iii) The only solutions of $\mathcal{L}f = 0$

are constants.

(iv) The stationary Liouville equation

$$\mathcal{L}^* \rho = 0$$

has a unique solution $\rho^\infty$ satisfying $\int_Z \rho^\infty(z)dz = 1$.

(v) For every bounded measurable $\phi : Z \mapsto \mathbb{R}$ we have

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T \phi(z(t)) dt = \int_Z \phi(z) \rho^\infty(z) dz, \quad \mu \text{ a.s.}$$

By choosing $\phi$ to be $I_A$, the indicator function of Borel set $A \subseteq Z$, we deduce from the last result that the measure $\mu$ defined by

$$\mu(A) = \lim_{T \to \infty} \frac{1}{T} \int_0^T I_A(z(t)) dt,$$

(4.4.1)

has density $\rho^\infty$. Thus $\mu(dz) = \rho^\infty(z)dz$.

Notice that an important aspect of ergodicity, encapsulated in the fifth item of the theorem, is that the system forgets its initial condition. It is this property that will be crucial in much of our exploitation of ergodicity.

4.5 Discussion and Bibliography

The complete proof of Theorem 4.2 can be found in [8, Sec. 31] or [31]. The generator of a system of ODEs and its properties are discussed in [88, Sec. 7.6]; in particular the results of the ergodicity Theorem 4.7 can be found in this book. The use of the generator to study Lyapunov functions, and obtain a priori estimates on solutions of ODEs, may be found in [141]. Sometimes the generator $\mathcal{L}$ which was defined in equation (4.3.1) and its adjoint $\mathcal{L}^*$ are called the generators of the Koopman and the Frobenious–Perron operators, respectively. The Liouville equation is the fundamental equation of non–equilibrium statistical mechanics; see e.g. [13]. The method of characteristics is discussed in numerous PDE books including [45].
4.6 Exercises

1. Let $\mathcal{Z} = \mathbb{T}^d$. Show that for all $f \in C^1(\mathcal{Z}, \mathbb{R})$ the formal $L^2$–adjoint of $\mathcal{L}$ defined in eqn. (4.3.1) is $\mathcal{L}^*$ defined in eqn. (4.3.3).

2. Let $H(p, q) : \mathbb{R}^{2d} \to \mathbb{R}$ be a smooth function and consider the (Hamiltonian) ODE
   \[ \dot{q} = \nabla_p H(p, q), \quad \dot{p} = -\nabla_q H(p, q) \]  
   (4.6.1)

   a. Write down the generators of the Koopman and Frobenious–Perron operators for (4.6.1).
   b. Write down the Liouville equation (4.6.1)
   c. Show that every smooth function of the Hamiltonian $H(p(t), q(t))$ solves the Liouville equation.
   d. Is the Hamiltonian system (4.6.1) ergodic?

3. Carry out the same program as in the previous exercise for the (gradient) ODE
   \[ \dot{q} = -\nabla q V(q), \]  
   (4.6.2)

   where $V(q) : \mathbb{R}^d \to \mathbb{R}$ is a smooth function.
Chapter 5

Markov Chains

5.1 Set-Up

In this section we set-up Markov chains on a countable state space which, without
loss of generality, we take as $\mathcal{I}$, a subset of the positive integers. $^1$ We start with
discrete time Markov chains and then construct continuous time Markov chains
from them.

A matrix $P$ with entries $p_{ij}$ is a stochastic matrix if

$$
\sum_j p_{ij} = 1 \quad \forall i \in \mathcal{I}
$$

and $p_{ij} \in [0, \infty)$ for all $(i, j) \in \mathcal{I} \times \mathcal{I}$.

Definition 5.1. The random sequence $\{z_n\}_{n\geq 0}$ is a homogeneous Markov chain
with initial distribution $\rho_0$ and transition matrix $P$ if

- $z_0$ has distribution $\rho_0$;
- for $n \geq 0$ $P(z_{n+1} = j | z_n = i) = p_{ij}$, independently of $z_0, \ldots, z_{n-1}$.

Note that $(P^k)_{ij}$ gives the probability that $z_k = j$ given $z_0 = i$.

Notice that a stochastic matrix satisfies

$$
P1 = 1, \quad (5.1.1)
$$

$^1$Unless stated otherwise, all sums are over $\mathcal{I}$ in this chapter. In later chapters we will sometimes
find it convenient to work with a doubly indexed state space found as the product of two subsets of
the integers.
where $1$ is the vectors of ones. Combining this with the fact that $P$ has positive entries implies the following fundamental estimate:

$$\|P\|_\infty = 1. \quad (5.1.2)$$

**Definition 5.2.** A continuous time Markov chain is a Markov stochastic process with state space $E = \mathcal{I}$.

We will only consider time homogeneous Markov chains in what follows and hence refrain from explicitly stating this fact in the remainder of these notes.

A discrete-time Markov chain can be used to construct a continuous time Markov chain as follows. Let the i.i.d sequence $\{\tau_n\}_{n \geq 0}$ be distributed as $\exp(\lambda)$ and define $\{t_n\}_{n \geq 0}$ by $t_{n+1} = t_n + \tau_n$. Let $\{z_n\}_{n \geq 0}$ be a discrete time Markov chain on $\mathcal{I}$, independent of the $\{\tau_n\}_{n \geq 0}$, and set

$$z(t) = z_n, \ t \in [t_n, t_{n+1}). \quad (5.1.3)$$

We call this a jump chain. Notice that $z(t)$ takes values in $\mathcal{I}$ and is a càdlàg process. The fact that $z(t)$ is Markov follows from the Markovian structure of $\{z_n\}_{n \geq 0}$, together with the properties of the exponential random variable.

Informally we may write $z(t)$ as the solution of the differential equation

$$\frac{dz}{dt} = \delta(t - t_j) \left( k(z(t^-)) - z(t^-) \right) \quad (5.1.4)$$

where $k(z)$ is distributed as $p(z, \cdot)$ and the $\{k(z(t_j^-)))\}_{j \geq 0}$ are drawn independently of one another, and independently of the $\{\tau_j\}$. This representation follows because, integrating over the jump times $t_j$ yields

$$z(t_j^+) - z(t_j^-) = \lim_{\varepsilon \to 0} \int_{t_j-\varepsilon}^{t_j+\varepsilon} \delta(t - t_j) \left( k(z(t^-)) - z(t^-) \right) dt$$

$$= k(z(t_j^-)) - z(t_j^-)$$

and so $z(t_j^+) = k(z(t_j^-))$ as desired. Making sense of this random differential equation, and in particular showing that it has a solution for all time $t > 0$, is intimately related to the question of showing that the jump times $t_j$ do not accumulate at a finite time. In the next section we assume that this is indeed the case. In the section following it, we return to the discussion of existence of solutions to the Markov chain.
5.2 The Generator

We start by finding a representation of the matrix $P(t)$ with entries

$$p_{ij}(t) = \mathbb{P}(z(t) = j|z(0) = i). \quad (5.2.1)$$

We will express $P(t)$ in terms of $P$ and $\lambda$ the parameters input into the jump chain. Note that, by properties of exponential random variables encapsulated in (3.2.4),

$$\mathbb{P}(k \text{~jumps in~}[0, t]) = e^{-\lambda t}((\lambda t)^k)/k!.$$ 

Thus

$$p_{ij}(t) = \sum_{k=0}^{\infty} e^{-\lambda t}((\lambda t)^k)/k! (P^k)_{ij},$$

since $(P^k)_{ij} = \mathbb{P}(z_k = j|z_0 = i)$. Hence

$$P(t) = e^{\lambda t}(P - I) = e^{Lt}$$

with $L = \lambda(P - I)$. The matrix $L$ is called the generator of the continuous time Markov chain and can be related to the abstract definition of generator for Markov processes as given in Chapter 3. Notice that, since $P$ is a stochastic matrix, the matrix $L$ satisfies all of the following criteria which we take as a definition.

**Definition 5.3.** A matrix $L : \mathcal{I} \to \mathcal{I}$ with entries $L_{ij}$ is the generator of a continuous time Markov chain if

- $\sum_j l_{ij} = 0 \quad \forall i \in \mathcal{I};$
- $l_{ij} \in [0, \infty) \quad \forall (i, j) \in \mathcal{I} \times \mathcal{I} \text{~with~} i \neq j.$

The definition implies that

$$-l_{ii} \in [0, \infty) \quad \forall i \in \mathcal{I}.$$ 

Notice that, given a generator $L$, it is possible to find a discrete time Markov chain, and a sequence of i.i.d exponential random variables so that the associated jump chain generates paths of the continuous time Markov chain with generator $L$. Specifically we fix any $\lambda > 0$, generate the $\tau_j$ as an i.i.d sequence with $\tau_0 \sim \exp(\lambda)$ and generate $z_n$ from the Markov chain $P = I + \lambda^{-1}L$. In the construction above
we insisted that the jump times were i.i.d from $\exp(\lambda)$. However it is also possible to allow each $\tau_j$ to be independent of the others, but distributed as $\exp(\lambda(z(t^-_j)))$. We discuss this in the next section.

The fundamental object that characterizes a continuous time Markov chain is its generator. We now give another way to see the relationship between the continuous time Markov chain with transition matrix $P(t)$, and the generator $L$. Consider a continuous time Markov chain $z(t)$, $t \geq 0$, taking values in the state space $\mathcal{I} \subseteq \{1, 2, \ldots \}$. Let $p_{ij}(t)$ be the transition probability from state $i$ to $j$ given by (5.2.1). The Markov property implies that for all $t$, $\Delta t \geq 0$,

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

This is the Chapman-Kolmogorov equation (3.3.3) in the discrete state space setting, so that integrals become sums. From this equation it follows that

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

where

$$\ell_{kj}(\Delta t) = \frac{1}{\Delta t} \times \left( \begin{array}{c@{\quad}c} p_{kj}(\Delta t) & k \neq j \\ p_{jj}(\Delta t) - 1 & k = j \end{array} \right). \quad (5.2.2)$$

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

$$\frac{dp_{ij}}{dt} = \sum_k p_{ik}\ell_{kj}. \quad (5.2.3)$$

Because $\sum_j p_{ij}(\Delta t) = 1$ it follows that $\sum_j \ell_{ij}(\Delta t) = 0$, and assuming that the limit exists,

$$\sum_j \ell_{ij} = 0. \quad (5.2.4)$$

This implies that

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

for the limit equation (5.2.3).

Introducing the matrices $P(t)$, $L$ with entries $p_{ij}$, $\ell_{ij}$, respectively, $i, j \in \mathcal{I}$, equation (5.2.3) reads, in matrix notation,

$$\frac{dP}{dt} = PL, \quad P(0) = I. \quad (5.2.5)$$
5.2. **THE GENERATOR**

As shown in (5.2.2) above, it is calculated from $P$ via the formula

$$L = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \left( P(\Delta t) - I \right).$$

The generator has constants in its null space by (5.2.4):

$$L 1 = 0.$$  \hfill (5.2.6)

Furthermore, the non-negativity of the $p_{ij}$ implies that $L$ has non-negative off-diagonals. The condition (5.2.4) thus implies that diagonal entries of $-L$ are also non-negative.

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

$$\frac{dP}{dt} = LP, \quad P(0) = I.$$  \hfill (5.2.7)

We refer to both (5.2.5) and (5.2.7) as the **master equation** of the Markov chain.

Let $\rho(t) = (\rho_0(t), \rho_1(t), \ldots)^T$ be the transpose of the $i^{th}$ row of $P(t)$, i.e., a column vector whose entries $\rho_j(t) = p_{ij}(t)$ are the probabilities that a system starting in state $i$ will end up, at time $t$, in each of the states $j \in I$. Let $e_i$ denote the $i^{th}$ unit vector, zero in all entries except the $i^{th}$, in which it is one. Directly from (5.2.5) we obtain the following theorem.

**Theorem 5.4.** The probability vector $\rho$ satisfies

$$\frac{d\rho}{dt} = L^T \rho, \quad \rho(0) = e_i.$$  \hfill (5.2.8)

If the initial state of the Markov chain is random, with probability vector $\rho(0) = \rho_0$ chosen independently of the transition probabilities in the Markov chain, then

$$\frac{d\rho}{dt} = L^T \rho, \quad \rho(0) = \rho_0.$$  \hfill (5.2.9)

**Proof.** The first result follows from (5.2.5). Let $\rho^{(i)}$ denote the solution of (5.2.8). If the initial condition is random with $\rho(0) = \rho_0$ then

$$\rho(t) = \sum_i \rho_0(i) \rho^{(i)}(t).$$

Differentiating and using (5.2.8) gives (5.2.9). \hfill \Box

Equation (5.2.9) is the Markov chain analogue of the Liouville and Fokker-Planck equations described in the previous and following chapter respectively. We refer to it as the **forward equation**.
Let $\phi : I \mapsto \mathbb{R}$ be a real valued function defined on the state space; it can be represented as a vector with entries $\phi_j, j \in I$. Then let $v(t) = (v_0(t), v_1(t), \ldots)^T$ denote the vector with $i^{th}$ entry

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

where $\mathbb{E}$ denotes expectation with respect to the Markov transition probabilities.

**Theorem 5.5.** The vector of expectations $v$ satisfies the equation

$$\frac{dv}{dt} = Lv, \quad v(0) = \phi. \quad (5.2.10)$$

**Proof.** The function $v_i(t)$ can be written explicitly in terms of the transition probabilities:

$$v_i(t) = \sum_j p_{ij}(t)\phi_j. \quad (5.2.11)$$

If we set $\phi = (\phi_0, \phi_1, \ldots)^T$ then this can be written in vector form as $v(t) = P(t)\phi$. Differentiating with respect to time and using the master equation (5.2.7) gives the desired result. \hfill \Box

Equation (5.2.10) is the Markov chain analogue of the method of characteristics and of the backward Kolmogorov equation described in the preceding and following chapters respectively. We refer to it as the **backward equation**.

### 5.3 Existence and Uniqueness

In the construction above we insisted that the jump times were i.i.d exponential random variables with fixed rate $\lambda$. However it is also possible to allow each $\tau_j$ to be independent of the others, but with rate depending on the state $z(t)$. Given a generator $L$ there is a canonical way of constructing a jump chain, with variably distributed jump times, to realize the Markov chain. Let $l(i) = -l_{ii}$ (no summation convention). For $j \neq i$ we set

$$p_{ij} = \begin{cases} l_{ij}/l(i), & l(i) \neq 0, \\ 0, & l(i) = 0 \end{cases}.$$

and for $j = i$ we set

$$p_{ii} = \begin{cases} 0 & l(i) \neq 0, \\ 1 & l(i) = 0. \end{cases}$$
Notice that, if \( l(i) \neq 0 \),
\[
\sum_j p_{ij} = \frac{1}{l(i)} \sum_{j \neq i} t_{ij} = -\frac{l_{ii}}{l(i)} = 1,
\]
and that, if \( l(i) = 0 \),
\[
\sum_j p_{ij} = 1.
\]
Hence \( p \) is a stochastic matrix.

We generate a sequence \( \{z_n\}_{n \geq 0} \) from the discrete time Markov chain with \( P \) as transition matrix. The jump chain associated with this choice of transition matrix \( P \) is then given by
\[
z(t) = z_n, \quad t \in [t_n, t_{n+1}).
\]  
(5.3.1)
where \( t_{n+1} = t_n + \tau_n \) and the \( \tau_n \) are independent random variables distributed as \( \exp(l(z_n)) \). Again \( z(t) \) takes values in \( \mathcal{I} \), is càdlàg, and may be described by the differential equation (5.1.4). We call this the \textit{canonical jump chain} associated to a Markov chain with generator \( L \).

The discrete time Markov chain with transition matrix \( P \) will generate sequences \( \{z_n\}_{n \geq 0} \) with probability one. However, the associated jump chain just described may not generate a sequence \( \{z(t)\}_{t \geq 0} \). Specifically, the sequence of jump times \( t_n \) may accumulate at a finite time. In this situation we say that the Markov chain \textit{explodes} – we do not have existence and uniqueness on \( t \in [0, \infty) \).

**Definition 5.6.** A \textit{continuous time Markov chain} is non-explosive if, with probability one, the jump times do not accumulate at a finite time.

It is important to understand conditions which ensure non-explosion. There are two useful conditions which ensure this:

**Theorem 5.7.** The Markov chain is non-explosive if either:

- \( \mathcal{I} \) is finite; or
- \( \sup_{i \in \mathcal{I}} l(i) < \infty \).

**Proof.** Let
\[
\zeta = \sum_{n=0}^{\infty} \tau_n.
\]
Set $T_n = l(z_n)\tau_n$ and notice that the $\{T_n\}$ form an i.i.d sequence with $T_1 \sim \exp(1)$. By the strong law of large numbers,
\[
\frac{1}{N+1} \sum_{n=0}^{N} T_n \to 1 \quad \text{a.s.}
\]
In both cases (i) and (ii) we have $l = \sup_{i \in I} l(i) < \infty$. Hence
\[
l\zeta \geq \sum_{n=0}^{\infty} T_n = \infty \quad \text{a.s.}
\]
so that $l\xi = \infty$. The result follows.

### 5.4 Ergodicity

For simplicity we assume that $I$ is a finite set. We start by discussing discrete time Markov chains. By (5.1.1), the matrix $P - I$ has a non-empty null-space, and hence its transpose does too. Hence there exists vector $\rho^\infty$ such that
\[
P^T \rho^\infty = \rho^\infty.
\]

**Theorem 5.8.** All eigenvalues of $P$ lie in the closed unit circle. The vector $\rho^\infty$ may be chosen so that all of its entries are non-negative.

The vector $\rho^\infty$ is known as the **invariant distribution**.

**Definition 5.9.** The discrete time Markov chain is said to be ergodic if the spectrum of $P$ lies strictly inside the unit circle, with the exception of a simple eigenvalue at one.

Using the construction of $L$ from $P$ in (5.2.2) we deduce that
\[
L1 = 0, \\
L^T \rho^\infty = 0.
\]

**Theorem 5.10.** All eigenvalues of $L$ lie in the left half-plane. The vector $\rho^\infty$ may be chosen so that all of its entries are non-negative.

The vector $\rho^\infty$ is known as the **invariant distribution**.
Definition 5.11. The continuous time Markov chain is said to be ergodic if the spectrum of $L$ lies strictly in the left-half plane, with the exception of a simple eigenvalue at zero.

Theorem 5.12. An ergodic continuous time Markov chain on finite state space $\mathcal{I}$ satisfies the following five properties:

i) $\text{Null}(L) = \text{span}\{1\}$;

ii) $\text{Null}(L^T) = \text{span}\{\rho^\infty\}$, $\rho^\infty(i) > 0 \forall i \in \mathcal{I}$;

iii) $\exists C, \lambda > 0$ such that solution of the forward equation satisfies

$$\|\rho(t) - \rho^\infty\|_1 \leq Ce^{-\lambda t} \ \forall t > 0;$$

iv) $\exists C, \lambda > 0$ such that solution of the backward equation satisfies

$$\|v(t) - \langle \rho^\infty, \phi \rangle_1 \|_\infty \leq Ce^{-\lambda t} \ \forall t > 0.$$

v) $\lim_{T \to \infty} \frac{1}{T} \int_0^T \phi(z(t))dt = \langle \rho^\infty, \phi \rangle$ a.s.

Since the state space is finite dimensional the convergence results hold in any norm; however the choices as stated are natural from a probabilistic viewpoint.

Notice that, by choosing $\phi = e_i$, we deduce from the final result that the $i^{th}$ component of $\rho^\infty$ can be found as the proportion of time that an arbitrary trajectory of the Markov chain on $t \in [0, \infty)$ spends in state $i$.

### 5.5 Discussion and Bibliography

A good background reference on Markov chains is Norris [110]. The text [41] has a wealth of material on the general setting for Markov processes, including Markov chains. The paper [57] describes a simple algorithm for simulating continuous time Markov chains.
5.6 Exercises

1. Consider the discrete time Markov chain with generator

\[ P = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix}. \]

Find the invariant density \( \rho^\infty \). Under what conditions is the Markov chain ergodic?

2. Consider the continuous time Markov chain with generator

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

Find the invariant density \( \rho^\infty \). Under what conditions is the Markov chain ergodic?

3. Construct the canonical jump chain associated to the generator given in the previous question.

4. Implement a numerical algorithm to simulate paths of the continuous time Markov chain from the previous example, by using the canonical jump chain.

5. Show that the jump chain (5.1.3) satisfies the Markov property.

6. Show that the canonical jump chain (5.3.1) satisfies the Markov property.
Chapter 6

Stochastic Differential Equations

6.1 Set-Up

In this chapter we give background material concerning ODEs driven by Gaussian white noise – SDEs. Let $W(t)$ denote $m$ dimensional Brownian motion, $h: \mathcal{Z} \to \mathbb{R}^d$ a smooth vector-valued function and $\gamma: \mathcal{Z} \to \mathbb{R}^{d \times m}$ a smooth matrix valued function. In the following we typically take $\mathcal{Z} = \mathbb{T}^d, \mathbb{R}^d$ or $\mathbb{R}^l \oplus \mathbb{T}^{d-l}$. Consider the Itô SDE

$$\frac{dz}{dt} = h(z) + \gamma(z) \frac{dW}{dt}, \quad z(0) = z_0. \quad (6.1.1)$$

We think of the term $\frac{dW}{dt}$ as representing Gaussian white noise: a delta-correlated Gaussian process. But such a process only exists as a distribution, and so the precise interpretation of (6.1.1) is as an integral equation for $z \in C(\mathbb{R}^+, \mathcal{Z})$:

$$z(t) = z(0) + \int_0^t h(z(s))ds + \int_0^t \gamma(z(s))dW(s). \quad (6.1.2)$$

In order to make sense of this equation we need to make sense of the stochastic integral against $W(s)$. We use the Itô interpretation of the stochastic integral as defined in Chapter 3. section. Because it is notationally convenient to do so, we will frequently write SDEs in the unintegrated form (6.1.1). Whenever we do this it should be interpreted as shorthand for the precise interpretation (6.1.2).

In later chapters we will often consider $z = (x^T, y^T)^T$, with $x \in \mathcal{X}, y \in \mathcal{Y}$. If $\mathcal{Z} = \mathbb{T}^d$ (resp. $\mathbb{R}^d$) then $\mathcal{X} = \mathbb{T}^d$ (resp. $\mathbb{R}^l$) and $\mathcal{Y} = \mathbb{T}^{d-l}$ (resp. $\mathbb{R}^{d-l}$). If $\mathcal{Z} = \mathbb{R}^l \oplus \mathbb{T}^{d-l}$ then $\mathcal{X} = \mathbb{R}^l$ and $\mathcal{Y} = \mathbb{T}^{d-l}$.
6.2 Existence and Uniqueness

In Theorem 4.2 we proved existence and uniqueness of solutions for ODE (i.e. when \( \gamma \equiv 0 \) in (6.1.1)) for globally Lipschitz vector fields \( h(x) \). A very similar theorem holds when \( \gamma \neq 0 \). As for ODE the conditions can be weakened, when \textit{a priori} bounds on the solution can be found; but we limit ourselves to the simple set-up of the following theorem, for expository purposes.

**Theorem 6.1.** Assume that both \( h(\cdot) \) and \( \gamma(\cdot) \) are globally Lipschitz on \( \mathcal{Z} \) and that \( z_0 \) is a random variable independent of the Brownian motion \( W(t) \) with

\[
\mathbb{E}|z_0|^2 < \infty.
\]

Then the SDE (6.1.1) has a unique solution \( z(t) \in C(\mathbb{R}^+; \mathcal{Z}) \) with

\[
\mathbb{E} \left[ \int_0^T |z(t)|^2 \, dt \right] < \infty \quad \forall T < \infty.
\]

We conclude the section with two remarks, both of which will play an important role in future chapters.

**Remark 6.2.** The Stratonovich analogue of (6.1.1) is

\[
\frac{dz}{dt} = h(z) + \gamma(z) \circ dW \quad , \quad z(0) = z_0.
\]

By using definitions (3.4.5) and (3.4.6) it can be shown that \( z \) satisfying the Stratonovich SDE (6.2.1) also satisfies the Itô SDE

\[
\frac{dz}{dt} = h(z) + \frac{1}{2} \nabla \cdot (\gamma(z)\gamma(z)^T) - \frac{1}{2} \gamma(z)\nabla \cdot (\gamma(z)^T) + \gamma(z) \frac{dW}{dt} \quad , \quad z(0) = z_0.
\]

See Exercise 1.

**Remark 6.3.** The Definition 3.3 of Brownian motion implies the interesting scaling property

\[
W(ct) = \sqrt{c}W(t)
\]

where the above should be interpreted as holding in law. From this it follows that, if \( s = ct \), then

\[
\frac{dW}{ds} = \frac{1}{\sqrt{c}} \frac{dW}{dt} ,
\]

again in law.
6.3. THE GENERATOR

Hence, if we scale time to \( s = ct \) in (6.1.1), then we get the equation

\[
\frac{dz}{ds} = h(z) + \sqrt{c\gamma(z)} \frac{dW}{ds}, \quad z(0) = z_0.
\]

(The precise interpretation is as an integral equation, as always). Notice that, whilst the SDE transforms unusually under \( s = ct \) the Fokker-Planck equation transforms in the standard way, because it sees the a quadratic term formed from the diffusion coefficient.

6.3 The Generator

With the functions \( h(z), \gamma(z) \) given in the SDE (6.1.1) we define

\[
\Gamma(z) = \gamma(z)\gamma(z)^T.
\]

The generator \( \mathcal{L} \) is then defined as

\[
\mathcal{L}v = h \cdot \nabla v + \frac{1}{2} \Gamma : \nabla \nabla v. \tag{6.3.1}
\]

We will also be interested in the formal \( L^2 \)-adjoint operator \( \mathcal{L}^* \)

\[
\mathcal{L}^*v = -\nabla \cdot (hv) + \frac{1}{2} \nabla \cdot (\Gamma v).
\]

The \textit{Itô formula} which follows is the basic result concerning the rate of change in time of functions \( V : Z \to \mathbb{R}^n \) evaluated the solution of an \( Z \)-valued SDE. Heuristically it delivers the following result:

\[
\frac{d}{dt} \{V(z(t))\} = \mathcal{L}V(z(t)) + \nabla V(z(t))\gamma(z(t)) \frac{dW}{dt}, \quad Y(0) = V(z_0).
\]

Note that if \( W \) were a smooth time-dependent function this formula would not be correct: there is an additional term in \( \mathcal{L}V \), proportional to \( \Gamma \), which arises from the lack of smoothness of Brownian motion. As for the SDE (6.1.1) itself, the precise interpretation of this expression is in integrated form.

**Lemma 6.4. Itô Formula** Assume that the conditions of Theorem 6.1 hold. Let \( x(t) \) solve (6.1.1) and let \( V \in C^2(Z, \mathbb{R}^n) \). Then the process \( Y(t) = V(z(t)) \) satisfies

\[
V(z(t)) = V(z(0)) + \int_0^t \mathcal{L}V(z(s))ds + \int_0^t \nabla V(z(s))\gamma(z(s))dW(s).
\]
Let \( \phi : \mathcal{Z} \rightarrow \mathbb{R} \) and consider the function

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

(6.3.2)

where the expectation is with respect to all Brownian driving paths. By averaging in the Itô formula, which removes the stochastic integral, and using the Markov property, we deduce the following important consequence of Lemma 6.4.

**Theorem 6.5.** Assume that \( \phi \) is chosen sufficiently smooth so that the backward Kolmogorov equation

\[
\frac{\partial v}{\partial t} = \mathcal{L}v, \quad (z, t) \in \mathcal{Z} \times (0, \infty),
\]

\[
v = \phi, \quad (z, t) \in \mathcal{Z} \times \{0\}
\]

(6.3.3)

has a unique classical solution \( v(x, t) \in C^{2,1}(\mathcal{Z} \times (0, \infty), \mathbb{R}^+) \). Then \( v \) is given by (6.3.2).

This is the analogue of (5.2.10) for Markov chains. If \( \gamma \equiv 0 \) in (6.1.1), so that the dynamics are deterministic, and \( \varphi^t \) is the flow on \( \mathcal{Z} \) so that \( z(t) = \varphi^t(z(0)) \), then the Kolmogorov equation (6.3.3) reduces to the hyperbolic equation (4.3.4) whose characteristics are the integral curves of the ODE (4.1.1).

A direct consequence of Result 6.5 is the Fokker-Planck equation for propagation of densities.

**Theorem 6.6.** Consider the (6.1.2) with \( z(0) \) a random variable with density \( \rho_0(z) \). Assume that the law of \( z(t) \) has a density \( \rho(x, t) \in C^{2,1}(\mathcal{Z} \times (0, \infty), \mathbb{R}^+) \). Then \( \rho \) satisfies the Fokker-Planck equation

\[
\frac{\partial \rho}{\partial t} = \mathcal{L}^* \rho, \quad (z, t) \in \mathcal{Z} \times (0, \infty),
\]

\[
\rho = \rho_0, \quad z \in \mathcal{Z} \times \{0\}.
\]

(6.3.4)

**Proof.** Let \( \mathbb{E}^\mu \) denote averaging with respect to the product measure induced by measure \( \mu \) with density \( \rho_0 \) on \( z(0) \) and independent driving Wiener measure on the SDE itself. By the previous result, averaging over random \( z(0) \) distributed with density \( \rho_0(z) \), we find

\[
\mathbb{E}^\mu(\phi(z(t))) = \int_{\mathcal{Z}} v(z, t)\rho_0(z)dz
\]

\[
= \int_{\mathcal{Z}} (e^{Lt}\phi)(z)\rho_0(z)dz
\]

\[
= \int_{\mathcal{Z}} (e^{L^*t}\rho_0)(z)\phi(z)dz.
\]
But since \( \rho(z, t) \) is the density of \( z(t) \) we also have

\[
\mathbb{E}^\rho(\phi(z(t))) = \int_{\mathbb{Z}} \rho(z, t) \phi(z) dz.
\]

Equating these two expressions for the expectation at time \( t \) shows that

\[
\rho(z, t) = (e^{L^* t} \rho_0)(z)
\]

in \( L^2 \). Hence the density \( \rho(z, t) \) satisfies the Fokker-Planck equation as required.

The Fokker-Planck equation is the continuous analogue of (5.2.11) for Markov chains, and of the Liouville equation (4.3.7) for ODE. Note that constants are in the null-space of \( L \) given by (6.3.1). Hence we expect \( L^* \) to have a non-trivial null-space. Let \( \rho \) be any steady solution of the Fokker-Planck equation, i.e.

\[
L^* \rho = 0. \tag{6.3.5}
\]

We have the following important result giving the Dirichlet form associated with the operator \( L \).

**Theorem 6.7.** Let \( \rho \) be any steady solution of the Fokker-Planck equation on \( \mathbb{T}^d \) with periodic boundary conditions. Let \( f \) be a smooth function on \( \mathbb{T}^d \). Then

\[
\int_{\mathbb{T}^d} (-L f(z)) f(z) \rho(z) dz = \frac{1}{2} \int_{\mathbb{T}^d} (\nabla f(z) \cdot \Gamma(z) \nabla f(z)) \rho(z) dz. \tag{6.3.6}
\]

**Proof.** We have

\[
L^*(f\rho) = -\nabla \cdot (h \rho f) + \frac{1}{2} \nabla \cdot \nabla \cdot \left( \Gamma \rho f \right)
\]

\[
= (L^* \rho) f + \left( -L f \right) \rho + \nabla \rho \cdot \Gamma \nabla f
\]

\[
+ \rho \Gamma : \nabla \nabla f + \nabla f \cdot (\nabla \cdot \Gamma) \rho.
\]

Now let \( g \) be another smooth, periodic function. We use the previous calculation to compute, using the fact that \( L \) and \( L^* \) are adjoint operators in \( L^2(\mathbb{T}^d) \) and by using

\[
\int_{\mathbb{T}^d} (\nabla g(z) \cdot \Gamma(z) \nabla g(z)) \rho(z) dz = \frac{1}{2} \int_{\mathbb{T}^d} \nabla \cdot (h \rho g) \rho(z) dz
\]

\[
= \frac{1}{2} \int_{\mathbb{T}^d} \nabla \cdot (h \rho g) \rho(z) dz
\]

\[
= \frac{1}{2} \int_{\mathbb{T}^d} \nabla \cdot (h \rho g) \rho(z) dz.
\]
the divergence theorem on the last term in the second line,

\[
\int_{\mathbb{T}^d} (\mathcal{L}g) f\rho \, dz = \int_{\mathbb{T}^d} g\mathcal{L}^*(f\rho) \, dz
\]

\[
= \int_{\mathbb{T}^d} g(-\mathcal{L}f)\rho + g\nabla\rho \cdot \Gamma \nabla f \, dz
\]

\[
+ \int_{\mathbb{T}^d} g\left(\nabla f \cdot (\nabla \cdot \Gamma)\rho + \rho\Gamma : \nabla\nabla f\right) \, dz
\]

\[
= \int_{\mathbb{T}^d} g(-\mathcal{L}f)\rho \, dz - \int_{\mathbb{T}^d} \nabla g \cdot \Gamma \nabla f \rho \, dz.
\]  (6.3.7)

Equation (6.3.7) implies

\[
\int_{\mathbb{T}^d} (-\mathcal{L}g) f\rho \, dz + \int_{\mathbb{T}^d} g(-\mathcal{L}f)\rho \, dz = \int_{\mathbb{T}^d} \nabla g \cdot \Gamma \nabla f \rho \, dz. \quad (6.3.8)
\]

Equation (6.3.6) follows from the above equation upon setting \( g = f \). □

Roughly speaking it shows that \(-\mathcal{L}\) is a positive operator, in an appropriate weighted \( L^2 \) space. Indeed let \( \rho \) be strictly positive on \( \mathbb{Z} \) and define a measure \( \mu(dx) = \rho(x)dx \). We can then introduce the weighted Hilbert space \( L^2(\mu) \) with inner product and norm as follows:

\[
(a, b)_\rho = \int_{\mathbb{T}^d} a(z) \cdot b(z) \rho(z) \, dz, \quad |a|_\rho^2 = (a, a)_\rho.
\]

Then the preceding theorem shows that

\[
(-\mathcal{L}f, f)_\rho = \frac{1}{2} |\nabla^T \nabla f|_\rho^2. \quad (6.3.9)
\]

Remark 6.8. In suitable functional settings, the previous theorem also applies on other choices of domain \( \mathbb{Z} \), not just on the torus. Specifically the function space should ensure that \( \mathcal{L}^* \) is the adjoint of \( \mathcal{L} \) and allow the divergence theorem calculation used to reach (6.3.7). Typically these conditions are realized on non-compact spaces by means of decay assumptions at infinity.

6.4 Ergodicity

For simplicity of exposition we state the rigorous results in the case where \( \mathbb{Z} = \mathbb{T}^d \). Consider the SDE (6.1.1) on \( \mathbb{T}^d \). Equip both the generator \( \mathcal{L} \) and its adjoint \( \mathcal{L}^* \) with periodic boundary conditions.
6.4. ERGODICITY

Definition 6.9. The SDE is said to be ergodic if the spectrum of the generator lies strictly in the left-half plane, with the exception of a simple eigenvalue at the origin.

In the following we use the short-hand notation $\rho(t)$ and $v(t)$ to denote the function-valued time-dependent solutions of the Fokker-Planck and Kolmogorov equations respectively. Thus we may view $\rho(t)$, $v(t)$ as being in a Banach space, for each fixed $t$, and measure their size through the $L^p$ norms. The notation 1 is used to denote functions which are constant and equal to one, a.e. in an $L^p$ sense.

Theorem 6.10. Equip $\mathcal{L}$, $\mathcal{L}^*$ on $\mathbb{T}^d$ with periodic boundary conditions and assume that $\Gamma(z)$ is strictly positive-definite, uniformly in $z \in \mathbb{T}^d$. Then the SDE (6.1.1) is ergodic and satisfies the following five properties:

- $\text{Null}(\mathcal{L}) = \text{span}\{1\}$;
- $\text{Null}(\mathcal{L}^*) = \text{span}\{\rho^\infty\}$, $\inf_{z \in \mathbb{T}^d} \rho^\infty(z) > 0$;
- $\exists C, \lambda > 0$ such that the solution of the Fokker-Planck equation with initial data a Dirac mass at arbitrary $z(0) \in \mathbb{T}^d$ satisfies
  \[ \|\rho(t) - \rho^\infty\|_1 \leq Ce^{-\lambda t} \quad \forall t > 0; \]
- $\exists C, \lambda > 0$ such that the solution of the backward Kolmogorov equation with initial data a continuous function $\phi$ satisfies
  \[ \left\| v(t) - \left( \int_{\mathbb{T}^d} \phi(z) \rho^\infty(z) \, dz \right) 1 \right\|_\infty \leq Ce^{-\lambda t} \quad \forall t > 0; \]
- For all $\phi \in C(\mathbb{T}^d)$
  \[ \lim_{T \to \infty} \frac{1}{T} \int_0^T \phi(z(t)) dt = \int_{\mathbb{T}^d} \phi(z) \rho^\infty(z) \, dz, \quad \text{a.s.} \]

Let $I_A$ denote the indicator function of Borel set $A \subseteq \mathbb{Z}$. This function is not continuous but may be approximated by a sequence of continuous functions. By choosing $\phi$ to be $I_A$ we deduce from the last result that the measure $\mu$ defined by
\[
\mu(A) = \lim_{T \to \infty} \frac{1}{T} \int_0^T I_A(z(t)) \, dt, \tag{6.4.1}
\]
has density $\rho^\infty$. Thus $\mu(dz) = \rho^\infty(z) dz$. 

Remark 6.11. Although we do not prove these results, a few remarks are in order. First the fact that the null-space of $L$ comprises constant, when the diffusion matrix $\Gamma$ is uniformly positive-definite and when periodic boundary conditions are used, may be seen by means of the maximum principle. Secondly the same fact also follows directly from (6.3.9) if it is assumed that $\rho^\infty$ is strictly positive.

Example 6.12. Consider one-dimensional Brownian motion on $T^1 = \mathbb{S}^1$:

\[
\frac{dz}{dt} = \sqrt{2\sigma}dW, \quad z(0) = z_0.
\]

The generator $L$ is the differential operator

\[
L = \sigma^2 \frac{d^2}{dz^2}
\]

equipped with periodic boundary conditions on $[0, 1]$. This operator is self-adjoint. The null space of both $L$ and $L^*$ comprises constant functions on $[0, 1]$. Both the backward Kolmogorov and the Fokker-Planck equation reduce to the heat equation

\[
\frac{\partial \rho}{\partial t} = \sigma^2 \frac{\partial^2 \rho}{\partial x^2}
\]

with periodic boundary conditions. Straightforward Fourier analysis shows that the solution converges to a constant at an exponential rate.

Although the theorem as stated holds with state space $T^d$, it readily extends to a variety of settings with the appropriate function space choice for $L$ and $L^*$. To illustrate this we include two examples on $\mathbb{R}^d$.

Example 6.13. Consider the Ornstein-Uhlenbeck (OU) process

\[
\frac{dz}{dt} = -z + \sqrt{2\sigma}dW, \quad z(0) = z_0, \quad(6.4.2)
\]

with $z(t) \in \mathbb{R}$. Here we consider $z_0$ is fixed and non-random. The solution is

\[
z(t) = e^{-t}z_0 + \sqrt{2\sigma} \int_0^t e^{-(t-s)} dW(s).
\]

Hence,

\[
\mathbb{E}z(t) = z_0 e^{-t}
\]
and, by the Itō isometry,

\[ E(z(t) - E z(t))^2 = 2\sigma E \left( \int_0^t e^{-2(t-s)} dW(s) \right)^2 = 2\sigma \int_0^t e^{-2(t-s)} ds = \sigma (1 - e^{-2t}). \]

In fact,

\[ z(t) \sim \mathcal{N}(e^{-t}z_0, \sigma(1 - e^{-2t})) \]

indicating convergence to the Gaussian invariant measure \( \mathcal{N}(0, \sigma) \) as \( t \to \infty \).

The Fokker-Planck equation here reduces to

\[
\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x} (-x \rho) + \frac{\partial}{\partial y} (-y \rho) = \sigma \frac{\partial^2 \rho}{\partial y^2} \quad (6.4.3)
\]

with initial condition being a Dirac mass centered at \( z_0 \). It is readily verified that the density associated with the Gaussian measure for \( z(t) \) is a solution of this linear PDE.

**Example 6.14.** Consider the equations

\[
\begin{align*}
\frac{dx}{dt} &= -ax + y, \quad x(0) = x_0, \\
\frac{dy}{dt} &= -y + \sqrt{2\sigma} dW, \quad y(0) = y_0,
\end{align*}
\]

for \((x(t), y(t)) \in \mathbb{R}^2\). We assume \( a > 0 \). The Fokker-Planck equation takes the form

\[
\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x} ((-ax + y) \rho) + \frac{\partial}{\partial y} (-y \rho) = \sigma \frac{\partial^2 \rho}{\partial y^2}.
\]

The previous example shows that \( y \) has Gaussian distribution and converges to a Gaussian invariant measure. Since

\[ x(t) = e^{-at}x(0) + \int_0^t e^{-a(t-s)} y(s) ds \]

and \( y \) is Gaussian we deduce that \( x \) too is Gaussian.

These considerations suggest that we seek a steady solution of the Fokker-Planck equation in the form

\[ \rho^\infty(x, y) \propto \exp\{-\alpha x^2 + \beta xy - \gamma y^2\}. \]
(The constant of proportionality should be chosen so that $\rho^\infty$ integrates to 1 on $\mathbb{R}^2$.) Substitution shows that

$$\alpha = \frac{a(a + 1)^2}{2\sigma}, \quad \beta = \frac{2a(a + 1)}{2\sigma}, \quad \gamma = \frac{(a + 1)}{2\sigma}.\]

Note that we have thus found the density of a Gaussian invariant measure for $(x, y)$.

**Example 6.15.** Consider the stochastic integral

$$I(t) = \int_0^t \eta(s) \, dW_1(s),$$

where $\eta(t)$ is the Ornstein–Uhlenbeck process defined in Example 6.13, namely

$$\frac{d\eta}{dt} = -\eta + \sqrt{2\sigma} \frac{dW_2}{dt}, \quad \eta(0) = \eta_0.$$

Here $W_1(t)$ and $W_2(t)$ are independent Brownian motions. The invariant measure for $\eta$ is an $\mathcal{N}(0, \sigma)$ Gaussian random variable. We assume that the initial condition is distributed according to this invariant measure. Hence, $\eta(t)$ is a stationary ergodic Markov process and Corollary 3.33 applies. Hence

$$\lim_{\varepsilon \to 0} \varepsilon I(t/\varepsilon^2) = \sqrt{\sigma}W(t),$$

where $W(t)$ is a standard Brownian motion in one dimension.

### 6.5 Discussion and Bibliography

For a discussion of SDEs from the viewpoint of the Fokker-Planck equation, see Risken [128] or Gardiner [53]. For a discussion of the generator $\mathcal{L}$, and the backward Kolmogorov equation, see Oksendal [111]. For a discussion concerning ellipticity, hypo-ellipticity and smoothness of solutions to these equations see Rogers and Williams [130]. The book by Khasminskii [66] has a good discussion of ergodicity. The book by Mao [100] has a good overview of stability theory and large time properties of SDEs. The Fokker-Planck equation is often referred to as the forward Kolmogorov equation in the mathematics literature.

We have only discussed strong solutions to SDE. The definition of a weak solution, together with existence and uniqueness theorems of weak solutions can be found in [130, Ch. 5]. The weak formulation of an SDE is equivalent to the martingale formulation. See [140]
6.6 Exercises

1. Derive the Itô SDE (6.2.2) from the Stratonovich SDE (6.2.1). Using the Itô form of the Stratonovich SDE, find the Fokker-Planck equation for the Stratonovich SDE.

2. Prove Result 6.5 (hint: use the Itô formula and the martingale property of the stochastic integral).

3. Consider the OU process defined in equation (6.4.2).
   (a) Calculate all moments of the process \( z(t) \).
   (b) Solve the corresponding Fokker–Planck equation (6.4.3) with \( \rho(0, x) = \delta(z) \) to show that the solution is
   \[
   \rho(x, t; z) = \sqrt{\frac{1}{\mu \sigma^2}} \frac{1}{(1 - e^{-2t})} \exp \left[ -\frac{(x - e^{-t}z)^2}{\sigma^2 (1 - e^{-2t})} \right].
   \]
   (c) Deduce the long–time behavior of the OU process from the above formula.

4. Consider the SDE
   \[
   m\ddot{x} = -\nabla V(x) - \gamma \dot{x} + \sqrt{\gamma D} \frac{dW}{dt},
   \]
   where \( m, D, \gamma \) are positive constants and \( V(x) : \mathbb{R}^d \rightarrow \mathbb{R} \) is a smooth function.
   (a) Write equation (6.6.1) as a first order system of SDEs in the form
   \[
   \frac{dz}{dt} = -\frac{1}{D} K \nabla H(z) + \frac{1}{\sqrt{m}} J \nabla H(z) + \sqrt{2K} \frac{dB}{dt},
   \]
   where \( z = (x^T, y^T)^T \), \( J \) (respectively \( K \)) is a skew (respectively symmetric) matrix which you should define and \( H(z) = \frac{1}{2} |y|^2 + V(x) \).
   (b) Write the corresponding generator and the Fokker–Planck equation.
   (c) Solve the stationary Fokker–Planck equation (hint: use separation of variables).
   (d) Solve equation (6.6.1) in one dimension for the cases \( V(x) \equiv 0 \) and \( V(x) = \frac{1}{2} x^2 \).
5. Consider a Markov chain $z$ with generator

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

Now let $x$ solve an SDE with coefficients depending on $x$:

$$\frac{dx}{dt} = f(x, z) + \alpha(x, z) \frac{dW}{dt}.$$

Write down the generator for the Markov process $(x, z)$. 
Chapter 7

Partial Differential Equations

7.1 Set-Up

In this chapter we describe the basic theory of elliptic, parabolic and hyperbolic PDE. In sections 7.3 and 7.4 We study the Dirichlet and the periodic boundary value problem for elliptic PDE in divergence form. Section 7.5 will be devoted to the Fredholm alternative, in a more general setting. In section 7.6 we describe the Cauchy problem for parabolic PDE and in section 7.7 the Cauchy problem for hyperbolic PDE. The function space settings are described in Chapter 2.

7.2 Elliptic PDE

The (homogeneous) Dirichlet problem is to find $u(x)$ solving

$$-\nabla \cdot (A(x) \nabla u) = f, \text{ for } x \in \Omega$$

$$u(x) = 0, \text{ for } x \in \partial \Omega,$$

where $A(x)$ is a positive definite matrix and $f \in H^{-1}(\Omega)$.

Let $\mathcal{Y} = T^d$ the $d-$ dimensional torus. The periodic problem is to find $u(y)$ solving

$$-\nabla_y \cdot (A(y) \nabla_y u) = f(y), \text{ } u(y) \text{ is } 1\text{-periodic},$$

where $A(y)$ is periodic positive definite matrix and $f \in H^*$ where $H^*$ is the dual of $H$; recall that $H$ is the set of mean zero $H^1_{per}(\mathcal{Y})$ functions – see equation (2.4.5).

The class of coefficients $A(x)$ that we will consider is provided in the following definition.
Definition 7.1. Let \( \alpha, \beta \in \mathbb{R} \), such that \( 0 < \alpha \leq \beta < \infty \). We define \( M(\alpha, \beta, \Omega) \) to be the set of \( d \times d \) matrices \( A(x) = \{a_{ij}(x)\}_{i,j=1}^{d} \in (L^{\infty}(\Omega))^{d \times d} \) such that, for every vector \( \xi \in \mathbb{R}^{d} \) and every \( x \in \Omega \)

(i) \( \xi^{T}A(x)\xi \geq \alpha|\xi|^{2} \),
(ii) \( |A(x)\xi| \leq \beta|\xi| \).

Furthermore, we define \( M_{per}(\alpha, \beta, \mathcal{Y}) \) to be the set of matrices in \( M(\alpha, \beta, \mathcal{Y}) \) with \( \mathcal{Y} \)-periodic coefficients.

In Section 7.3 we will study the Dirichlet problem. The periodic problem will be studied in Section 7.4. The Fredholm alternative for elliptic operators of the form

\[ A = -\nabla \cdot (A(y)\nabla y) + b(y) \cdot \nabla y \]

with periodic boundary conditions will be studied in Section 7.5.

The elliptic operators we study have the form

\[ A = -\nabla \cdot (A(x)\nabla) + b(x) \cdot \nabla + c(x) \]

Operators of this form, and the corresponding PDE, and said to be in divergence form. On the other hand, elliptic differential operators of the form

\[ A = -A(x) : \nabla \nabla + b(x) \cdot \nabla + c(x), \]

are in non–divergence form.

### 7.3 The Dirichlet Problem for Elliptic PDE

First we give the precise definition of a solution. For this we will need to introduce the bilinear form

\[ a[\phi, \psi] = \int_{\Omega} \nabla \psi(x)^{T}A(x)\nabla \phi(x) \, dx, \quad (7.3.1) \]

for \( \phi, \psi \in H_{0}^{1}(\Omega) \). We will use the notation \( \langle \cdot, \cdot \rangle \) for the pairing between \( H_{0}^{1}(\Omega) \) and its dual \( H^{-1}(\Omega) \) (see Chapter 2).

**Definition 7.2.** We will say that \( u \in H_{0}^{1}(\Omega) \) is a weak solution of the boundary value problem (7.2.1) if

\[ a[u, v] = \langle f, v \rangle \quad \forall \ v \in H_{0}^{1}(\Omega). \quad (7.3.2) \]
The Lax–Milgram Theorem 2.41 enables us to prove existence and uniqueness of weak solutions for the class of matrices \( A(x) \) given by Definition 7.1.

**Theorem 7.3.** The Dirichlet problem (7.2.1) with \( A \in M(\alpha, \beta, \Omega) \), \( f \in H^{-1}(\Omega) \) and \( g = 0 \) has a unique weak solution \( u \in H^1_0(\Omega) \). Moreover, the following estimate holds:

\[
\|u\|_{H^1_0(\Omega)} \leq \frac{1}{\alpha} \|f\|_{H^{-1}(\Omega)}. \tag{7.3.3}
\]

**Proof.** We have to verify the conditions of the Lax–Milgram Theorem. We start with coercivity. We use the positive definiteness of the matrix \( A \) to obtain:

\[
a[u, u] = \int_{\Omega} A(x) \nabla u \cdot \nabla u \, dx \\
\geq \alpha \int_{\Omega} |\nabla u|^2 \, dx = \alpha \|u\|^2_{H^1_0(\Omega)}.
\]

Finally we proceed with continuity. We use the \( L^\infty \) bound on the coefficients \( \{a_{ij}(x)\}_{i,j=1}^d \), together with the Cauchy–Schwarz inequality to estimate:

\[
a[u, v] = \int_{\Omega} \nabla v A \nabla u \, dx \\
\leq \beta \int_{\Omega} |\nabla u| |\nabla v| \, dx \\
\leq \beta \|\nabla u\|_{L^2(\Omega)} \|\nabla v\|_{L^2(\Omega)} \\
= \beta \|u\|_{H^1_0(\Omega)} \|v\|_{H^1_0(\Omega)}.
\]

The bilinear form \( a(u, v) \) satisfies the conditions of the Lax–Milgram Theorem and hence there exists a unique solution \( u \in H^1_0(\Omega) \) of (7.3.2).

Now we prove estimate (7.3.3). We have:

\[
\alpha \|u\|^2_{H^1_0(\Omega)} \leq a[u, u] = \langle f, u \rangle \\
\leq \|f\|_{H^{-1}(\Omega)} \|u\|_{H^1_0(\Omega)},
\]

from which the estimate follows. \( \Box \)

If \( f \in L^2 \) then the following bound is also useful.

**Remark 7.4.** Consider the problem (7.2.1) with \( A \in M(\alpha, \beta, \Omega) \) and \( f \in L^2(\Omega) \). Then

\[
\|u\|_{H^1_0(\Omega)} \leq \frac{C_{\Omega}}{\alpha} \|f\|_{L^2(\Omega)},
\]

where \( C_{\Omega} \) is the Poincaré constant for the domain \( \Omega \) defined in Theorem 2.21.
CHAPTER 7. PARTIAL DIFFERENTIAL EQUATIONS

The bound (7.3.3) enables us to obtain information on the solution of 1–parameter families of Dirichlet problems.

**Theorem 7.5.** Assume that there are $\alpha, \beta$ such that the one parameter family of matrices $A^\varepsilon$ belongs to $M(\alpha, \beta, \Omega)$ for all $\varepsilon > 0$. Consider the Dirichlet problem

$$\begin{align*}
-\nabla \cdot (A^\varepsilon(x)\nabla u^\varepsilon) &= f, \quad \text{for } x \in \Omega \\
u^\varepsilon(x) &= 0, \quad \text{for } x \in \partial \Omega,
\end{align*}$$

with $f(x) \in H^{-1}(\Omega)$. Then there is a constant $C$ independent of $\varepsilon$ such that

$$\|u^\varepsilon\|_{H^1_0(\Omega)} \leq C;$$

(7.3.5)

furthermore, there exists a subsequence $\{\varepsilon_n\}_{n \geq 0}$ and a $u(x) \in H^1_0(\Omega)$ such that

$$u^\varepsilon_n(x) \to u(x) \quad \text{strongly in } L^2(\Omega).$$

**Proof.** Estimate (7.3.3) implies (7.3.5). The Rellich compactness theorem, Theorem 2.19, implies that there exists a function $u \in H^1_0(\Omega)$ and a subsequence $\{\varepsilon_n\} \subseteq \varepsilon$ such that $u^\varepsilon_n(x) \to u(x)$ strongly in $L^2(\Omega)$. 

**Remark 7.6.** When studying homogenization for elliptic PDE, in Chapter 13, we will be interested in finding the equation satisfied by $u$.

### 7.4 The Periodic Problem for Elliptic PDE

It is intuitively clear that the solution of the periodic problem can be determined only up to a constant. To ensure uniqueness we need to fix this constant: we work in $H$, the set of mean zero $H^1_{per}(Y)$ functions discussed above. We will use the notation $a_1[\cdot, \cdot]$ to denote the bilinear form

$$a_1[u, v] = \int_Y \nabla_y v A \nabla_y u \, dy \quad \forall u, v \in H.$$  

(7.4.1)

Recall that we denote the pairing between $H$ and its dual $H^*$ by $\langle \cdot, \cdot \rangle_{H,H^*}$. The structure of $H^*$ means that (7.2.2) has a unique solution only when $f$ has mean zero.

**Definition 7.7.** We will say that $u \in H$ is a weak solution of the boundary value problem (7.2.2) if

$$a_1[u, v] = \langle f, v \rangle_{H^*, H} \quad \forall v \in H.$$  

(7.4.2)
Existence and uniqueness of weak solutions to (7.2.2) holds within the space $H$. Indeed, we have the following theorem.

**Theorem 7.8.** The problem (7.2.2) with $A \in M_{\text{per}}(\alpha, \beta, \mathcal{Y})$ and $f \in H^*$ has a unique weak solution $u \in H$. Moreover, the following estimate holds:

$$\|u\|_H \leq \frac{1}{\alpha} \|f\|_{H^*}. \quad (7.4.3)$$

The proof is almost identical to that of Theorem 7.3 and so we omit it. The fact that (7.2.2) has a unique solution only when $f$ has mean zero can also be shown by means of the Fredholm alternative – a topic that we no turn to.

### 7.5 The Fredholm Alternative for Second Order Uniformly Elliptic Operators in Divergence Form

In this section we prove that elliptic differential operators of the form

$$\mathcal{A} = -\nabla \cdot (A(x)\nabla) + b(x) \cdot \nabla, \quad (7.5.1)$$

with periodic coefficients and equipped with periodic boundary conditions, satisfy the Fredholm alternative. Notice that Theorem 2.42 does not apply directly to operator $\mathcal{A}$ since it is an unbounded operator. The main idea will be to study the resolvent operator

$$R_{\mathcal{A}}(\lambda) = (\mathcal{A} + \lambda I)^{-1}, \quad (7.5.2)$$

where $I$ stands for the identity operator on $L^2_{\text{per}}(\mathcal{Y})$ and $\lambda > 0$. We will prove that it is compact; consequently, Fredholm theory can be used.

Our assumptions on the coefficients of $\mathcal{A}$ are

$$A(x) \in M_{\text{per}}(\alpha, \beta, \mathcal{Y}), \quad (7.5.3a)$$

$$A(x) = A^T(x), \quad (7.5.3b)$$

$$b(x) \in (C^1_{\text{per}}(\mathcal{Y}))^d. \quad (7.5.3c)$$

The $L^2$–adjoint of $\mathcal{A}$ is

$$\mathcal{A}^* U = -\nabla (A(x)\nabla U) - \nabla \cdot (b(x)U) \quad (7.5.4)$$

also equipped with periodic boundary conditions. We want to study the PDE

$$\mathcal{A} u = f, \quad u(x) \text{ is } 1\text{–periodic} \quad (7.5.5)$$
and its adjoint
\[ \mathcal{A}^* U = F, \quad U(x) \text{ is } 1\text{–periodic}, \] (7.5.6)
for \( f, F \in L^2_{\text{per}}(\mathcal{Y}) \).

Let \( a[\cdot, \cdot] : H^1_{\text{per}}(\mathcal{Y}) \times H^1_{\text{per}}(\mathcal{Y}) \to \mathbb{R} \) and \( a^*[\cdot, \cdot] : H \times H \to \mathbb{R} \) denote the bilinear forms associated with the operators \( \mathcal{A} \) and \( \mathcal{A}^* \), i.e.
\[
a[u, v] = \int_{\mathcal{Y}} \nabla v A \nabla u \, dx + \int_{\mathcal{Y}} b \cdot \nabla u v \, dx, \quad \forall u, v \in H^1_{\text{per}}(\mathcal{Y})
\]
and
\[
a^*[u, v] = \int_{\mathcal{Y}} \nabla v A \nabla u \, dx - \int_{\mathcal{Y}} \nabla \cdot (bv) u \, dx \quad \forall u, v \in H^1_{\text{per}}(\mathcal{Y})
\]
respectively. As in Section 7.4, we will say that \( u \) and \( U \) are weak solutions of the PDE (7.5.5) and (7.5.6) provided that
\[
a[u, v] = (f, v) \quad \forall v \in H^1_{\text{per}}(\mathcal{Y}) \quad \text{(7.5.7)}
\]
and
\[
a^*[U, V] = (F, V) \quad \forall V \in H^1_{\text{per}}(\mathcal{Y}), \quad \text{(7.5.8)}
\]
respectively.

The main result of this section is contained in the next theorem.

**Theorem 7.9. (Fredholm Alternative for Periodic Elliptic PDE)**

i) Assume conditions (7.5.3). Then precisely one of the following statements holds: either

(a) for every \( f \in L^2_{\text{per}}(\mathcal{Y}) \) there exists a unique solution of (7.5.5); or else

(b) there exists a nontrivial weak solution of the homogeneous problem
\[ \mathcal{A} u = 0, \quad u(x) \text{ is } 1\text{–periodic}. \] (7.5.9)

ii) When assertion (b) holds we have that
\[ 1 \leq \dim(\text{Null}(\mathcal{A})) = \dim(\text{Null}(\mathcal{A}^*)) < \infty. \]

iii) Finally, the boundary value problem (7.5.5) has a weak solution if and only if
\[ (f, U) = 0, \quad \forall U \in \text{Null}(\mathcal{A}). \]
For the proof of this theorem we will need two lemmas.

**Lemma 7.10.** Assume conditions (7.5.3). Then there exist constants \( \nu, \mu > 0 \) such that
\[
|a[u, v]| \leq \nu \|u\|_{H^1} \|v\|_{H^1},
\]
and
\[
\frac{\alpha}{2} \|u\|_{H^1}^2 \leq a[u, u] + \mu \|u\|_{L^2}^2,
\]
for all \( u, v \in H^1_{\text{per}}(\mathcal{Y}) \).

**Proof.** 1. We use the \( L^\infty \) bounds on the coefficients \( A, b \), together with the Cauchy–Schwarz inequality to deduce:
\[
|a(u, v)| \leq \left| \int_\mathcal{Y} A \nabla u \nabla v \, dx + \int_\mathcal{Y} b \cdot \nabla uv \, dx \right| \\
\leq \|A\|_{L^\infty} \int_\mathcal{Y} |\nabla u| |\nabla v| \, dx + \|b\|_{L^\infty} \int_\mathcal{Y} |\nabla u| |v| \, dx \\
\leq C \left( \|\nabla u\|_{L^2} \|\nabla v\|_{L^2} + \|\nabla u\|_{L^2} \|v\|_{L^2} \right) \\
\leq C\|u\|_{H^1} \|v\|_{H^1}.
\]

2. We use now the uniform ellipticity of \( A \) to compute:
\[
\alpha \|\nabla u\|_{L^2}^2 \leq \int_\mathcal{Y} (\nabla u)^T A \nabla u \, dx \\
= a[u, u] - \int_\mathcal{Y} b \nabla uu \, dx \\
\leq a[u, u] + \|b\|_{L^\infty} \int_\mathcal{Y} |\nabla u| |u| \, dx. \quad (7.5.10)
\]
Now we make use of the algebraic inequality
\[
ab \leq \delta a^2 + \frac{1}{4\delta} b^2, \quad \forall \delta > 0.
\]
Using this in the second term on the right hand side of (7.5.10) we obtain
\[
\int_\mathcal{Y} |\nabla u| |u| \, dx \leq \delta \|\nabla u\|_{L^2}^2 + \frac{1}{4\delta} \|u\|_{L^2}^2. \quad (7.5.11)
\]
We chose \( \delta \) so that
\[
\alpha - \|b\|_{L^\infty} \delta = \frac{\alpha}{2}.
\]
We use inequality (7.5.11) with \( \delta \) chosen as above in (7.5.10) to obtain
\[
\frac{\alpha}{2} \|
abla u\|_{L^2}^2 \leq a[u, u] + \frac{1}{4\delta} \|b\|_{L^\infty} \|u\|_{L^2}^2.
\]
We add now \( \frac{\alpha}{2} \|u\|_{L^2}^2 \) on both sides of the above inequality to obtain
\[
\frac{\alpha}{2} \|u\|_{H^1}^2 \leq a[u, u] + \mu \|u\|_{L^2}^2,
\]
with
\[
\mu = \frac{1}{2\alpha} \|b\|_{L^\infty} + \frac{\alpha}{2}.
\]

**Lemma 7.11.** Assume conditions (7.5.3). Take \( \mu \) from Lemma 7.10. Then for every \( \lambda \geq \mu \) and each function \( f \in L^2_{per}(\mathcal{Y}) \) there exists a unique weak solution \( u \in H^1_{per}(\mathcal{Y}) \) of the problem
\[
(A + \lambda I)u = f, \quad u(x) \text{ is 1–periodic.} \tag{7.5.12}
\]

**Proof.** Let \( \lambda \geq \mu \). Define the operator
\[
A_\lambda := A + \lambda I. \tag{7.5.13}
\]
The bilinear form associated to \( A_\lambda \) is
\[
a_\lambda[u, v] = a[u, v] + \lambda(u, v) \quad \forall u, v \in H^1_{per}(\mathcal{Y}). \tag{7.5.14}
\]
Now, Lemma 7.10 together with our assumption that \( \lambda \geq \mu \) imply that the bilinear form \( a_\lambda[u, v] \) is continuous and coercive on \( H^1_{per}(\mathcal{Y}) \). Hence the Lax–Milgram Theorem applies\(^1\) and we deduce the existence and uniqueness of a solution \( u \in H^1_{per}(\mathcal{Y}) \) of the equation
\[
a_\lambda[u, v] = (f, v), \quad \forall v \in H^1_{per}(\mathcal{Y}). \tag{7.5.15}
\]
This is precisely the weak formulation of the boundary value problem (7.5.12).

**Proof of Theorem 7.9.** 1. By Lemma 7.11 there exists, for every \( g \in L^2_{per}(\mathcal{Y}) \), a unique solution \( u \in H^1_{per}(\mathcal{Y}) \) of
\[
a_\mu[u, v] = (g, v), \quad \forall v \in H^1_{per}(\mathcal{Y}). \tag{7.5.16}
\]
\(^1\)We have that \( (f, v) = (f, v) \) defines a bounded linear functional on \( H^1_{per}(\mathcal{Y}) \).
7.5. **FREDHOLM ALTERNATIVE**

We use the resolvent operator defined in (7.5.2) to write the solution of (7.5.16) in the following form:

\[ u = R_A(\mu)g. \]  
(7.5.17)

Consider now equation (7.5.5). We add the term \( \mu u \) on both sides of this equation to obtain

\[ A_\mu u = \mu u + f, \]

where \( A_\mu \) is defined in (7.5.13). The weak formulation of this equation is

\[ a_\mu [u, v] = (\mu u + f, v), \quad \forall v \in H^1_{\text{per}}(\mathcal{Y}). \]

We can rewrite this as an integral equation (see (7.5.17))

\[ u = R_A(\mu)(\mu u + f), \]

or, equivalently,

\[ (I - K)u = h, \]

where

\[ K := \mu R_A(\mu), \quad h = R_A(\mu)f. \]

2. Now we claim that the operator \( K : L^2_{\text{per}}(\mathcal{Y}) \to L^2_{\text{per}}(\mathcal{Y}) \) is compact. Indeed, let \( u \) be the solution of (7.5.16) which is given by (7.5.17). We use the second estimate in Lemma 7.10, the definition of the bilinear form (7.5.14) and the Cauchy–Schwarz inequality in (7.5.16) to obtain

\[ \frac{\alpha}{2} \| u \|_{H^1}^2 \leq a_\mu [u, u] = (g, u) \leq \| g \|_{L^2} \| u \|_{L^2} \leq \| g \|_{L^2} \| u \|_{H^1}. \]

Consequently,

\[ \| u \|_{H^1} \leq C \| g \|_{L^2}. \]

We use now (7.5.17), the definition of \( K \) and the above estimate to deduce that

\[ \mu \| u \|_{H^1} = \| Kg \|_{H^1} \leq C \mu \| g \|_{L^2}. \]  
(7.5.18)

By the Rellich compactness theorem \( H^1_{\text{per}}(\mathcal{Y}) \) is compactly embedded in \( L^2_{\text{per}}(\mathcal{Y}) \) and consequently estimate (7.5.18) implies that \( K \) maps bounded sets in \( L^2_{\text{per}}(\mathcal{Y}) \) into compact ones in \( L^2_{\text{per}}(\mathcal{Y}) \). Hence, it is a compact operator.

3. We apply now the Fredholm alternative (Theorem 2.42) to the operator \( K \): either
a. there exists a unique $u \in L_{\text{per}}^2(\mathcal{Y})$ such that

$$(I - K)u = h,$$  \hspace{1cm} (7.5.19)

or

b. there exists a non trivial solution $u \in L_{\text{per}}^2(\mathcal{Y})$ of the homogeneous equation

$$(I - K)u = 0.$$  \hspace{1cm} (7.5.20)

Let us assume that (7.5.19) holds. From the preceding analysis we deduce that there exists a unique weak solution $u \in H_{\text{per}}^1(\mathcal{Y})$ of (7.5.5). Assume now that (7.5.20) holds. Let $N$ and $N^*$ denote the dimensions of null spaces of $I - K$ and $I - K^*$, respectively. From Theorem 2.42 we know that $N = N^*$. Moreover, it is straightforward to prove that

$$u \in \mathcal{N}(I - K) = 0 \iff a[u, \phi] = 0, \ \forall u \in H_{\text{per}}^1(\mathcal{Y})$$

and

$$v \in \mathcal{N}(I - K^*) = 0 \iff a^*[v, \phi] = 0, \ \forall \phi \in H_{\text{per}}^1(\mathcal{Y}),$$

Thus, the Fredholm alternative for $K$ implies the Fredholm alternative for $A$ (within the context of weak solutions.)

4. Now we prove the third part of the theorem. Let $v \in \mathcal{N}(I - K^*)$. By Theorem 2.42 we know that (7.5.20) has a solution if and only if

$$(h, v) = 0 \ \forall v \in \mathcal{N}(I - K^*).$$

We compute

$$(h, v) = (R_A(\mu)f, v) = \frac{1}{\mu}(Kf, v)$$

$$= \frac{1}{\mu}(f, K^*v) = \frac{1}{\mu}(f, v).$$

Hence, problem (7.5.5) has a weak solution if and only if

$$(f, v) = 0 \ \forall v \in \mathcal{N}(A^*).$$

This completes the proof of the theorem. \hfill \Box
Example 7.12. Let \( f \in (L^2_{\text{per}}(Y)) \) and assume that \( A(x) \) satisfies assumptions (7.5.3a) and (7.5.3b). Then the problem
\[
\alpha_1[u, v] = (f, v), \quad \forall v \in H^1_{\text{per}}(Y),
\]
where \( \alpha_1[\cdot, \cdot] \) was defined in (7.4.1), has a unique solution \( u \in H \) if and only if
\[
\langle f, 1 \rangle = 0.
\] (7.5.21)
Indeed, consider the homogeneous adjoint equation
\[ A^*U = 0. \]
Clearly, the constant function (say, \( U = 1 \)) is a solution of this equation. The uniform ellipticity of the matrix \( A(x) \) implies that
\[
\int_Y |\nabla U| \, dx = 0,
\]
and hence, the constant solution is unique. Since assumptions (7.5.3a) and (7.5.3b) are satisfied, Theorem 7.9 applies and the result follows.

7.6 Parabolic PDE

In this section we describe the basic theory of parabolic PDE, in non-divergence form. Specifically we study the initial value (Cauchy) problem
\[
\frac{\partial u}{\partial t} = \mathcal{L}u + F(x, t), \quad \text{for } (x, t) \in \mathbb{R}^d \times [0, T],
\]
\[
u(x, 0) = f(x), \quad \text{for } x \in \mathbb{R}^d,
\]
where
\[
\mathcal{L} := b(x) \cdot \nabla + \frac{1}{2} A(x) : \nabla \nabla.
\] (7.6.2)

We will make the following assumptions.
\[
b(x) \in (C^\infty_b(\mathbb{R}^d), \mathbb{R}^d); \]
\[
A(x) \in (C^\infty_b(\mathbb{R}^d), \mathbb{R}^{d \times d});
\]
\[
F(x, t) \in C^\infty_0(\mathbb{R}^d \times [0, T]);
\]
\[
A \in M(\alpha, \beta, \mathbb{R}^d).
\] (7.6.3d)
CHAPTER 7. PARTIAL DIFFERENTIAL EQUATIONS

Theorem 7.13. Under assumptions (7.6.3) there exists a unique solution \( u(x, t) \in C^{1,2}([0, t] \times \mathbb{R}^d) \cup C^\infty([0, t] \times \mathbb{R}^d) \) to the Cauchy problem (7.6.1). Furthermore the following estimate holds.

\[
\|u\|_{L^\infty(\mathbb{R}^d \times [0,t])} \leq \|f\|_{L^\infty(\mathbb{R}^d)} + \int_0^t \|F(\cdot, s)\|_{L^\infty(\mathbb{R}^d)} \, ds. \tag{7.6.4}
\]

Estimate (7.6.4) is a consequence of the maximum principle for parabolic PDE.

The solution \( u(x, t) \) in the above theorem is a classical solution. The weak formulation of the Cauchy problems (7.6.1) is obtained by multiplying the equation by a smooth, compactly supported function and integrating by parts. See Exercise 5.

7.7 Hyperbolic PDE

In this chapter we investigate some basic properties of solutions to the Cauchy problem for the linear transport PDE

\[
\frac{\partial u}{\partial t} + \nabla \cdot (a(x)u) = 0, \quad (x, t) \in \mathbb{R}^d \times \mathbb{R}^+, \tag{7.7.1a}
\]

\[
u = f(x), \quad (x, t) \in \mathbb{R}^d \times \{0\}, \tag{7.7.1b}
\]

where \( a(x) : \mathbb{R}^d \to \mathbb{R}^d \) is a smooth vector field and \( f(x) \) is a smooth scalar field. Notice that equation (7.7.1a) is the Liouville equation for the differential equation

\[
\frac{dx}{dt} = a(x).
\]

In this chapter we will define an appropriate concept of solution for (7.7.1), we will state the basic existence and uniqueness theorem and derive some basic energy estimates.

As is always the case, the weak formulation of (14.2.4) involves multiplication by a smooth function, integration over \( \mathbb{R}^+ \times \mathbb{R}^d \) and integration by parts. We have the following definition

Definition 7.14. Assume that \( a(x) \in C^\infty_b(\mathbb{R}^d, \mathbb{R}^d) \) and that \( f(x) \in C^\infty_b(\mathbb{R}^d) \). A function \( u(x, t) \in L^\infty(\mathbb{R}^+ \times \mathbb{R}^d) \) is a weak solution of (7.7.1) provided that

\[
\int_{\mathbb{R}^d} \int_{\mathbb{R}^+} \left( \frac{\partial \phi(x, t)}{\partial t} + a(x) \cdot \nabla_x \phi(x, t) \right) u(x, t) \, dx \, dt + \int_{\mathbb{R}^d} f(x) \phi(x, 0) \, dx = 0, \tag{7.7.2}
\]

for every \( \phi(x, t) \in C^\infty_0(\mathbb{R}^+ \times \mathbb{R}^d) \).
Theorem 7.15. Assume that \( a(x) \in C^\infty_b(\mathbb{R}^d, \mathbb{R}^d) \) and that \( f(x) \in C^\infty_b(\mathbb{R}^d) \). Then there exists a unique weak solution of (7.7.1).

The proof uses the method of characteristics introduced in Chapter 4.

7.8 Discussion and Bibliography

The material of this section is pretty standard and can be found in many books on partial differential equations and functional analysis, such as [25, 56, 45, 129]. Our treatment of the Dirichlet problem follows closely [45, Ch. 6]. Our discussion about periodic boundary conditions is based on [30, Ch. 4]. The Fredholm theory for the Dirichlet problem is developed in [45, Sec. 6.2.3].

In the case where the data of the problem is regular enough so that the Dirichlet problem (7.2.1) admits a classical solution (i.e. a function \( u \in C^2(\Omega) \cap C^0(\overline{\Omega}) \) satisfying (7.2.1)), then the weak and classical solutions coincide. See e.g. [56].

We saw in this chapter that the analysis of operators in divergence form is based on energy methods within appropriate function spaces. On the other hand, for PDE in non–divergence form techniques based on the maximum principle are more suitable. The maximum principle for elliptic PDE is studied in [56] and for parabolic PDE in [125]. Of course, provided that the coefficients are \( C^1 \), we can rewrite a divergence form PDE in non–divergence form and vice versa, by introducing terms that involve first order derivatives. Operators in non–divergence form appear more natural in the probabilistic theory of diffusion, as generators of Markov processes. The dimensionality of the null space of the \( L^2 \)–adjoint of a non–divergence form second order uniformly elliptic operator is related to the ergodic theory of Markov processes, see [116, 120].

Turning now to parabolic PDE, the proof of Theorem 7.13 can be found in [51]; see also[52]. Similar theorems hold for parabolic PDE with time–dependent coefficients. Parabolic PDE can also be studied using probabilistic methods. Indeed, under appropriate assumptions on the coefficients, the solution of (7.6.1) admits a probabilistic interpretation, after noting that it is the backward Kolmogorov equation for an SDE, when \( F \equiv 0 \). Probabilistic proofs of existence and uniqueness of solutions to parabolic PDE can be found in [28]. In this book the case of unbounded coefficients is also studied.
CHAPTER 7. PARTIAL DIFFERENTIAL EQUATIONS

7.9 Exercises

1. Use the Lax–Milgram lemma to prove Theorem 7.8.

2. Use Theorem 7.9 to derive Theorem 7.8.

3. State and prove a result analogous to that discussed in Remark 7.4 for the periodic problem (7.2.2).

4. Prove the Fredholm alternative for operator $A$ defined in (7.5.1) under assumptions analogous to (7.5.3), but adapted to the case of Dirichlet boundary conditions.

5. Consider the parabolic PDE

   \[
   \frac{\partial u}{\partial t} = \mathcal{L} u + F(x), \quad \text{for} \ (x,t) \in \Omega \times [0,T],
   \]

   \[
   u(x,t) = 0, \quad \text{for} \ x \in \partial \Omega,
   \]

   \[
   u(x,0) = f(x), \quad \text{for} \ x \in \Omega.
   \]

   where

   \[
   \mathcal{L} := \Delta.
   \]

   Formulate a notion of weak solution analogous to that developed in the Elliptic case in the previous chapter.

6. Prove that the equation (7.9.1) has a unique steady solution $\overline{u}(x)$. Prove that $u(x, t) \rightarrow \overline{u}(x)$ as $t \rightarrow \infty$.

7. Use the method of characteristics to solve the equation

   \[
   \frac{\partial u(x,t)}{\partial t} + a(x) \frac{\partial u(x,t)}{\partial x} = b(x) \quad (x,t) \in \mathbb{R} \times \mathbb{R}^+, \quad u(x,0) = 0 \quad (x,t) \in \mathbb{R} \times \{0\}.
   \]

8. Use the method of characteristics to solve the Burger’s equation

   \[
   \frac{\partial u(x,t)}{\partial t} + u(x,t) \frac{\partial u(x,t)}{\partial x} = 0 \quad (x,t) \in \mathbb{R} \times \mathbb{R}^+, \quad u(x,0) = 0 \quad (x,t) \in \mathbb{R} \times \{0\}.
   \]
Part II

Perturbation Expansions
Chapter 8

Invariant Manifolds for ODE

8.1 Introduction

Perhaps the simplest situation where variable reduction occurs in dynamical systems is that of attractive invariant manifolds. These manifolds \textit{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.

8.2 Full Equations

We consider ODE and write \( z \) solving (4.1.1) as \( z = (x^T, y^T)^T \), where

\[
\begin{align*}
\frac{dx}{dt} &= f(x, y), \\
\frac{dy}{dt} &= \frac{1}{\varepsilon} g(x, y),
\end{align*}
\]

and \( \varepsilon \ll 1 \). Here \( x \in \mathcal{X} \) and \( y \in \mathcal{Y} \).

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

\[
\frac{d}{dt} \varphi^t_{\xi}(y) = g(\xi, \varphi^t_{\xi}(y)), \quad \varphi^0_{\xi}(y) = y.
\]

We assume that

\[
\lim_{{t \to \infty}} \varphi^t_{\xi}(y) = \eta(\xi)
\]
exists, is independent of \( y \) and that the convergence is uniform in \( \xi \). Roughly speaking \( y(t) \) solving (8.2.1) is given by 
\[ y(t) \approx e^{t/\varepsilon} x(0) \]
for times \( t \) which are small compared with 1 so that \( x \) has not evolved very much. If we then look at scales which are large compared with \( \varepsilon \), so that \( y \) is close to its equilibrium point (for example if \( t = O(\varepsilon^2) \)), then 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 heuristics more precise.

Notice that the generator \( \mathcal{L} \) for (8.2.1) has the form
\[ \mathcal{L} = \frac{1}{\varepsilon} \mathcal{L}_0 + \mathcal{L}_1 \]  
(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:
\[ \frac{\partial v}{\partial t} = \mathcal{L}_0 u, \quad v(y, 0) = \phi(y) \]  
(8.2.5)
Result 4.4 shows that
\[ v(y, t) = \phi(\varphi^{t}_x(y)) . \]
Thus
\[ v(y, t) \to \phi(\eta(x)), \quad t \to \infty \]  
(8.2.6)
This is related to ergodicity, as equation (8.2.6) shows that the functional \( v(y, t) \) exhibits no dependence on initial data, asymptotically as \( t \to \infty \). Compare with the discussion on ergodicity in Chapter 4 and Theorem 4.7 in particular.

8.3 Simplified Equations

Define the vector field \( F_0(x) \) by
\[ F_0(x) = f(x, \eta(x)) . \]

Result 8.1. For \( \varepsilon \ll 1 \) and \( t = O(1) \) \( x \) solving (4.1.1) is approximated by \( X \) solving
\[ \frac{dX}{dt} = F_0(X) \]  
(8.3.1)
This is the leading order approximation in $\varepsilon$. Keeping the next order yields the refined approximation
\[
\frac{dX}{dt} = F_0(X) + \varepsilon F_1(X),
\] (8.3.2)
where
\[
F_1(x) = \nabla_y f(x, \eta(x)) (\nabla_y g(x, \eta(x)))^{-1} \nabla_x \eta(x) f(x, \eta(x)).
\]
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 $y = \Psi(x)$. Such a graph is invariant 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
\[
\Psi(x) = \Psi_0(x) + \varepsilon \Psi_1(x) + O(\varepsilon^2).
\]
This is our first example of a multiscale expansion.

Substituting and equating successive powers of $\varepsilon$ to zero yields the hierarchy
\[
O\left(\frac{1}{\varepsilon}\right) \quad g(x, \Psi_0(x)) = 0,
O(1) \quad \nabla_y g(x, \Psi_0(x)) \Psi_1(x) = \nabla \Psi_0(x) f(x, \Psi_0(x)).
\]
Notice that equations (8.2.2,8.2.3) together imply that $g(\xi, \eta(\xi)) = 0$ for all $\xi$. Hence the $O\left(\frac{1}{\varepsilon}\right)$ equation above may be satisfied by choosing $\Psi_0(x) = \eta(x)$, giving the approximation (8.3.1). Since the rate of convergence in (8.2.3) is assumed to be uniform it is natural to assume that $y = \eta(\xi)$ is a hyperbolic equilibrium point of (8.2.2), so that $\nabla_y g(x, \eta(x))$ is invertible. Setting $\Psi_0(x) = \eta(x)$ in the $O(1)$ equation, and inverting yields
\[
\Psi_1(x) = \nabla_y g(x, \eta(x))^{-1} \nabla \eta(x) f(x, \eta(x)).
\]
Thus

\[ f(x, \Psi(x)) = f(x, \Psi_0(x) + \varepsilon \Psi_1(x) + \mathcal{O}(\varepsilon^2)) = f(x, \Psi_0(x)) + \varepsilon \nabla_y f(x, \Psi_0(x)) \Psi_1(x) + \mathcal{O}(\varepsilon^2) = f(x, \eta_0(x)) + \varepsilon \nabla_y f(x, \eta(x)) \Psi_1(x) + \mathcal{O}(\varepsilon^2) \]

and the refined approximation (8.3.2) follows.

### 8.5 Applications

#### 8.5.1 Linear Fast Dynamics

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

\[
\begin{align*}
\frac{dx}{dt} &= f(x, y) \\
\frac{dy}{dt} &= -y + \varepsilon g(x) + \tilde{g}(x)
\end{align*}
\]

(8.5.1)

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

\[
\varphi_t^\varepsilon(y) = e^{-t}y + \int_0^t e^{s-t} \tilde{g}(\xi) ds = e^{-t}y + (1 - e^{-t}) \tilde{g}(\xi).
\]

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 (8.3.2) is

\[
\frac{dX}{dt} = f(X, \tilde{g}(X)) \left( 1 - \varepsilon \frac{df}{dy}(X, \tilde{g}(X)) \tilde{g}'(X) \right).
\]

#### 8.5.2 Large Time Dynamics

The statement of the result concerning simplified dynamics concerns 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

\[
\begin{align*}
\frac{dx_1}{dt} &= -x_2 - x_3 \\
\frac{dx_2}{dt} &= x_1 + \frac{1}{5}x_2 \\
\frac{dx_3}{dt} &= \frac{1}{5} + y - 5x_3 \\
\frac{dy}{dt} &= -\frac{y}{\varepsilon} + \frac{x_1x_3}{\varepsilon},
\end{align*}
\] (8.5.2)

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

\[
\begin{align*}
\frac{dX_1}{dt} &= -X_2 - X_3 \\
\frac{dX_2}{dt} &= X_1 + \frac{1}{5}X_2 \\
\frac{dX_3}{dt} &= \frac{1}{5} + X_3(X_1 - 5).
\end{align*}
\] (8.5.3)

A comparison of the numerically generated attractors for both systems shows that, when projected onto the common space \( \mathcal{X} = \mathbb{R}^3 \), they are very close.

### 8.5.3 Centre Manifold

Consider the equations

\[
\begin{align*}
\frac{dx}{dt} &= \lambda x + \sum_{i=0}^{2} a_i x^i y^{2-i}, \\
\frac{dy}{dt} &= x - y + \sum_{i=0}^{2} b_i x^i y^{2-i}.
\end{align*}
\]

At \( \lambda = 0 \) an eigenvalue of the system linearized at the origin passes through zero and so we expect to find a bifurcation and a resulting centre manifold. To construct this manifold set

\[
x \rightarrow \varepsilon x, y \rightarrow \varepsilon y, \lambda \rightarrow \varepsilon \lambda, t \rightarrow \varepsilon^{-1}t.
\]
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. 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 centre manifold

\[ \frac{dX}{dt} = \lambda X + AX^2 \]

where \( A = \sum_{i=0}^{2} a_i. \)

### 8.6 Discussion and Bibliography

Early studies of the reduction of ODE with attracting slow manifold into differential-algebraic equations includes the independent work of Levinson and of Tikhonov (see O’Malley [113] and Tikhonov et al. [149]). 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 [27], Wiggins [154]), slow manifolds (Kreiss [83]) and inertial manifolds (Constantin et al. [32]).

### 8.7 Exercises

1. Consider the equations

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

Show that the scaling

\[ x \to \varepsilon x, \ y \to \varepsilon^2 y, \ \lambda \to \varepsilon^2 \lambda, \ t \to \varepsilon^{-2} t \]
8.7. EXERCISES

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

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

where $A = a_0 + a_1 b_2$.

2. Consider the equations

$$\frac{dx}{dt} = Ax + \epsilon f_0(x, y),$$

$$\frac{dy}{dt} = -\frac{1}{\epsilon} By + g_0(x, y),$$

for $\epsilon \ll 1$ 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 $x$. 
Chapter 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.

9.2 Full Equations

We work in the set-up of Chapter 5 and consider the backward equation for $v_i(t) = \mathbb{E}(\phi(z(t))|z(0) = i)$, namely

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

We assume that the generator $Q$ takes the form

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

with $\varepsilon \ll 1$. 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 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 $I := \mathcal{I}_x \times \mathcal{I}_y$ with $\mathcal{I}_x, \mathcal{I}_y \subseteq \{1, 2, \cdots \}$. We let $q((i, k), (j, l))$ denote the element of the generator.

---

1In 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.
CHAPTER 9. AVERAGING FOR MARKOV CHAINS

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 \(A_0\) has a one-dimensional null-space and for each \(i \in \mathcal{I}_x\)\(^3\)

\[
\sum_l a_0(k, l; i) = 0, \quad (i, k) \in \mathcal{I}_x \times \mathcal{I}_y,
\]

\[
\sum_k \rho^\infty(k; i)a_0(k, l; i) = 0, \quad (i, l) \in \mathcal{I}_x \times \mathcal{I}_y.
\] (9.2.3)

This is the index form of equations (5.4.2) with \(Q\) 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 a probability density for each \(i \in \mathcal{I}_x\).

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

\[
q_0((i, k), (j, l)) = a_0(k, l; i)\delta_{ij},
\]

\[
q_1((i, k), (j, l)) = a_1(i, j; k)\delta_{kl}.
\] (9.2.4)

Here \(\delta_{ij}\) is the usual Kronecker delta. In \(Q_0\) (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 these are indeed generators notice that non-diagonal entries \((i, k) \neq (j, l)\) are non-negative because \(A_0\) and \(A_1\) are generators. Also

\[
\sum_{j,l} q_0((i, k), (j, l)) = \sum_{j,l} a_0(k, l; i)\delta_{ij}
\]

\[
= \sum_l a_0(k, l; i)
\]

\[
= 0
\]

\(^2\)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.
9.3. SIMPLIFIED EQUATIONS

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:

\[ \bar{q}_{1}(i, j) = \sum_{k} \rho^{\infty}(k; i) a_{1}(i, j; k). \]

(9.3.1)

Notice that \( \bar{q}_{1}(i, j) \geq 0 \) for \( i \neq j \) because \( \rho^{\infty}(k; i) \geq 0 \) for \( i \neq j \). Furthermore

\[ \sum_{j} \bar{q}_{1}(i, j) = \sum_{k} \rho^{\infty}(k; i) \left( \sum_{j} q_{1}(i, j; k) \right) = 0. \]

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 \( t = O(1) \) the implied dynamics in \( x \) is approximately Markovian with generator \( \bar{Q}_{1} \) and governed by the backward equation

\[ \frac{d\nu_{0}}{dt} = \bar{Q}_{1}\nu_{0}. \]

(9.3.2)

As discussed above, the generator of a Markov chain on \( \mathcal{I}_{x} \) alone, and the dynamics in \( \mathcal{I}_{y} \) has been eliminated through averaging. We now provide justification for this elimination of variables, by means of perturbation expansion.
CHAPTER 9. AVERAGING FOR MARKOV CHAINS

9.4 Derivation

The method used is to show that backward equation for the full Markov chain in 
\((x, y) \in \mathcal{I}_x \mathcal{I}_y\) can be approximated by the backward equation

\[
\frac{dv_0}{dt} = Q_1 v_0
\]

for \(x\) alone.

We consider equation (9.2.1) under (9.2.2). We have the backward equation

\[
\frac{dv}{dt} = \left(\frac{1}{\varepsilon} Q_0 + Q_1\right) v.
\]

Unlike the previous section, where we solved 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

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

Substituting and equating powers of \(\varepsilon\) we find

\[
\mathcal{O}\left(\frac{1}{\varepsilon}\right) \quad Q_0 v_0 = 0,
\] (9.4.2a)

\[
\mathcal{O}(1) \quad Q_0 v_1 = -Q_1 v_0 + \frac{dv_0}{dt}.
\] (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

\[
v_0(i, k, t) = v_0(i, t) \mathbf{1}(k)
\] (9.4.3)

where \(\mathbf{1}(k) = 1, \quad \forall k \in \mathcal{I}_y\). The operator \(Q_0\) is singular and hence for (9.4.2b) to have a solution, the Fredholm alternative implies that we must impose the solvability condition

\[
-Q_1 v_0 + \frac{dv_0}{dt} \perp \text{Null } \{Q_0^T\}.
\] (9.4.4)

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

\[
\sum_{k,l} \rho^\infty(k; i)c(i) q_0((i, k), (j, l)) = 0,
\] (9.4.5)
for any vector $c = \{c(i)\}$ on $I_x$. Using (9.4.3) we find that
\[
\frac{dv_0}{dt} - Q_1v_0 = \frac{dv_0}{dt}(i, t)1(k) - \sum_{j,l} a_1(i, j; k)\delta_{kl}v_0(j, t)1(l)
= \left(\frac{dv_0}{dt}(i, t) - \sum_{j} a_1(i, j; k)v_0(j, t)\right)1(k).
\]

Imposing the solvability condition (9.4.4) by means of (9.4.5) we obtain
\[
\sum_{k,i} \rho^\infty(k; i)c(i)\left(\frac{dv_0}{dt}(i, t) - \sum_{j} a_1(i, j; k)v_0(j, t)\right),
\]
which implies that
\[
\sum_{i} c(i)\left(\frac{dv_0}{dt}(i, t) - \sum_{j} Q_1(i, j)v_0(j, t)\right) = 0.
\]
Since $c$ is an arbitrary vector on $I_x$ 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 $I_x = I_y = \{1, 2\}$. Thus we have a four state Markov chain on $I = I_x \times I_y$. We assume that the generators of the Markov chains on $I_y$ and $I_x$ 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 $i \in I_x$ (resp. $k \in I_y$) is a fixed parameter. The parameters $\theta_i, \phi_k, \alpha_k, \beta_k$ 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
\[
Q_0 = \begin{pmatrix} -\theta_1 & \theta_1 & 0 & 0 \\ \phi_1 & -\phi_1 & 0 & 0 \\ 0 & 0 & -\theta_2 & \theta_2 \\ 0 & 0 & \phi_2 & -\phi_2 \end{pmatrix}
\] (9.5.1)
CHAPTER 9. AVERAGING FOR MARKOV CHAINS

and

\[
Q_1 = \begin{pmatrix}
-\alpha_1 & 0 & \alpha_1 & 0 \\
0 & -\alpha_2 & 0 & \alpha_2 \\
\beta_1 & 0 & -\beta_1 & 0 \\
0 & \beta_2 & 0 & -\beta_2
\end{pmatrix}.
\] (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

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

9.6 Discussion and Bibliography

Two recent monographs where multiscale problems for Markov chains are studied are [156], [155]. See also [136] 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 [50]

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 \(\bar{Q}_1\) and interpret your findings in the light of your answer to the previous question.
Chapter 10

Averaging for ODE and SDE

10.1 Introduction

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

10.2 Full Equations

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

$$
\begin{align*}
\frac{dx}{dt} &= f(x, y), \\
\frac{dy}{dt} &= \frac{1}{\varepsilon}g(x, y) + \frac{1}{\sqrt{\varepsilon}}\beta(x, y) \frac{dV}{dt},
\end{align*}
$$

(10.2.1)

with $\varepsilon \ll 1$ and $V$ a standard Brownian motion. Here $x \in X$, $y \in Y$, $z \in Z$ as discussed in sections 4.1 and 6.1.

Our starting point is to analyze the behavior of the fast dynamics with $x$ being a fixed parameter. 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, over an appropriate invariant measure.
Let \( \varphi^\xi(y) \) be the solution operator of the fast dynamics with \( x \) a fixed parameter. To be precise, for fixed \( \xi \),

\[
\frac{d}{dt} \varphi^\xi(y) = g(\xi, \varphi^\xi(y)) + \beta(\xi, \varphi^\xi(y)) \frac{dV}{dt}, \quad \varphi^\xi_0(y) = y.
\] (10.2.2)

As in Chapter 8, \( y(t) \) solving (10.2.2) is given by \( y(t) \approx \varphi^{t/\varepsilon}(0) \) for times \( t \) which are small compared with 1, so that \( x \) has not evolved very much. If (10.2.2) is ergodic then \( y(t) \) will traverse its \( \xi \) dependent invariant measure in any time large compared to \( \varepsilon \) and it is natural to average \( y \) in the \( x \) equation, against this invariant measure with \( \xi = x \). We now make these heuristics precise.

In the case where \( \beta \equiv 0 \) then \( \varphi^\xi(y) \) coincides the solution of (8.2.2). When \( \beta \neq 0 \), note that \( \varphi^\xi(y) \) depends on the Brownian motion \( \{V(s)\}_{s \in [0,t]} \) and hence is a random variable. Rather than assuming convergence to a fixed point, as we did in (8.2.3), we assume here that \( \varphi^\xi(y) \) is ergodic.

To be precise we assume that the measure defined by

\[
\mu_x(A) = \lim_{T \to \infty} \frac{1}{T} \int_0^T I_A(\varphi^\xi(y)) \, dt, \quad A \subseteq \mathcal{Y},
\] (10.2.3)

exists, for \( I_A \) the indicator function of arbitrary Borel sets \( A \subseteq \mathcal{Y} \). When working with an SDE (\( \beta \neq 0 \)) that it is natural to assume that \( \mu_x(\cdot) \) has a density \( \mu_x(dy) = \rho_\infty(y; x) \, dy \). However we will illustrate by means of an example arising in Hamiltonian mechanics that this assumption is not necessary. Note that the situation in Chapter 8 corresponds to \( \mu_x(dy) = \delta(y - \eta(x)) \, dy \), that is the ergodic invariant measure is a Dirac mass centered on the invariant manifold.

We define the generators

\[
\mathcal{L}_0 = g(x, y) \cdot \nabla_y + \frac{1}{2} B(x, y) : \nabla_y \nabla_y
\]
\[
\mathcal{L}_1 = f(x, y) \cdot \nabla_x
\] (10.2.4)

where \( B(x, y) = \beta(x, y) \beta(x, y)^T \). In terms of the generator \( \mathcal{L}_0 \), our ergodicity assumption becomes the statement that \( \mathcal{L}_0 \) has one dimensional null-space characterized by

\[
\mathcal{L}_0 1(y) = 0, \quad \mathcal{L}_0^* \rho_\infty(y; x) = 0.
\] (10.2.5)

Here \( 1(y) \) denotes constants in \( y \).
10.3. SIMPLIFIED EQUATIONS

In the case where $\mathcal{Y} = \mathbb{T}^d$ the operators $L_0$ and $L_0^*$ are equipped with periodic boundary conditions. The interrelations amongst the characterizations of ergodicity that we use here are made precise in the case where $B(x, y)$ is strictly positive-definite, uniformly in $(x, y) \in \mathcal{X} \times \mathcal{Y}$, in Theorem 6.10. In more general situations, such as when $\mathcal{Y} = \mathbb{R}^d$, 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.

10.3 Simplified Equations

Define a vector field $F$ by

$$F(x) = \int_{\mathcal{Y}} f(x, y) \mu_x(dy).$$  \hspace{1cm} (10.3.1)

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

$$\frac{dX}{dt} = F(X).$$ \hspace{1cm} (10.3.2)

Note that, in the case we considered in Chapter 8, where $\mu_x(dy) = \delta(dy - \eta(x))dy$, we obtain

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

This is precisely the vector field in (8.3.1) and so the simplified equations in Chapter 8 are a special case of those derived here. We will derive Result 10.1 in the case where $\beta$ is non-zero and we will use the density given by $\mu_x(dy) = \rho^\infty(y; x)dy$.

The expression for $F$ given above is useful for theoretical purposes, and when the measure $\mu$ is known explicitly. However, for numerical purposes the following representation is often useful:

**Result 10.2.** An alternative representation of $F(x)$ is via a time-average:

$$F(x) = \lim_{T \to \infty} \frac{1}{T} \int_0^T f(x, \varphi_t^x(y(0)))dt.$$  \hspace{1cm} (10.3.3)

This representation is found by using (10.2.3) to evaluate (10.3.1).
10.4 Derivation

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

\[ v(x, y, t) = \mathbb{E}\left( \phi(x(t), y(t)) | x(0) = x, y(0) = y \right). \]

The backward equation (6.3.3) for the SDE (10.2.1) is

\[ \frac{\partial v}{\partial t} = \frac{1}{\varepsilon} L_0 v + L_1 v. \quad (10.4.1) \]

Here \( L_0, L_1 \) are given by (10.2.4) and \( z \) in (6.3.3) is \((x, y)\) here. Note that \( L_0 \) is a differential operator in \( y \), in which \( x \) appears as a parameter. Thus we must equip it with boundary conditions. For simplicity we consider the case where \( \mathcal{Y} = T^d \) so that periodic boundary conditions are used. 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 \( L_0 \).

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

\[ \mathcal{O}(1/\varepsilon) \quad L_0 v_0 = 0, \quad v_0 \text{ is } 1\text{-periodic}, \quad (10.4.2a) \]

\[ \mathcal{O}(1) \quad L_0 v_1 = -L_1 v_0 + \frac{\partial v_0}{\partial t}, \quad v_1 \text{ is } 1\text{-periodic}. \quad (10.4.2b) \]

Equation (10.4.2a) implies that \( v_0 \) is in the null space of \( L_0 \) and hence, by (10.2.5) and ergodicity, is a function only of \((x, t)\). The Fredholm alternative for (10.4.2b) shows that

\[ -L_1 v_0 + \frac{\partial v_0}{\partial t} \perp \text{Null } \{ L_0^* \}. \]

By (10.2.5) this implies that

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

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

\[ \frac{\partial v_0}{\partial t} - (\int_{\mathcal{Y}} f(x, y) \mu(y) dy) \cdot \nabla_x v_0(x, y) = 0 \]
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 4.4 shows that we have the required result.

## 10.5 Applications

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

### 10.5.1 A Skew-Product SDE

Consider the equations
\begin{align*}
\frac{dx}{dt} &= (1 - y^2)x, \\
\frac{dy}{dt} &= -\frac{\alpha}{\varepsilon}y + \sqrt{\frac{\lambda}{\varepsilon}} dV.
\end{align*}
(10.5.1)

Here $\mathcal{X} = \mathcal{Y} = \mathbb{R}$. It is of interest to know whether $x$ will explode in time, or remain bounded. We can get insight into this question in the limit $\varepsilon \to 0$ by deriving the averaged equations. The ergodic measure for $y$ is a Gaussian $\mathcal{N}(0, \frac{\lambda}{2\alpha})$. Note that this measure does not depend on $x$ and hence has density $\rho_\infty(y)$ only. The average vector field $F$ is here defined by
$$F(x) = \left(1 - \int_\mathbb{R} \rho_\infty(y)y^2 dy\right)x$$
where $\rho_\infty$ is the density associated with Gaussian $\mathcal{N}(0, \frac{\lambda}{2\alpha})$. Thus
$$\int_\mathbb{R} \rho_\infty(y)y^2 dy = \frac{\lambda}{2\alpha}$$
and
$$F(x) = \left(1 - \frac{\lambda}{2\alpha}\right)x.$$

Hence trajectories of $x$ will explode if $\lambda < 2\alpha$ and will contract if $\lambda > 2\alpha$. If $\lambda = 2\alpha$ then the averaged vector field is zero. In this situation we need to rescale.
time \( t \mapsto t/\varepsilon \) and to obtain the problem
\[
\begin{align*}
\frac{dx}{dt} &= \frac{1}{\varepsilon}(1 - y^2)x, \\
\frac{dx}{dt} &= -\frac{\alpha}{\varepsilon^2}y + \sqrt{\frac{2\alpha}{\varepsilon^2}} \frac{dv}{dt}.
\end{align*}
\]
In this time rescaling nontrivial dynamics occur. SDE of this form are the topic of Chapter 11, and this specific example is considered in section 11.6.

### 10.5.2 Hamiltonian Mechanics

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,
\[
H(x, p, y, v) = \frac{1}{2}(p^2 + v^2) + V(x) + \frac{\omega(x)\varepsilon^2}{2}y^2,
\]
where \((x, y)\) are the coordinates and \((p, v)\) are the conjugate momenta of the two particles, \(V(x)\) is a non-negative potential and \(\omega(x)\) is assumed to satisfy \(\omega(x) \geq \tilde{\omega} > 0\) for all \(x\). The corresponding equations of motion are
\[
\begin{align*}
\frac{dx}{dt} &= p, \\
\frac{dp}{dt} &= -V'(x) - \frac{\omega'(x)}{2\varepsilon^2}y^2, \\
\frac{dy}{dt} &= v, \\
\frac{dv}{dt} &= -\frac{\omega(x)}{\varepsilon^2}y.
\end{align*}
\]
We let \(E\) denote the value of the energy \(H\) at time \(t = 0\). We assume that the energy \(E\) is bounded independently of \(\varepsilon\) and this implies that \(y^2 \leq 2\varepsilon^2 E/\tilde{\omega}\). 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

\[
\begin{align*}
\frac{dx}{dt} &= p \\
\frac{dp}{dt} &= -V'(x) - \frac{\omega'(x)}{2} \eta^2 \\
\frac{d\eta}{dt} &= \frac{1}{\varepsilon} v \\
\frac{dv}{dt} &= -\frac{\omega(x)}{\varepsilon} \eta.
\end{align*}
\]

(10.5.2)

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

\[
\begin{align*}
\frac{d^2 x}{dt^2} + V'(x) + \frac{1}{2} \omega'(x) \eta^2 &= 0 \\
\frac{d^2 \eta}{dt^2} + \frac{1}{\varepsilon^2} \omega(x) \eta &= 0.
\end{align*}
\]

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

Consider the fast dynamics, with \( (x, p) \) frozen. The total energy for this fast dynamics is

\[
\frac{1}{2} v^2 + \frac{\omega(x)}{2} \eta^2.
\]

The initial energy of the fast system, which is conserved, is found by subtracting the energy associated with the frozen variables from the total energy of the original system, and we denote it by

\[
E_{\text{fast}} = E - \frac{1}{2} p^2 - V(x).
\]

For fixed \( x, p \) the dynamics in \( \eta, v \) is confined to this energy shell, which we denote by \( \mathcal{Y}(x, p) \).

Harmonic oscillators satisfy an equi-partition property whereby, on average, the energy is equally distributed between its kinetic and potential contributions.
(the virial theorem). Thus the average of the kinetic energy of the fast oscillator against the ergodic measure \( \mu_{x,p} \) is 
\[
\int_{\gamma(x,p)} \frac{\omega(x)}{2} \eta^2 \mu_{x,p}(d\eta, dv) = \frac{1}{2} \left( E - \frac{1}{2} p^2 - V(x) \right).
\]
Thus 
\[
\int_{\gamma(x,p)} \frac{1}{2} \eta^2 \mu_{x,p}(d\eta, dv) = \frac{1}{2} \omega(x) \left( E - \frac{1}{2} p^2 - V(x) \right).
\]
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 for \( p \) can be approximated by its time-average, giving rise to a closed system of equations for \((X, P) \approx (x, p)\). These are
\[
\frac{dX}{dt} = P,
\frac{dP}{dt} = -V'(X) - \frac{\omega'(X)}{2\omega(X)} \left( E - \frac{1}{2} p^2 - V(X) \right),
\]
with initial data \( E, X(0) = X_0 \) and \( P(0) = P_0 \). It is verified below that \((X, P)\) satisfying (10.5.3) conserve the following adiabatic invariant
\[
J = \frac{1}{\omega^{1/2}(X)} \left( E - \frac{1}{2} p^2 - V(X) \right).
\]
Thus, (10.5.3) reduces to the simpler form
\[
\frac{dX}{dt} = P,
\frac{dP}{dt} = -V'(X) - J_0 \left[ \omega^{1/2}(X) \right]',
\]
where \( J_0 \) is given by
\[
J_0 = \frac{1}{\omega^{1/2}(X_0)} \left( E - \frac{1}{2} p^2_0 - V(X_0) \right).
\]
This means that the influence of the stiff potential on the slow variables is to replace the potential \( V(x) \) by an effective potential,
\[
V_{\text{eff}}(x) = V(x) + J_0 \omega^{1/2}(x).
\]
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 10.2.

To verify that $J$ is indeed conserved in time, note that, from the definition of $J$,

$$\frac{d}{dt} \left( \omega^\frac{1}{2} (X) J \right) = \frac{d}{dt} \left( E - \frac{1}{2} P^2 - V(X) \right)$$

$$= \frac{P \omega'(X)}{2 \omega(X)} \left( E - \frac{1}{2} P^2 - V(X) \right)$$

$$= \frac{P \omega'(X)}{2 \omega^\frac{1}{2} (X)} J.$$

But, from the equations of motion for $(X, P)$,

$$\frac{d}{dt} \left( \omega^\frac{1}{2} (X) J \right) = \frac{1}{2} \frac{\omega'(X)}{\omega^\frac{1}{2} (X)} \frac{dX}{dt} J + \omega^\frac{1}{2} (X) \frac{dJ}{dt}$$

$$= \frac{P \omega'(X)}{2 \omega^\frac{1}{2} (X)} J + \omega^\frac{1}{2} (X) \frac{dJ}{dt}.$$

Equating the two expressions gives

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

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

### 10.6 Discussion and Bibliography

A detailed account of the averaging method, as well as numerous examples, can be found in Sanders and Verhulst [132]. See also [7]. An English language review of the Russian literature can be found in Lochak and Meunier [90]. An overview of the topic of slow manifolds, especially in the context of Hamiltonian problems, may be found in [92]. The averaging method applied to equations (10.2.1) is analyzed in an instructive manner in [115], where the Liouville equation is used to construct a rigorous proof of the averaged limit. The paper [150] provides an overview of variable elimination in a wealth of problems with scale separation.

Anosov’s theorem is the name given to the averaging principle in the context of ODE – (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 [90] 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 [78, 79, 80, 81].

The situations in which the fast dynamics tend to fixed points, periodic solutions, or chaotic solutions can be treated in a unified manner through the introduction of Young measures. Artstein and co-workers considered a class of singularly perturbed system of type (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

$$\frac{dX}{dt} = \int \{X, y\} \mu_X(dy),$$

and $\mu_X$ is the ergodic measure on $\mathcal{Y}$; 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 concentrated at a point.) A general theorem along these lines is proved in [9].

There are many generalizations of this idea. The case of non-autonomous fast dynamics, as well as a case with infinite dimension are covered in [10]. 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 (non-deterministic) differential inclusion [11].

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

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

$$\frac{dy}{dt} = \frac{1}{\varepsilon} g(x, y) + \frac{1}{\sqrt{\varepsilon}} \beta(x, y) \frac{dV}{dt}. \tag{10.6.1}$$

The simplified equation is then an SDE, not an ODE. This situation is a sub-case of the set-up we consider in the next chapter which can be obtained by setting $f_0 = 0$ in that chapter, letting $f_1 = f$ 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 time-scales was initiated by Rubin and Ungar [131]. More recently the ideas of Neistadt, based on normal form theory, have been applied to such problems [16];
this approach is very powerful, yielding very tight, exponential, error estimates between the original and limiting variables. A recent approach to the problem, using the techniques of time-homogenization [21], is the paper [22]. The example we present is taken from that paper. The heuristic derivation we have given here is made rigorous in [22], using time-homogenization techniques, and it is also generalized to higher dimension. Resonances become increasingly important as the co-dimension, \( m \), increases, limiting the applicability of the averaging approach to such two-scale Hamiltonian systems (Takens [142]).

### 10.7 Exercises

1. Find the averaged equation resulting from the SDE (10.6.1).

2. Consider the equations

\[
\begin{align*}
\frac{dx}{dt} &= -\nabla_x \Phi(x, y) + \sqrt{2\sigma} \frac{dU}{dt} \\
\frac{dy}{dt} &= -\frac{1}{\varepsilon} \nabla_y \Phi(x, y) + \sqrt{\frac{2\sigma}{\varepsilon}} \frac{dV}{dt}.
\end{align*}
\]

Show that the averaged equation for \( X \) has the form

\[
\frac{dX}{dt} = -\nabla \Psi(X) + \sqrt{2\sigma} \frac{dU}{dt}
\]

where the **Fixman potential** \( \Psi \) is given by

\[
\exp\left(-\frac{1}{\sigma} \Psi(x)\right) = \int_Y \exp\left(-\frac{1}{\sigma} \Phi(x, y)\right) dy.
\]

3. Let \( y \) be a two state continuous time Markov chain with generator

\[
L = \frac{1}{\varepsilon} \begin{pmatrix} -a & a \\ b & -b \end{pmatrix}.
\]

Without loss of generality label the state-space \( I = \{-1, +1\} \). Consider the SDE

\[
\frac{dx}{dt} = f(x, y) + \alpha(x, y) \frac{dU}{dt}.
\]

Write the generator of the process and use multiscale analysis to derive the averaged SDE of the form

\[
\frac{dX}{dt} = F(X) + A(X) \frac{dU}{dt}.
\]
4. Let $z$ 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 $z$, given by

$$\frac{dx}{dt} = f(x, y) + \sqrt{2\sigma} \frac{dU}{dt}$$

$$\frac{dy}{dt} = -\frac{1}{\varepsilon} \omega(z)(y - m(z)) + \sqrt{2\sigma \varepsilon} \frac{dV}{dt}.$$ 

Write the generator for this process and use multiscale analysis to derive the averaged coupled Markov chain and SDE of the form

$$\frac{dX}{dt} = F(X, z) + \sqrt{2\sigma} \frac{dU}{dt}.$$ 

5. Generalize the previous exercise to the case where the transition rates of the Markov chain, determined by $a$ and $b$, depend upon $x$ and $y$. 
Chapter 11

Homogenization for ODE and SDE

11.1 Introduction

In this chapter we continue our study of systems of SDE 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 is the longer, \textit{diffusive} time scale $O(1/\varepsilon^2)$. This is the time scale of interest when the effective drift defined in equation (10.3.1) vanishes, due, for example, to the symmetries of the problem. This is the centering condition, equation (11.2.2) below. In contrast to the case considered in the previous chapter, in the diffusive time scale the effective equation is stochastic.

11.2 Full Equations

Consider the stochastic differential equations

\[
\begin{align*}
\frac{dx}{dt} &= \frac{1}{\varepsilon} f_0(x, y) + f_1(x, y) + \alpha(x, y) \frac{dU}{dt} \\
\frac{dy}{dt} &= \frac{1}{\varepsilon^2} g(x, y) + \frac{1}{\varepsilon} \beta(x, y) \frac{dV}{dt}
\end{align*}
\]  

(11.2.1)

Both the $x$ and $y$ equations contain fast dynamics, but the dynamics in $y$ is an order of magnitude faster than in $x$. Here $x \in \mathcal{X}$, $y \in \mathcal{Y}$, $z \in \mathcal{Z}$ as discussed in sections 4.1 and 6.1.
As in the previous chapter, we assume that the dynamics in $y$, with $x$ viewed as frozen, is ergodic. Specifically we assume that the frozen dynamics (10.2.2) generates an ergodic measure $\mu_x(\cdot)$ defined by (4.4.1). In addition we also assume that $f_0(x, y)$ averages to zero under this measure, so that

$$\int f_0(x, y) \mu_x(dy) = 0. \quad (11.2.2)$$

It can then be shown that the term $f_0$ will, in general, give rise to $O(1)$ effective drift and noise contributions in an approximate equation for $x$.

For equation (11.2.1) the backward Kolmogorov equation (6.3.3) with $\phi = \phi(x)$ is,

$$\begin{align*}
\frac{\partial v}{\partial t} &= \frac{1}{\varepsilon^2} L_0 v + \frac{1}{\varepsilon} L_1 v + L_2 v \\
v(x, y, 0) &= \phi(x), \quad (11.2.3)
\end{align*}$$

where

$$\begin{align*}
L_0 v &= g \cdot \nabla_y v + \frac{1}{2} B : \nabla_y \nabla_y v \\
L_1 v &= f_0 \cdot \nabla_x v \\
L_2 v &= f_1 \cdot \nabla_x v + \frac{1}{2} A : \nabla_x \nabla_x v
\end{align*} \quad (11.2.4a-c)$$

with

$$\begin{align*}
A(x, y) &= \alpha(x, y) \alpha(x, y)^T, \\
B(x, y) &= \beta(x, y) \beta(x, y)^T.
\end{align*}$$

By using perturbation theory we eliminate the $y$ dependence in this equation, giving rise to simplified equations for $x$.

In terms of the generator $L_0$, our ergodicity assumption becomes the statement that $L_0$ has one dimensional null-space characterized by

$$\begin{align*}
L_0 1(y) &= 0, \\
L_0^* \rho_\infty(y; x) &= 0. \quad (11.2.5)
\end{align*}$$

Here $1(y)$ denotes constants in $y$.

As in the previous chapter, in the case where $\mathcal{Y} = \mathbb{T}^d$ the operators $L_0$ and $L_0^*$ are equipped with periodic boundary conditions. Then in the case where $B(x, y)$
is strictly positive-definite, uniformly in \((x, y) \in X \times Y\) see Theorem 6.10 justifies the statement that the null-space of \(L_0^*\) is one-dimensional. In more general situations, such as when \(Y = \mathbb{R}^d\), 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.

### 11.3 Simplified equations

We define the cell problem as follows:

\[
L_0 \Phi(x, y) = -f_0(x, y), \quad \int \Phi(x, y) \rho_\infty(y; x) \, dy = 0.
\] (11.3.1)

Because \(f_0\) satisfies (11.2.2), this equation has a unique solution by the Fredholm alternative. We may then define a vector field \(F\) by

\[
F(x) = \int \left( f_1(x, y) + \nabla_x \Phi(x, y) f_0(x, y) \right) \rho_\infty(y; x) \, dy
\]

\[
= F_1(x) + F_0(x)
\] (11.3.2)

and a diffusion matrix \(A(x)\) by

\[
A(x) A(x)^T = A_1(x) + \frac{1}{2} \left( A_0(x) + A_0(x)^T \right)
\] (11.3.3)

where

\[
A_0(x) := \int 2f_0(x, y) \odot \Phi(x, y) \rho_\infty(y; x) \, dy,
\] (11.3.4)

\[
A_1(x) := \int A(x, y) \rho_\infty(y; x) \, dy.
\] (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 \(A_0(x)\) is positive semi-definite. This is done in section 11.5.

**Result 11.1.** For \(\varepsilon \ll 1\) and \(t = \mathcal{O}(1)\) solving (11.2.1) is approximated by \(X\) solving

\[
\frac{dX}{dt} = F(X) + A(X) \frac{dW}{dt}.
\] (11.3.6)

Alternative representations of \(F(x)\) and \(A(x)A(x)^T\) may be found by using time averaging. The first of such results is as follows. Note that the expressions
CHAPTER 11. HOMOGENIZATION FOR ODE AND SDE

for $A_0$ and $F_0$ 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 $t$ enables us to express the effective equations without reference to solution of the cell problem $\Phi$, and requires a decay of correlations in order to be well-defined.

**Result 11.2.** Alternative representations of the vector field $F$ and diffusion matrix $A$ can be found through the following integrals over time:

\[ F_1(x) = \lim_{T \to \infty} \frac{1}{T} \int_0^T f_1(x, \varphi_x^t(y)) dt, \]
\[ A_1(x) = \lim_{T \to \infty} \frac{1}{T} \int_0^T A(x, \varphi_x^t(y)) dt; \]

and

\[ A_0(x) = 2 \int_0^\infty \left( \lim_{T \to \infty} \frac{1}{T} \int_0^T f_0(x, \varphi_x^s(y)) \otimes f_0(x, \varphi_x^{t+s}(y)) ds \right) dt. \] (11.3.7)

If the generator $L_0$ is independent of $x$ then

\[ F_0(x) = \int_0^\infty \left( \lim_{T \to \infty} \frac{1}{T} \int_0^T \nabla_y f_0(x, \varphi_x^{t+s}(y)) f_0(x, \varphi_x^s(y)) ds \right) dt. \]

A second representation of $A_0(x)$ and $F_0(x)$, intermediate between the pure ergodic average and pure time-average, is as follows. We let $E^{\mu_x}$ denote the stationary measure on path space for $\varphi_x^t(\cdot)$. To be precise this is the product measure formed from use of $\mu_x(\cdot)$ on initial data and standard independent Wiener measure on driving Brownian motions. With this notation we have:

**Result 11.3.** 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 $E^{\mu_x}$:

\[ A_0(x) = 2 \int_0^\infty E^{\mu_x} \left( f_0(x, y) \otimes f_0(x, \varphi_x^s(y)) \right) dt \] (11.3.8)

and, if the generator $L_0$ is independent of $x$, then

\[ F_0(x) = \int_0^\infty E^{\mu_x} \left( \nabla_y f_0(x, \varphi_x^s(y)) f_0(x, y) \right) dt. \] (11.3.9)

The following result will be useful to us in deriving the previous two representations of $A_0(x)$ and $F_0(x)$. It uses ergodicity to represent solution of the cell problem, and related Poisson equations, as time-integrals.
Result 11.4. Assume that \( \mathcal{L}_0 \) is the generator of an ergodic process in \( \mathcal{Y} \) with invariant density \( \rho^\infty(y) \). Let \( h \) be a function orthogonal to the null-space of \( \mathcal{L}_0^* \) so that
\[
\int_{\mathcal{Y}} h(y) \rho^\infty(y) dy = 0.
\]
Now define
\[
H(y) = -\int_0^\infty (e^{\mathcal{L}_0 t}) h(y) dt.
\]
Then
\[
\mathcal{L}_0 H = h, \quad \int_{\mathcal{Y}} H(y) \rho^\infty(y) dy = 0.
\]

Proof. We first note that
\[
(\mathcal{L}_0 H)(y) = -\int_0^\infty \mathcal{L}_0 (e^{\mathcal{L}_0 t} h)(y) dt
\]
\[
= -\int_0^\infty \frac{\partial}{\partial t} (e^{\mathcal{L}_0 t} h)(y) dt
\]
\[
= h(y) - \lim_{t \to \infty} (e^{\mathcal{L}_0 t} h)(y)
\]
\[
= h(y).
\]
In the last identity we used ergodicity and the fact that \( h \) averages to zero under the density \( \rho^\infty \). This follows from Theorem 6.10 because \( v(y, t) = (e^{\mathcal{L}_0 t} h)(y) \) solves the backward Kolmogorov equation.
Note that also \( \int H \rho^\infty dy = 0 \) since
\[
\int_{\mathcal{Y}} H(y) \rho^\infty(y) dy = -\int_0^\infty \left( \int_{\mathcal{Y}} (e^{\mathcal{L}_0 t} h)(y) \rho^\infty(y) dy \right) dt
\]
\[
= -\int_0^\infty \left( \int_{\mathcal{Y}} h(y) e^{\mathcal{L}_0^* t} \rho^\infty(y) dy \right) dt
\]
\[
= 0.
\]
The last line follows because \( \rho^\infty(y) \) is stationary for the Fokker-Planck dynamics. \( \square \)
11.4 Derivation

We seek a multiscale expansion for the solution with the form

$$v = v_0 + \varepsilon v_1 + \varepsilon^2 v_2 + \cdots$$  \hspace{1cm} (11.4.1)

Note that $L_0$ is a differential operator in $y$, in which $x$ appears as a parameter. Thus we must equip it with boundary conditions. For simplicity we consider the case where $\mathcal{Y} = \mathbb{T}^d$ so that periodic boundary conditions are used. 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 $L_0$. Substituting this expansion into (11.2.3) and equating powers of $\varepsilon$ gives a hierarchy of equations, the first three of which are

$$O(1/\varepsilon^2) \quad L_0 v_0 = 0, \text{ } v_0 \text{ is 1–periodic,}$$  \hspace{1cm} (11.4.2a)

$$O(1/\varepsilon) \quad L_0 v_1 = -L_1 v_0 \cdot v_1 \text{ is 1–periodic},$$  \hspace{1cm} (11.4.2b)

$$O(1) \quad L_0 v_2 = \frac{\partial v_0}{\partial t} - L_1 v_1 - L_2 v_0 \cdot v_2 \text{ is 1–periodic.}$$  \hspace{1cm} (11.4.2c)

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

Recall that we assume that $L_0$ generates an ergodic process and hence that $L_0$ and $L_0^*$ have a one-dimensional null space. These are characterized in (11.2.5). Hence equation (11.4.2a) implies that $v_0(x, t)$ only. The solution of the next two equations requires the Fredholm alternative. In both cases the right hand side must be orthogonal to $\text{Null}\{L_0^\ast\}$. By ergodicity we know that

$$\text{Null}\{L_0^\ast\} = \text{span}\{\rho_\infty(y; x)\}.$$  \hspace{1cm} (11.4.2d)

Equation (11.4.2b) is solvable for $v_1$, by virtue of (11.2.2) which ensures that the operator $L_1$, given by (11.2.4c) averages to zero. From the structure of $L_1$, we have that (11.4.2b) may be written

$$L_0 v_1 = -f_0(x, y) \cdot \nabla_x v_0(x, t)$$  \hspace{1cm} (11.4.3)

Since $L_0$ is a differential operator in $y$ alone with $x$ appearing as a parameter, we may separate variables in (11.4.3) and write $v_1(x, y, t) = \Phi(x, y) \cdot \nabla_x v_0(x, t)$. 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

$$L_0 v_0 = 0, \text{ } v_0 \text{ is 1–periodic,}$$  \hspace{1cm} (11.4.4a)

$$L_0 v_1 = -L_1 v_0 \cdot v_1 \text{ is 1–periodic,}$$  \hspace{1cm} (11.4.4b)

$$L_0 v_2 = \frac{\partial v_0}{\partial t} - L_1 v_1 - L_2 v_0 \cdot v_2 \text{ is 1–periodic.}$$  \hspace{1cm} (11.4.4c)
(11.3.1). Condition (11.2.2) 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 have that the right hand side takes the form
\[ \frac{\partial v_0}{\partial t} - \mathcal{L}_2 v_0 - \mathcal{L}_1 \Phi \cdot \nabla_x v_0. \]
Hence solvability of (11.4.2c) requires
\[ \frac{\partial v_0}{\partial t} = \int \rho^\infty(y; x) \mathcal{L}_2 v_0(x, t) dy + \int \rho^\infty(y; x) \mathcal{L}_1 \left( \Phi(x, y) \cdot \nabla_x v_0(x, t) \right) dy \]
\[ = I_1 + I_2. \quad (11.4.4) \]
We consider the two terms on the right hand side separately. The first is
\[ I_1 = \int \rho^\infty(y; x) \left( f_1(x, y) \cdot \nabla_x + \frac{1}{2} A(x, y) : \nabla_x \nabla_x \right) v_0(x, t) dy \]
\[ = F_1(x) \cdot \nabla_x v_0(x, t) + \frac{1}{2} A_1(x) : \nabla_x \nabla_x v_0(x, t) \]
Now for the second term note that
\[ \mathcal{L}_1 \left( \Phi \cdot \nabla_x v_0 \right) = f_0 \otimes \Phi : \nabla_x \nabla_x v_0 + (\nabla_x \Phi f_0) \cdot \nabla_x v_0. \]
Hence \( I_2 = I_3 + I_4 \) where
\[ I_3 = \int \rho^\infty(y; x) \left( \nabla_x \Phi(x, y) f_0(x, y) \right) \cdot \nabla_x v_0(x, t) dy \]
and
\[ I_4 = \int \rho^\infty(y; x) \left( f_0(x, y) \otimes \Phi(x, y) : \nabla_x \nabla_x v_0(x, t) \right) dy. \]
Thus
\[ I_2 = F_0(x) \cdot \nabla_x v(x, t) + \frac{1}{2} A_0(x) : \nabla_x \nabla_x v_0(x, t). \]
Combining our simplifications of the right hand side of (11.4.4) we obtain, since only the symmetric part of \( A_0 \) is required to calculate the Frobenius inner-product with another symmetric matrix, the following expression:
\[ \frac{\partial v_0}{\partial t} (x, t) = F(x) \cdot \nabla_x v_0(x, t) + \frac{1}{2} A(x) A(x)^T : \nabla_x \nabla_x v_0(x, t). \]
This is the backward equation corresponding to the reduced dynamics given in (11.3.6).
We finish this section by deriving the alternative expressions for \( A(x), F(x) \) through time integration, given in Results 11.2 and 11.3. The expressions for \( F_1(x) \) and \( A_1(x) \) in Result 11.2 are immediate from ergodicity, simply using the fact that the time average equals the average against \( \rho^\infty \). By use of Result 11.4, the solution to the cell problem can be written as

\[
\Phi(x, y) = \int_0^\infty (\epsilon \mathcal{L}_0 t) f_0(x, y) dt = \int_0^\infty \mathbb{E} f_0(x, \varphi^t_x(y)) dt \tag{11.4.5}
\]

where \( \mathbb{E} \) denotes expectation with respect to Wiener measure. Now

\[
F_0(x) = \int_Y \rho^\infty(y; x) \nabla_x \Phi(x, y) f_0(x, y) dy.
\]

In the case where \( \mathcal{L}_0 \) is \( x \)-independent so that \( \varphi^t_x(\cdot) = \varphi^t(\cdot) \) is also \( x \)-independent, we may use (11.4.5) to see that

\[
F_0(x) = \int_Y \rho^\infty(y; x) \int_0^\infty \mathbb{E} \nabla_x f_0(x, \varphi^t_x(y)) f_0(x, y) dt.
\]

Recalling that \( \mathbb{E}^{\mu_x} \) denotes the stationary measure on path space for \( \varphi^t_x(\cdot) \) and changing the order of integration we find that

\[
F_0(x) = \int_0^\infty \mathbb{E}^{\mu_x} \left( \nabla_x f_0(x, \varphi^t_x(y)) f_0(x, y) \right) dt \tag{11.4.6}
\]

as required for the expression in Result 11.3. Now we replace averages over \( \mathbb{E}^{\mu_x} \) by time averaging to obtain

\[
F_0(x) = \int_0^\infty \left( \lim_{T \to \infty} \frac{1}{T} \int_0^T \nabla_x f_0(x, \varphi^{t+s}_x(y)) f_0(x, \varphi^s_x(y)) ds \right) dt,
\]

and so we obtain the desired formula for Result 11.2.

A similar calculation to that yielding (11.4.6) gives (11.3.8) for \( A_0(x) \) in Result 11.3. Replacing average against \( \mathbb{E}^{\mu_x} \) by time-average we arrive at the desired formula for \( A_0(x) \) in Result 11.2.

**11.5 Properties of the Effective Equation**

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 non-negative definite. We now prove that this is indeed the case.
Theorem 11.5. Consider the case where $\mathcal{Y} = \mathbb{T}^d$ and $\mathcal{L}_0$ is equipped with periodic boundary conditions. The real-valued matrix $A$ is well-defined by (11.3.3):

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

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

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

By Theorem 6.7 we have

$$\langle \xi, A_1(x)\xi + A_0(x)\xi \rangle = \int_\mathcal{Y} \left( |\alpha(x, y)^T \xi|^2 - 2(L_0 \phi(x, y)) \phi(x, y) \right) \rho^\infty(y; x) dy$$

$$= \int_\mathcal{Y} \left( |\alpha(x, y)^T \xi|^2 + |\beta(x, y)^T \nabla_y \phi(x, y)|^2 \right) \rho^\infty(y; x) dy$$

$$\geq 0. \quad \square$$

Two important remarks are in order.

Remark 11.6. Techniques similar to those used in the proof of the previous theorem, using (6.3.8) instead of the Dirichlet form itself, show that

$$\frac{1}{2} \left( A_0(x) + A_0(x)^T \right) = \int_\mathcal{Y} \left( \nabla \Phi(x, y) \beta(x, y) \otimes \nabla \Phi(x, y) \beta(x, y) \right) \rho^\infty(y; x) dy.$$  \hspace{1cm} (11.5.1)

Remark 11.7. By virtue of Remark 6.8 we see that the proceeding theorem can be extended to settings other than $\mathcal{Y} = \mathbb{T}^d$.

11.6 Applications

11.6.1 Fast Ornstein-Uhlenbeck Noise

Consider the equations

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

$$\frac{dy}{dt} = -\frac{\alpha}{\varepsilon^2}y + \sqrt{\frac{2\alpha}{\varepsilon^2}} dV.$$
Recall that these equations arise from the first application in section 10.5, in the case where $\lambda = 2\alpha$, and after time-rescaling to produce non-zero effects. The function $V(t)$ is a standard unit Brownian motion.

We want to find the fluctuations induced by the fast process $y$ on $x$. To this end we need to study the variable $\varphi^t(y)$ solving (10.2.2), noting that here there is no dependence on $x$ in this fast process – it is simply on OU process. A straightforward calculation shows that

$$\varphi^t(y) = e^{-\alpha t}y + \sqrt{2\alpha} \int_0^t e^{-\alpha(t-s)} dV(s),$$

$$\varphi^t(y)^2 = e^{-2\alpha t}y^2 + \sqrt{2\alpha ye^{-\alpha t}} \int_0^t e^{-\alpha(t-s)} dV(s) + 2\alpha \left( \int_0^t e^{-\alpha(t-s)} dV(s) \right)^2.$$  

(11.6.1)

In this example

$$\mathcal{L}_0 = -\alpha y \frac{\partial}{\partial y} + \alpha \frac{\partial^2}{\partial y^2}. \quad (11.6.2)$$

Because this operator is $x-$independent, the invariant density in the null-space of $\mathcal{L}^*$ is $\rho^\infty(y)$, and is independent of $x$. In fact it is the density corresponding to an $\mathcal{N}(0, 1)$ random variable. By the Itô isometry,

$$\mathbb{E}\left( \int_0^t e^{-\alpha(t-s)} dV(s) \right)^2 = \int_0^t e^{-2\alpha(t-s)} ds,$$

$$= \frac{1}{2\alpha} \left( 1 - e^{-2\alpha t} \right).$$

To construct the measure $\mathbb{E}^{\mu^x}$ we take $y$ and $V$ to be independent. Thus

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

Furthermore

$$\mathbb{E} \int \rho^\infty(y)y^2 \varphi^t(y)^2 = e^{-2\alpha t} \int \rho^\infty(y)y^4 dy + 2\alpha \mathbb{E} \left( \int_0^t e^{-\alpha(t-s)} dV(s) \right)^2$$

$$= 3e^{-2\alpha t} + 1 - e^{-2\alpha t}$$

$$= 1 + 2e^{-2\alpha t}.$$
11.6. APPLICATIONS

Here \( f_0(x, y) = (1 - y^2)x \). Combining these calculations in (11.3.9) gives

\[
F_0(x) = x \int_0^\infty \mathbb{E}^{\mu_t} \left( (1 - \varphi^t(y)^2)(1 - y^2) \right) dt = x \int_0^\infty 2e^{-2\alpha t} dt = \frac{x}{\alpha}. \tag{11.6.3}
\]

Similarly from (11.3.8) we obtain

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

It follows that the effective equation is

\[
\frac{dX}{dt} = \frac{X}{\alpha} + \sqrt{\frac{2}{\alpha}} X \frac{dW}{dt}.
\]

11.6.2 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 11.2 since it uses time-integrals, and makes no reference to averaging over Brownian motion (which is not present in this example). Consider the equations

\[
\begin{align*}
\frac{dx}{dt} &= x - x^3 + \frac{\lambda}{\varepsilon} y_2 \\
\frac{dy_1}{dt} &= \frac{10}{\varepsilon^2} (y_2 - y_1) \\
\frac{dy_2}{dt} &= \frac{1}{\varepsilon^2} (28y_1 - y_2 - y_1y_3) \\
\frac{dy_3}{dt} &= \frac{1}{\varepsilon^2} (y_1y_2 - \frac{8}{3} y_3)
\end{align*}
\]

(11.6.4)

The vector \( y = (y_1, y_2, y_3)^T \) solves the Lorenz equations, at parameter values where the solution is chaotic [154]. Thus the equation for \( x \) is a scalar ODE driven by a chaotic signal with characteristic time \( \varepsilon^2 \). Because \( f_0(x, y) \propto y_2 \) and because \( f_1(x) \) only, the candidate equation for the approximate dynamics is

\[
\frac{dX}{dt} = X - X^3 + \sigma \frac{dW}{dt}, \tag{11.6.5}
\]
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 (11.3.7) giving

$$\sigma^2 = 2\lambda^2 \int_0^\infty \left( \lim_{T \to \infty} \int_0^T \psi_s(y) \psi_{t+s}(y) \, ds \right) \, dt.$$ 

This is the integrated autocorrelation function of $y_2$.

Another way to derive this result is as follows. Gaussian white noise, the time-derivative of Brownian motion, may be thought of as a delta-correlated stationary process. On the assumption that $y_2$ has a correlation function which decays in time, and noting that this has time-scale $\varepsilon^2$, the autocorrelation of $\frac{1}{\varepsilon}y_2(s)$ 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 $\sigma^2$.

### 11.6.3 Stratonovich Corrections

When white noise is approximated by a smooth process typically this leads to Stratonovich interpretations of stochastic integrals. We use multiscale analysis to illustrate this idea by means of a simple example.

Consider the equations

$$\frac{dx}{dt} = \frac{1}{\varepsilon} v(x)y + \sigma \frac{dU}{dt},$$

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

with $U, V$ standard Brownian motions. Heuristically we deduce from the $y$ equation that

$$\frac{y}{\varepsilon} \approx \sqrt{\frac{2}{\alpha}} \frac{dV}{dt} + \mathcal{O}(\varepsilon).$$

(11.6.6)

Thus we might conjecture a limiting equation of the form

$$\frac{dX}{dt} = \sqrt{\frac{2}{\alpha} v(X)} \frac{dV}{dt} + \sigma \frac{dU}{dt}$$

(11.6.7)

which is equivalent in distribution to the equation

$$\frac{dX}{dt} = \left( \frac{2}{\alpha} v(X)^2 + \sigma^2 \right)^{\frac{1}{2}} \frac{dW}{dt},$$

with $W$ is a standard unit Brownian motion.
11.6. APPLICATIONS

We will show that this heuristic is incorrect: this is because, whenever white noise is approximated by a smooth process, the limiting equation should be interpreted in the Stratonovich sense

\[
\frac{dX}{dt} = \sqrt{\frac{2}{\alpha}} v(X) \circ \frac{dV}{dt} + \sigma \frac{dU}{dt}.
\]

Thus, in Itô form the limiting equation is

\[
\frac{dX}{dt} = \left(\frac{2}{\alpha} v(X)^2 + \sigma^2\right)^{1/2} \frac{dW}{dt} + \frac{1}{\alpha} v'(X)v(X).
\] (11.6.8)

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

The cell problem is

\[
\mathcal{L}_0 \Phi(x, y) = -v(x) y
\]

with \(\mathcal{L}_0\) given by (11.6.2). The solution is readily seen to be

\[
\Phi(x, y) = \frac{1}{\alpha} v(x)y, \quad \nabla_x \Phi(x, y) = \frac{1}{\alpha} v'(x)y.
\]

The invariant density is

\[
\rho^\infty(y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right)
\]

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

\[
F(x) = \int_\mathbb{R} \frac{1}{\alpha} v'(x)v(x)y^2 \rho^\infty(y) dy
= \frac{1}{\alpha} v'(x)v(x).
\]

Also (11.3.3) gives

\[
A(x)^2 = \int_\mathbb{R} \left(\sigma^2 + \frac{2}{\alpha} v(x)^2 y^2\right) \rho^\infty(y) dy
= \sigma^2 + \frac{2}{\alpha} v(x)^2.
\]

Hence the desired result is established.
CHAPTER 11. HOMOGENIZATION FOR ODE AND SDE

11.6.4 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, and subject to additional molecular bombardment. If the particle has unit mass \( m \) then it is natural to study the equation

\[
\begin{align*}
md\frac{d^2 x}{dt^2} &= \frac{1}{\varepsilon} f(x)y - \frac{dx}{dt} + \sigma \frac{dU}{dt} \\
\frac{dy}{dt} &= -\frac{\alpha y}{\varepsilon^2} + \sqrt{\frac{2\alpha}{\varepsilon^2}} \frac{dV}{dt}.
\end{align*}
\]  

(11.6.9)

(11.6.10)

The first of the two previous equations is Stokes’ law, stating that the force on the particle is the sum of a drag force, \( \frac{1}{\varepsilon} f(x)y - \frac{dx}{dt} \), and a force due to molecular bombardment, \( \sigma \frac{dU}{dt} \). As in the previous example, \( y \) is a fluctuating OU process. For simplicity we consider the case of unit mass \( m \).

Using the heuristic (11.6.6) it is natural to conjecture the limiting equation

\[
\frac{d^2 X}{dt^2} = \sqrt{\frac{2}{\alpha}} f(X) \frac{dV}{dt} - \frac{dX}{dt} + \sigma \frac{dU}{dt}.
\]

This is equivalent to

\[
\frac{d^2 X}{dt^2} + \frac{dX}{dt} + \left( \sigma^2 + \frac{2}{\alpha} f(X)^2 \right)^{\frac{1}{2}} \frac{dW}{dt}
\]

in distribution. 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 corrections coincide; there is no Itô correction to the Stratonovich integral. We verify the result by using the multiscale techniques introduced in this chapter.

We first write (11.6.9) as the first order system

\[
\begin{align*}
\frac{dx}{dt} &= r, \\
\frac{dr}{dt} &= -r + \frac{1}{\varepsilon} f(x)y + \sigma \frac{dU}{dt}, \\
\frac{dy}{dt} &= -\frac{1}{\varepsilon^2} \alpha y + \frac{\sqrt{2\alpha}}{\varepsilon} \frac{dV}{dt}.
\end{align*}
\]

Here \((x, r)\) are slow variables \((x \text{ in (11.2.1)) and } r \text{ in (11.2.1))}\) and \(y\) the fast variables \((y \text{ in (11.2.1))}\).

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 $L_0$ given by (11.6.2). The solution is

$$\Phi(x, r, y) = \left( \frac{1}{\alpha} f(x) y \right), \quad \nabla_{(x, r)} \Phi(x, y) = \left( \frac{1}{\alpha} f'(x) y \right).$$

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

$$F(X, R) = \left( \begin{array}{c} R \\ -R \end{array} \right). \quad (11.6.11)$$

From (11.3.3) we have

$$A(X, R)A(X, R)^T = \int_R \left( \begin{array}{cc} 0 & 0 \\ 0 & \sigma^2 \end{array} \right) + 2 \left( \begin{array}{cc} 0 & 0 \\ 0 & \frac{1}{\alpha} f(X)^2 \end{array} \right) \rho^\infty(y) dy.$$ 

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

$$A(X, R)A(X, R)^T = \left( \begin{array}{cc} 0 & 0 \\ 0 & \sigma^2 + \frac{1}{\alpha} f(X)^2 \end{array} \right).$$ 

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

$$A(X, R) = \left( \sigma^2 + \frac{1}{\alpha} f(X)^2 \right) \frac{1}{2}.$$ 

Thus from (11.6.11) and (11.6.12) we obtain the limiting equation

$$\frac{dX}{dt} = R$$

$$\frac{dR}{dt} = -R + \left( \sigma^2 + \frac{1}{\alpha} f(X)^2 \right)^\frac{1}{2} \frac{dW}{dt}$$

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

**11.6.5 Kubo Formula**

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

$$\frac{d^2x}{dt^2} = f(x) - \frac{dx}{dt} + \sigma \frac{dU}{dt}. \quad (11.6.12)$$
Here \( U \) is a standard unit Brownian motion. We will study the effective diffusive behavior of the particle \( x \) on large length and time-scales, in the case where \( v \) is periodic with mean zero. We show that the diffusion coefficient is given by the Kubo formula, expressing the diffusion coefficient as the integral of the velocity autocorrelation.

To this end we rescale the equation of motion by setting \( x \rightarrow x/\varepsilon \) and \( t \rightarrow t/\varepsilon^2 \) to obtain
\[
\varepsilon^2 \frac{d^2 x}{dt^2} = \frac{1}{\varepsilon} f \left( \frac{x}{\varepsilon} \right) - \frac{dx}{dt} + \frac{\sigma}{\varepsilon} dU.
\]
Introducing the variables \( y = \varepsilon \frac{dx}{dt} \) and \( z = x/\varepsilon \) we obtain the system
\[
\begin{align*}
\frac{dx}{dt} &= \frac{1}{\varepsilon} y, \\
\frac{dy}{dt} &= -\frac{1}{\varepsilon^2} y + \frac{1}{\varepsilon^2} f(z) + \frac{\sigma}{\varepsilon} dW, \\
\frac{dz}{dt} &= \frac{1}{\varepsilon^2} y.
\end{align*}
\]

The process \((y, z)\) has 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 \( x \), \( \Phi(x, y, z) \), 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 \( \psi^t(y, r) \) to be the component of \( \varphi^t(y, r) \) projected onto the \( y \) coordinate. By Result 11.2 we have that
\[
A_0(x) = 2 \int_0^\infty \left( \lim_{T \to -\infty} \frac{1}{T} \int_0^T \psi^s(y) \psi^{s+t}(y) ds \right) dt.
\]
The expression
\[
C(t) = \lim_{T \to -\infty} \frac{1}{T} \int_0^T y(s)y(s+t) ds
\]
is the autocorrelation of the velocity \( y(\cdot) \) of the particle. Thus the effective equation is
\[
\frac{dX}{dt} = \sqrt{2D} \frac{dW}{dt},
\]
11.6. APPLICATIONS

a Brownian motion with diffusion coefficient

\[ D = \int_0^\infty C(t) \, dt, \]

the integrated velocity autocorrelation.

11.6.6 Neither \textit{Itô} nor Stratonovich

Consider Stokes’ law (11.6.9) for a particle of small mass \( m = \tau_0 \varepsilon^2 \) where \( \tau_0 = \mathcal{O}(1) \).

\[ \tau_0 \varepsilon^2 \frac{dq}{dt} = -\frac{dx}{dt} + \frac{1}{\varepsilon} f(x) \eta, \quad (11.6.13a) \]

\[ \frac{d\eta}{dt} = \frac{1}{\varepsilon^2} g(\eta) + \frac{1}{\varepsilon} \sqrt{2\sigma(\eta)} \frac{dW}{dt}. \quad (11.6.13b) \]

For simplicity we have set \( \sigma = 0 \).

We interpret (11.6.13b) in the \textit{Itô} sense. Notice that \( \eta(t) \) is a more general stochastic process, not necessarily the Ornstein-Uhlenbeck process. We assume that \( g(\eta), \sigma(\eta) \) are such that there exists a unique stationary solution of the Fokker-Planck equation for (11.6.13b), so that \( \eta \) is ergodic.

We write (11.6.13) as a first order system,

\[ \frac{dx}{dt} = \frac{1}{\varepsilon} \sqrt{\tau_0} v, \]

\[ \frac{dv}{dt} = \frac{f(x) \eta}{\varepsilon^2 \sqrt{\tau_0}} - \frac{v}{\tau_0 \varepsilon^2}, \quad (11.6.14) \]

\[ \frac{d\eta}{dt} = -\frac{g(\eta)}{\varepsilon^2} + \frac{\sqrt{2\sigma(\eta)}}{\varepsilon} \frac{dW}{dt}. \]

Equations (11.6.14) are of the form (11.2.1) and, under the assumption that the fast process 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 and the cell problem

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

and

\[ -\mathcal{L}_0 h = \frac{v}{\sqrt{\tau_0}}, \quad (11.6.15) \]

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

$$h(x, v, y) = \left(-\sqrt{\tau_0} v + f(x) \hat{h}(\eta)\right).$$ (11.6.16)

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

$$-\mathcal{L}_\eta \hat{h} = \eta.$$ Here \( L_\eta \) denotes the generator of \( \eta \). We assume that \( \eta(t) \) is an ergodic Markov process with unique invariant measure with density \( \rho_\eta(\eta) \) with respect to Lebesgue; the centering condition which ensures the well posedness of the

$$\int \eta \rho_\eta(\eta) \, d\eta = 0.$$ The homogenized SDE is

$$\frac{dX}{dt} = B(X) + \sqrt{D(X)} \frac{dW}{dt}.\quad (11.6.17)$$

where

$$B(x) := \int \int \left( \frac{v}{\sqrt{\tau_0}} \hat{h}(\eta) \frac{\partial f(x)}{\partial x} \right) \rho(x, v, \eta) \, dv \, d\eta$$

and

$$D(x) := 2 \int \int \left( -v^2 + \frac{v}{\sqrt{\tau_0}} \hat{h}(\eta) f(x) \right) \rho(x, v, \eta) \, dv \, d\eta.$$ Notice that (11.6.17) is neither of the Itô nor of the Stratonovich form.

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

$$\frac{d\eta}{dt} = -\frac{\alpha}{\sqrt{\tau_0}} \eta + \sqrt{2\lambda} \frac{dW}{dt},$$ (11.6.18)

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

$$\frac{dX}{dt} = \frac{\lambda}{\alpha^2(1 + \tau_0 \alpha)} f(X) f'(X) + \sqrt{2\lambda \alpha^2} \frac{dW}{dt}.\quad (11.6.19)$$

This equation can be written in the form

$$X(t) = x_0 \int_0^t \frac{2\lambda}{\alpha^2} f(X) \bar{\sigma} \, dW(t),$$

where the definition of the stochastic integral through Riemann sums depends on the value of \( \tau_0 \). Note that in the limit \( \tau_0 \to \infty \) we recover the Itô stochastic integral, whereas in the limit \( \tau_0 \to 0 \) we recover the Stratonovich stochastic integral.
11.6. APPLICATIONS

11.6.7 The Levy Area Correction

In Section 11.6.3 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 might appear in the limit. This spurious drift is related to the non–convergence of the Levy area (see the discussion in Section 11.7).

Consider the fast–slow system

\[ \dot{x}_1 = \frac{1}{\varepsilon} y_1, \]  
\[ \dot{x}_2 = \frac{1}{\varepsilon} y_2, \]  
\[ \dot{x}_3 = \frac{1}{\varepsilon} (x_1 y_2 - x_2 y_1), \]  
\[ \dot{y}_1 = -\frac{1}{\varepsilon^2} y_1 - \alpha \frac{1}{\varepsilon^2} y_2 + \frac{1}{\varepsilon} \dot{W}_1, \]  
\[ \dot{y}_2 = -\frac{1}{\varepsilon^2} y_2 + \alpha \frac{1}{\varepsilon^2} y_1 + \frac{1}{\varepsilon} \dot{W}_2, \]  

where \( \alpha > 0 \). Multiscale analysis leads to the following homogenized system:

\[ \dot{x}_1 = \frac{1}{1 + \alpha^2} \left( \dot{W}_1 - \alpha \dot{W}_2 \right), \]  
\[ \dot{x}_2 = \frac{1}{1 + \alpha^2} \left( \dot{W}_2 + \alpha \dot{W}_1 \right), \]  
\[ \dot{x}_3 = \frac{1}{1 + \alpha^2} \left( (\alpha x_1 - x_2) \dot{W}_1 + (\alpha x_2 + x_1) \dot{W}_2 \right) + \frac{\alpha}{1 + \alpha^2}. \]

Notice that the additional constant drift that appears in equation (11.6.21c) is not consistent with the Stratonovich interpretation of the stochastic integral. Let us write equations (11.6.20d) and (11.6.20e) in the form

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

where \( y = (y_1, y_2), W = (W_1, W_2), I \) is the identity matrix in \( \mathbb{R}^2 \) and \( J \) is the anti-symmetric (symplectic) matrix

\[ J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \]
It becomes clear that the anti-symmetric part in the equation for the fast process is responsible for the presence of the spurious drift in the homogenized equation. In particular, when \( \alpha = 0 \) the homogenized equation becomes
\[
\begin{align*}
\dot{x}_1 &= W_1, \\
\dot{x}_2 &= W_2, \\
\dot{x}_3 &= -x_2 W_1 + x_1 W_2.
\end{align*}
\]

11.7 Discussion and Bibliography

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 [85]. The perturbation expansion which underlies the approach is clearly exposed in [116]. The use of representations in Result 11.1 is discussed in [116]. The representations in Results 11.2 and 11.3 for the effective drift and diffusion can be used in the design of coarse time-stepping algorithms – see [151] and [37].

Studying the derivation of effective stochastic models when the original system is an ODE, as we did in fast chaotic noise application above, is a subject investigated in some generality in [117]. The idea outlined in that example is carried out in discrete time by Beck [14] who also uses a skew-product structure to enable the analysis; the ideas can then be rigorously justified in some cases. In the paper [93] 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 [97].

Related work can be found in [59] and similar ideas in continuous time are addressed in [74, 73] 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 [29], and then make power series expansions in \( \varepsilon \) of the resulting problem.

This situation, and more general related ones, is covered in a series of papers by Papanicolaou and co-workers—see [119, 116, 117, 115], building on original work of Khasminskii [64, 65]. See also [74, 73, 14, 96, 119, 117, 115, 74, 73] for further material.

A rigorous explanation of the example presented in Section 11.6.7 based on the theory of rough paths can be found in [91].
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 14 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.

Applications to climate models, where the atmosphere evolves quickly relative to the slow oceanic variations, are surveyed in Majda et al. [96]; we have followed the presentation in [116, 96] quite closely here. Further applications to the atmospheric sciences may be found in [98, 99]. See also [136].

### 11.8 Exercises

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

\[
\frac{dy}{dt} = g(y) + \frac{dW}{dt}?
\]

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

b. Find the homogenized SDE arising from the system

\[
\begin{align*}
\frac{dx}{dt} &= \frac{1}{\varepsilon} f(x, y) \\
\frac{dy}{dt} &= \frac{1}{\varepsilon} g(y) + \frac{1}{\varepsilon} \frac{dW}{dt},
\end{align*}
\]

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

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.

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

\[
\frac{dy}{dt} = g(y) + \frac{dW}{dt}.
\]

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
\[
\frac{dx}{dt} = f(x, y) \\
\frac{dy}{dt} = \frac{1}{\varepsilon} g(y) + \frac{1}{\sqrt{\varepsilon}} \frac{dW}{dt},
\]
in the case where \( g \) is divergence-free.

c. Find the homogenized SDE arising from the system
\[
\frac{dx}{dt} = \frac{1}{\varepsilon} f(x, y) \\
\frac{dy}{dt} = \frac{1}{\varepsilon^2} g(y) + \frac{1}{\varepsilon} \frac{dW}{dt},
\]
in the case where \( g \) is divergence-free.

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

3. Consider the Stokes law (11.6.9) in the case where the mass is small:
\[
\varepsilon^a \frac{d^2x}{dt^2} = \frac{1}{\varepsilon} f(x) y - \frac{dx}{dt} + \sigma \frac{dU}{dt} \\
\frac{dy}{dt} = -\frac{\alpha y}{\varepsilon^2} + \sqrt{\frac{2\alpha}{\varepsilon^2}} \frac{dV}{dt}.
\]
Derive the limiting equation satisfied by \( x \) in the cases \( a = 1, 2 \) and \( a = 3 \). Comment on your findings.

4. Consider the equation of motion
\[
\frac{dx}{dt} = f(x) + \sigma \frac{dW}{dt},
\]
where \( v \) is periodic with mean zero. It is of interest to understand how \( x \) behaves on large length and timescales. To this end we rescale the equation of motion by setting \( x \rightarrow x/\varepsilon \) and \( t \rightarrow t/\varepsilon^2 \) and introduce \( z = x\varepsilon \). Write down a pair of coupled SDE for \( x \) and \( z \). Generalize the methods developed in this chapter to enable eliminate \( z \) and obtain an effective equation for \( x \).

5. Carry out the analysis presented in Section 11.6.6 in arbitrary dimensions. Does the limiting equation have the same structure as in the one dimensional case?

6. Derive equation (11.6.19) from (11.6.17) when \( \eta(t) \) is given by (11.6.18).
Chapter 12

Homogenization for Elliptic Equations

12.1 Introduction

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

12.2 Full Equations

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

\[-\nabla \cdot \left( A \left( \frac{x}{\varepsilon} \right) \nabla u^\varepsilon \right) = f, \text{ for } x \in \Omega, \tag{12.2.1a}\]
\[u^\varepsilon (x) = 0, \text{ for } x \in \partial \Omega. \tag{12.2.1b}\]

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 coefficient \(A(y)\) is smooth, 1–periodic and uniformly elliptic. Furthermore,
will take the function $f(x)$ to be smooth and independent of $\varepsilon$. The assumptions that we make are

$$f \in C^\infty([0,1]^d, \mathbb{R}^d); \quad (12.2.2a)$$

$$A \in C^\infty(Y, \mathbb{R}^d); \quad (12.2.2b)$$

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

In the above $Y = [0, 1]^d$ denotes the (reference) unit cell (which is actually the $d$-dimensional unit torus $\mathbb{T}^d$.) For simplicity we assume that the matrix $A$ is symmetric, although this is not strictly necessary. With this assumption, (12.2.2c) states that the matrix-valued function $A(y)$ is uniformly positive-definite. 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: in many applications one expects that the coefficient $A(y)$ will have jumps when passing from one phase to the other.

### 12.3 Simplified Equations

Define the homogenized coefficient by the formula

$$\overline{A} = \int_Y \left( A(y) + A(y)\nabla \chi(y)^T \right) dy \quad (12.3.1)$$

where the field $\chi(y)$ satisfies the cell problem

$$-\nabla_y \cdot \left( A(y)\nabla \chi(y)^T \right) = \nabla \cdot A(y)^T, \quad \chi(y) \text{ is 1–periodic.} \quad (12.3.2)$$

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

$$-\overline{A} : \nabla u = f, \quad \text{for } x \in \Omega, \quad (12.3.3a)$$

$$u(x) = 0, \quad \text{for } x \in \partial \Omega. \quad (12.3.3b)$$

Notice that the field $\chi$ is undetermined up to a constant vector. However, since only $\nabla \chi$ enters the homogenized equation, the value of this constant is irrelevant.
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 expect that the solution is also a periodic function of its argument $\frac{x}{\varepsilon}$. Hence, we assume the following ansatz for the solution $u^\varepsilon$:

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

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 the assumption of scale separation enters: 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.

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

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

In other words, the total derivative (abusing slightly notation) of a function $f(x) := f \left( x, \frac{x}{\varepsilon} \right)$ can be expressed as

$$\nabla_x f^\varepsilon(x) = \nabla_x f(x, y) \bigg|_{y=\frac{x}{\varepsilon}} + \frac{1}{\varepsilon} \nabla_y f(x, y) \bigg|_{y=\frac{x}{\varepsilon}},$$

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 re-write the differential operator

$$A^\varepsilon := -\nabla_x \cdot (A(y) \nabla_x)$$
in the form
\[ A^\varepsilon = \frac{1}{\varepsilon^2} A_0 + \frac{1}{\varepsilon} A_1 + A_2, \]  
(12.4.2)

where
\[ A_0 := -\nabla_y \cdot (A(y)\nabla_y), \]  
(12.4.3a)
\[ A_1 := -\nabla_y \cdot (A(y)\nabla_x) - \nabla_x \cdot (A(y)\nabla_y), \]  
(12.4.3b)
\[ A_2 := -\nabla_x \cdot (A(y)\nabla_x). \]  
(12.4.3c)

Notice that the coefficients in all the operators defined above are periodic functions of \( y \).

Equation (12.2.1), on account of (12.4.2), becomes:
\[ \left( \frac{1}{\varepsilon^2} A_0 + \frac{1}{\varepsilon} A_1 + A_2 \right) u^\varepsilon = f, \quad \text{for } x \in \Omega, \]  
(12.4.4a)
\[ u^\varepsilon(x) = 0, \quad \text{for } x \in \partial\Omega. \]  
(12.4.4b)

We substitute (12.4.1) into (12.4.4) to deduce:
\[ \frac{1}{\varepsilon^2} A_0 u_0 + \frac{1}{\varepsilon} (A_0 u_1 + A_1 u_0) + (A_0 u_2 + A_1 u_1 + A_2 u_0) + O(\varepsilon) = f. \]  
(12.4.5)

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

\[ O(1/\varepsilon^2) \quad A_0 u_0 = 0, \quad u_0 \text{ is 1–periodic}, \]  
(12.4.6a)
\[ O(1/\varepsilon) \quad A_0 u_1 = -A_1 u_0, \quad u_1 \text{ is 1–periodic}, \]  
(12.4.6b)
\[ O(1) \quad A_0 u_2 = -A_1 u_1 - A_2 u_0 + f, \quad u_2 \text{ is 1–periodic}. \]  
(12.4.6c)

Note that the elliptic operator \( A_0 \), when equipped with periodic boundary conditions, is self-adjoint, and has a one dimensional null-space comprising constants in \( y \). From (12.4.6a) we deduce that \( u_0(x, y) = u(x) \) – the solution is \( y \) independent. The remaining two equations are of the form
\[ A_0 v = h, \quad v \text{ is 1–periodic}, \]  
(12.4.7)

with \( v = v(x, y) \) and similarly \( h = h(x, y) \), and \( A_0 \) a differential operator in \( y \) in which \( x \) appears as a parameter. By the Fredholm alternative this equation has a solution if and only if
\[ \int_y h(x, y)dy = 0. \]
Among all solutions of (12.4.7) which satisfy this solvability condition, we will choose the unique solution whose integral over \( \mathcal{Y} \) vanishes:

\[ \mathcal{A}_0 v = h, \quad v \text{ is 1–periodic, } \int_{\mathcal{Y}} v \ dy = 0. \]

Let us proceed now with (12.4.6b) which becomes

\[ \mathcal{A}_0 u_1 = \left( \nabla_y \cdot A^T \right) \cdot \nabla_x u, \quad u_1 \text{ is 1–periodic, } \int_{\mathcal{Y}} u_1 \ dy = 0. \]  \( 12.4.8 \)

The solvability condition is satisfied because

\[ \int_{\mathcal{Y}} \left( \nabla_y \cdot A^T \right) \cdot \nabla_x u \ dy = \nabla_x u \cdot \int_{\mathcal{Y}} \nabla_y \cdot A^T \ dy = 0, \]

by the divergence theorem and periodicity of \( A(\cdot) \). We seek a solution of (12.4.8) using separation of variables:

\[ u_1(x, y) = \chi(y) \cdot \nabla_x u(x). \]  \( 12.4.9 \)

Upon substituting (12.4.9) into (12.4.8) we obtain the cell problem (12.3.2) for \( \chi \).

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.

Now we consider equation (12.4.6c). In order for this equation to be well posed it is necessary and sufficient for the right hand side to average to zero. Since we have assumed that the function \( f(x) \) is independent of \( y \) the solvability condition implies:

\[ \int_{\mathcal{Y}} \left( A_2 u_0 + A_1 u_1 \right) \ dy = f. \]

\( 12.4.10 \)

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

\[ \int_{\mathcal{Y}} A_2 u_0 \ dy = \int_{\mathcal{Y}} -\nabla_x \cdot (A(y) \nabla_x u) \ dy = -\nabla_x \left[ \left( \int_{\mathcal{Y}} A(y) \ dy \right) \nabla_x u(x) \right] = -\left( \int_{\mathcal{Y}} A(y) \ dy \right) : \nabla_x \nabla_x u(x). \]

\( 12.4.11 \)
Moreover
\[ \int_Y A_1 u_1 \, dy = \int_Y (-\nabla_y \cdot (A(y) \nabla_x u_1) - \nabla_x \cdot (A(y) \nabla_y u_1)) \, dy =: I_1 + I_2. \]

The first term \( I_1 = 0 \) by the divergence theorem. Now we consider \( I_2 \):
\[ I_2 = \int_Y -\nabla_x \cdot (A(y) \nabla_y u_1) \, dy = -\int_Y A(y) : \nabla_x \nabla_y (\chi \cdot \nabla_x u) \, dy = -\left( \int_Y (A(y) \nabla_y \chi(y)^T) \, dy \right) : \nabla_x \nabla_x u. \quad (12.4.12) \]

We substitute (12.4.12) and (12.4.11) in (12.4.10) to obtain the homogenized equation
\[ -\overline{A} : \nabla_x \nabla u = f, \quad \text{for } x \in \Omega, \]
\[ u(x) = 0, \quad \text{for } x \in \partial \Omega, \]
where the homogenized coefficient \( \overline{A} \) is given by the formula:
\[ \overline{A} = \int_Y (A(y) + A(y) \nabla \chi(y)^T) \, dy. \]

## 12.5 Properties of the Homogenized Coefficients

In this section we study some basic properties of the effective coefficients. In particular, we show that the effective coefficients matrix \( \overline{A} \) is positive definite, which implies that the homogenized differential operator is uniformly elliptic. 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 symmetric bilinear form
\[ a_1(\psi, \phi) = \int_Y \langle \nabla_y \phi, A(y) \nabla_y \psi \rangle \, dy, \quad (12.5.1) \]
defined for all functions \( \phi, \psi \in C^1(Y) \). We start by obtaining an alternative, equivalent formulation for the cell problem. The formulation is closely related to the
12.5. PROPERTIES OF THE HOMOGENIZED COEFFICIENTS

The weak form of the elliptic PDE 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 vector $\chi$ is continuously differentiable and periodic: $\chi_\ell(y) \in C^1_{\text{per}}(Y)$.

As usual, $e_\ell$ denotes the unit vector with $i^{\text{th}}$ entry $\delta_{il}$. Also let $y_\ell$ denote the $\ell^{\text{th}}$ component of the vector $y$. Note that $e_\ell = \nabla_y y_\ell$. We let $a_\ell(y)$ denote the $\ell^{\text{th}}$ column of $A(y)$. Then the cell problem can be written as

$$
\nabla_y \cdot \left( A \nabla_y \chi_\ell + a_\ell \right) = 0, \quad \ell = 1, \ldots, d.
$$

Using this we obtain the following useful lemma.

**Lemma 12.2.** The cell problem (12.3.2) can be written in the form

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

**Proof.** We multiply the cell problem as formulated above by an arbitrary function $\phi \in C^1_{\text{per}}(Y)$ and integrate by parts over the unit cell to obtain

$$
\int_Y \nabla_y \phi \cdot \left( A \nabla_y \chi_\ell + a_\ell \right) dy = 0
$$

$$
\Rightarrow \int_Y \nabla_y \phi \cdot \left( A \nabla_y \chi_\ell + Ae_\ell \right) dy = 0
$$

$$
\Rightarrow \int_Y \nabla_y \phi \cdot \left( A \nabla_y \chi_\ell + A \nabla_y y_\ell \right) dy = 0.
$$

This gives the desired result by symmetry of the bilinear form.

Using this lemma we give an alternative representation formula for the homogenized coefficients. Note that, since $A$ is symmetric, so is the bilinear form $a_1$. Hence the following lemma shows that $\bar{A}$ is also symmetric.

**Lemma 12.3.** The effective matrix $\bar{A}$ has components given by

$$
\pi_{ij} = a_1(\chi_j + y_j, \chi_i + y_i), \quad i, j = 1, \ldots, d.
$$

**Proof.** We use formula (12.3.1), together with the cell problem written in the form
(12.5.2) and symmetry of \( a_1(\cdot, \cdot) \), to give
\[
\pi_{ij} = e_i \cdot \overline{A} e_j \\
= \int_Y \left( e_i \cdot A e_j + e_i \cdot A \nabla \chi^T e_j \right) dy \\
= \int_Y \nabla_y y_i \cdot A \left( \nabla_y y_j + \nabla_y \chi_j \right) dy \\
= a_1(\chi_j + y_j, y_i).
\]
But
\[
a_1(\chi_j + y_j, \chi_i) = 0
\]
by taking \( \phi = \chi_i \) in the previous lemma. Hence the result follows by the bilinearity of \( a_1 \).

In addition to being symmetric, the matrix \( \overline{A} \) is also positive definite.

**Theorem 12.4.** The matrix of homogenized coefficients \( \overline{A} \) is symmetric positive definite.

**Proof.** The previous proof establishes symmetry thus we need to show that there exists a constant \( \alpha > 0 \) such that
\[
\langle \xi, \overline{A} \xi \rangle \geq \alpha |\xi|^2, \quad \forall \xi \in \mathbb{R}^d.
\]
Let \( \xi \in \mathbb{R}^d \) be an arbitrary vector. We use the representation formula (12.5.3) to deduce that:
\[
\langle \xi, \overline{A} \xi \rangle = a_1(w, w),
\]
with \( w = \xi.(\chi + y) \). We use now the uniform positive definiteness of \( A(y) \) to obtain
\[
a_1(w, w) \geq \alpha \int_Y |\nabla_y w|^2 dy \geq 0.
\]
Thus \( \overline{A} \) is nonnegative.

To show that it is positive definite we argue as follows. Let us assume that
\[
\langle \xi, \overline{A} \xi \rangle = 0
\]
for some \( \xi \). Then \( \nabla_y w = 0 \) and \( w = c \), a constant vector; consequently
\[
\xi \cdot y = c - \xi \cdot \chi.
\]
The right hand side of this equation is 1–periodic 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 12.5.** 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 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)

\[
\bar{a}_{ij} = \int_Y a_{ii} \frac{\partial \chi_j}{\partial y_i} \, dy, \quad i \neq j.
\]

This expression is not necessarily equal to zero. This 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, the first in one dimension. Essentially, the one–dimensional case is the only case where 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 the one dimensional case. 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:

\[
\begin{align*}
-\frac{d}{dx} \left( a \left( \frac{x}{\varepsilon} \right) \frac{du^\varepsilon}{dx} \right) &= f, \quad x \in (0, 1), \\

u^\varepsilon(0) &= u^\varepsilon(1) = 0.
\end{align*}
\]  

(12.6.1a, 12.6.1b)

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

\[
\alpha \leq a(y) \leq \beta, \quad \forall y \in [0, 1].
\]  

(12.6.2)

We also assume that \( f \) is smooth.

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

\[
-\frac{d}{dy} \left( a(y) \frac{d\chi}{dy} \right) = \frac{da(y)}{dy}, \quad y \in (0, 1),
\]  

(12.6.3a)

\[
\chi \text{ is } 1\text{–periodic, } \int_0^1 \chi(y) \, dy = 0.
\]  

(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

\[
\bar{a} = \int_0^1 \left( a(y) + a(y) \frac{d\chi(y)}{dy} \right) \, dy
\]  

\[
= \left\langle a(y) \left( 1 + \frac{d\chi(y)}{dy} \right) \right\rangle.
\]  

(12.6.4)

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

\[
\langle f(y) \rangle := \int_{\mathcal{Y}} f(y) \, dy,
\]

for the average over \( \mathcal{Y} \).

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

\[
a(y) \frac{d\chi}{dy} = -a(y) + c_1.
\]  

(12.6.5)
The constant $c_1$ is undetermined at this point. The inequality (12.2.2c) 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:

$$
\chi(0) = \chi(1) \quad \Rightarrow \quad 0 = 1 - c_1 \int_0^1 \frac{1}{a(y)} \, dy
$$

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

Thus, from (12.6.5),

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

Notice that $c_2$ is not required. We substitute this expression in equations (12.6.4) to obtain:

$$
\pi = \langle (a(y)^{-1})^{-1} \rangle.
$$

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 inverse homogenized coefficient is the average of the inverse of the unhomogenized coefficient.

### 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

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

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

$$
\chi_\ell = \chi_\ell(y_1), \quad \ell = 1, 2.
$$
Substitution into (12.3.2) we conclude that the cell problem becomes
\[
- \frac{d}{dy_1} \left( a_{11}(y_1) \frac{d\chi_\ell(y_1)}{dy_1} \right) = \frac{da_{1\ell}(y_1)}{dy_1}, \quad \ell = 1, 2
\] (12.6.9)
with periodic boundary conditions. Similarly, the formula for the homogenized coefficients (12.3.1) becomes:
\[
\bar{a}_{ij} = \int_0^1 \left( a_{ij}(y_1) + a_{i1}(y_1) \frac{d\chi_j(y_1)}{dy_1} \right) dy_1, \quad i, j = 1, 2.
\] (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
\[
\frac{d\chi_\ell}{dy_1} = -\frac{a_{1\ell}}{a_{11}} + c_1 \frac{1}{a_{11}}, \quad \ell = 1, 2
\] (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:
\[
\frac{d\chi_1}{dy_1} = -1 + \frac{1}{\langle a_{11}(y)^{-1} \rangle a_{11}(y)}.
\] (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
\[
\frac{d\chi_2(y_1)}{dy_1} = -\frac{a_{12}(y_1)}{a_{11}(y_1)} + \frac{\langle a_{12}(y_1)/a_{11}(y_1) \rangle}{\langle a_{11}^{-1}(y_1) \rangle} \frac{1}{a_{11}(y_1)}.
\] (12.6.13)

Now we can compute the homogenized coefficients. We start with \( \bar{\sigma}_{11} \). The calculation is the same as in the one-dimensional case:
\[
\bar{\sigma}_{11} = \langle a_{11}(y_1)^{-1} \rangle^{-1}.
\] (12.6.14)
12.6. APPLICATIONS

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

\[
\overline{\sigma}_{12} = \int_0^1 \left( a_{12}(y_1) + a_{11}(y_1) \frac{d\chi_2(y_1)}{dy_1} \right) dy
\]

\[
= \int_0^1 \left( a_{12}(y_1) + a_{11}(y_1) \left( -\frac{a_{12}(y_1)}{a_{11}(y_1)} + \frac{\langle a_{12}(y_1)/a_{11}(y_1) \rangle}{\langle a_{11}^{-1}(y_1) \rangle} \frac{1}{a_{11}(y_1)} \right) \right) dy
\]

\[
= \int_0^1 \left( a_{12}(y_1) - a_{12}(y_1) + \frac{\langle a_{12}(y_1)/a_{11}(y_1) \rangle}{\langle a_{11}^{-1}(y_1) \rangle} \right) dy
\]

\[
= \frac{\langle a_{12}(y_1)/a_{11}(y_1) \rangle}{\langle a_{11}^{-1}(y_1) \rangle}.
\]

Hence

\[
\overline{\sigma}_{12} = \frac{\langle a_{12}(y_1) \rangle}{\langle a_{11}^{-1}(y_1) \rangle} \langle a_{11}^{-1}(y_1) \rangle^{-1}.
\]

(12.6.15)

By the symmetry of \( \overline{A} \) we deduce and of \( A \) we have

\[
\overline{\sigma}_{21} = \frac{\langle a_{21}(y_1) \rangle}{\langle a_{11}^{-1}(y_1) \rangle} \langle a_{11}^{-1}(y_1) \rangle^{-1}.
\]

(12.6.16)

Finally we consider \( \overline{\sigma}_{22} \):

\[
\overline{\sigma}_{22} = \int_0^1 \left( a_{22}(y_1) + a_{21}(y_1) \frac{d\chi_2(y_1)}{dy_1} \right) dy
\]

\[
= \int_0^1 \left( a_{22}(y_1) + a_{21}(y_1) \left( -\frac{a_{12}(y_1)}{a_{11}(y_1)} + \frac{\langle a_{12}(y_1)/a_{11}(y_1) \rangle}{\langle a_{11}^{-1}(y_1) \rangle} \frac{1}{a_{11}(y_1)} \right) \right) dy
\]

\[
= \int_0^1 \left( a_{12}(y_1) - a_{12}(y_1) + \frac{a_{21}(y_1)}{a_{11}(y_1)} \frac{\langle a_{12}(y_1)/a_{11}(y_1) \rangle}{\langle a_{11}^{-1}(y_1) \rangle} \right) dy
\]

\[
= \frac{\langle a_{21}(y_1) \rangle}{\langle a_{11}^{-1}(y_1) \rangle} \left\langle \frac{a_{12}(y_1)}{a_{11}(y_1)} \right\rangle \langle a_{11}^{-1}(y_1) \rangle^{-1} + \frac{a_{22}(y_1) - a_{12}(y_1)a_{21}(y_1)}{a_{11}(y_1)}.
\]

Consequently:

\[
\overline{\sigma}_{22} = \frac{\langle a_{21}(y_1) \rangle}{\langle a_{11}(y_1) \rangle} \left\langle \frac{a_{12}(y_1)}{a_{11}(y_1)} \right\rangle \langle a_{11}^{-1}(y_1) \rangle^{-1} + \frac{a_{22}(y_1) - a_{12}(y_1)a_{21}(y_1)}{a_{11}(y_1)}.
\]

(12.6.17)

It is evident from formulæ (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 is developed and used systematically in [17], where references to the earlier literature can be found. See also [76] and the references therein. The one dimensional problem (see Section 12.6.1) was studied in [138], 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 non–periodic setting; see [68, Ch. 5], [30, Ch. 5]. 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 [107] and the references to the original papers therein. The two–dimensional case that we treated in Section 12.6.2 can be easily extended to the $d$–dimensional one, $d \geq 2$, i.e. to the case where $a_{ij}(y) = a_{ij}(y_1)$, $i, j = 1, \ldots, d$. See [107].

The results that we presented in this chapter can be extended and generalized in various ways. Below we discuss about some generalizations of the basic theory.

**Different boundary conditions.** The elliptic boundary value problem (12.2.1) that we considered in the previous section was 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 would enable us to obtain the homogenized equation for Neumann or mixed boundary conditions. This is not surprising since the derivation of the homogenized equation was based on the analysis of "local" problems of the form (12.4.7). This local problem cannot really "see" the boundary.

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 $O(\varepsilon)$ introduces boundary layers [68, ch. 3]. Boundary layers affect the convergence rate, i.e. the rate with 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.

\footnote{The presence of boundary and initial layers is a common feature in all problems of singular perturbations. See e.g. [68] and [77] for further details.}
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 [12] for details.

**Locally Periodic Coefficients.** In the Dirichlet problem that we analyzed in Section 12.4 we assumed that the coefficients \( \{a_{ij}^\varepsilon(x)\}_{i,j=1}^d \) depend only on the microscale, i.e.

\[
A^\varepsilon(x) = A\left(\frac{x}{\varepsilon}\right), \quad i,j = 1, \ldots d,
\]

with \( A(y) \) being 1-periodic functions. However, the method of multiple scales would also be 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

\[
-\nabla_x \cdot \left( A\left(x, \frac{x}{\varepsilon}\right) \nabla_x u^\varepsilon \right) = f, \quad \text{for } x \in \Omega \quad (12.7.1a)
\]

\[
u^\varepsilon(x) = 0, \quad \text{for } x \in \partial \Omega. \quad (12.7.1b)
\]

Now the homogenized coefficients are functions of \( x \):

\[
-\nabla_x \cdot \left( \bar{A}(x) \nabla_x u \right) = f, \quad \text{for } x \in \Omega \quad (12.7.2a)
\]

\[
u(x) = 0, \quad \text{for } x \in \partial \Omega, \quad (12.7.2b)
\]

and the cell problem reads:

\[
-\nabla_y \cdot \left( A(x, y) \nabla_y \chi_T(x, y) \right) = \nabla_x \bar{A}(x, y)^T. \quad (12.7.3)
\]

The homogenized coefficients are given by the formula:

\[
\bar{A}(x) = \int_Y \left( A(x, y) + A(x, y) \nabla_x \chi(x, y)^T \right) dy. \quad (12.7.4)
\]

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 \( x \in \Omega \).
Semilinear PDE

It is possible to study semilinear elliptic PDE with the form

$$-\nabla \cdot \left( A \left( \frac{x}{\varepsilon} \right) \nabla u_{\varepsilon} \right) = f(u_{\varepsilon}), \quad \text{for } x \in \Omega, \quad (12.7.5a)$$

$$u_{\varepsilon}(x) = 0, \quad \text{for } x \in \partial \Omega. \quad (12.7.5b)$$

The homogenized equation takes the form

$$-\overline{A} : \nabla \nabla u = f(u), \quad \text{for } x \in \Omega, \quad (12.7.6a)$$

$$u(x) = 0, \quad \text{for } x \in \partial \Omega. \quad (12.7.6b)$$

Higher Order Correctors

In section (12.2) we studied homogenization for the Dirichlet boundary value problem (12.2.1) using the method of multiple scales. We derived the homogenized equation (12.3.3) and the cell problem (12.3.2). We also computed the first order correction

$$u_1 \left( x, \frac{x}{\varepsilon} \right) = \chi \left( \frac{x}{\varepsilon} \right) \cdot \nabla_x u(x) + \hat{u}_1(x). \quad (12.7.7)$$

We can proceed with solving equation (12.4.6) and computing the second corrector field $u_2(x, y)$. This is given by

$$u_2(x, y) = \sum_{i,j=1}^{d} \Theta(y) : \nabla_x \nabla_x u + \hat{u}_2(x) \quad (12.7.8)$$

where the second order corrector field $\{ \Theta^{ij}(y) \}_{i,j=1}^{d}$ satisfies

$$A_0 \Theta = B. \quad (12.7.9)$$

Here $B(y)$ is given by

$$B(y) := -\overline{A} + A(y) + A(y) \nabla_y \chi(y)^T + \chi(y) \otimes \nabla_y \left( A(y)^T \right).$$

Reiterated Homogenization. The method of multiple scales can be extended to situations where there are $k$ length scales in the problem, i.e. when the matrix $A$ has the form

$$A = A \left( \frac{x}{\varepsilon}, \frac{x}{\varepsilon^2}, \ldots, \frac{x}{\varepsilon^k} \right),$$
and \( A \) is 1–periodic in all of its arguments.

This is known as reiterated homogenization – [17, Sec. 1.8]. A rigorous analysis of reiterated homogenization in a quite general setting is presented in [5]. Reiterated homogenization has recently found applications in the problem of advection and diffusion of passive tracers in fluids. See, e.g. [122, 103, 104] for details. When there are infinitely many scales in the problem, without a clear separation, the homogenization result breaks down. See [15].

**Bounds on the Effective coefficients.** 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 is can be done only numerically. It is possible, however, to obtain bounds on the magnitude of the effective coefficients. Various tools, for example variational formulas for the effective coefficients have been developed. We refer to [106, 144] for various results in this direction.

**Numerical methods for Elliptic PDE with rapidly oscillating coefficients.** 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 the other are based on homogenization, have been developed over the last few years. We refer to [69, 35, 1, 39] and the references therein on this topic.

**The Method of Multiple Scales for Evolution PDE.** The method in this chapter readily extends to intial/boundary value problem such as the following parabolic (diffusion) PDE:

\[ \frac{\partial u^\varepsilon}{\partial t} - \nabla_x \cdot \left( A \left( \frac{x}{\varepsilon} \right) \nabla_x u^\varepsilon \right) = f(x, t) \quad \text{in} \; \Omega \times (0, T), \]  
\[ u^\varepsilon(x, t) = 0 \quad \text{on} \; \partial \Omega \times (0, T) \]  
\[ u^\varepsilon(x, 0) = u_{\text{in}}(x) \quad \text{in} \; \Omega. \]

Similarly, one can also study the problem of homogenization for hyperbolic (wave) equations:

\[ \frac{\partial^2 u^\varepsilon}{\partial t^2} - \nabla_x \cdot \left( A \left( \frac{x}{\varepsilon} \right) \nabla_x u^\varepsilon \right) = f(x, t) \quad \text{in} \; \Omega \times (0, T), \]
Another time-dependent situation of interest arises when the coefficients of the evolution PDE oscillate in time as well as space. Consider the following parabolic PDE
\[
\frac{\partial u^\varepsilon}{\partial t} - \nabla_x \cdot \left( A \left( \frac{x}{\varepsilon}, \frac{t}{\varepsilon^2} \right) \nabla_x u^\varepsilon \right) = f(x, t) \quad \text{in} \; \Omega \times (0, T), \tag{12.7.12a}
\]
\[
u^\varepsilon(x, t) = 0 \quad \text{on} \; \partial \Omega \times (0, T), \tag{12.7.12b}
\]
\[
u^\varepsilon(x, 0) = v_{in}(x) \quad \text{in} \; \Omega. \tag{12.7.12c}
\]

Now we take the coefficients $A(y, \tau)$ to be 1–periodic in both $y$ and $\tau$. This means that we have to introduce two fast variables: $y = \frac{x}{\varepsilon}$ and $\tau = \frac{t}{\varepsilon^2}$. More information on homogenization for evolution equations with space–time dependent coefficients can be found in [17, Ch. 3].

We study parabolic problems in the next chapter. However, in contrast to what is covered in this chapter, we consider spatial operators which are not in divergence form.

### 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). Solve equations (12.7.7) (after checking that the solvability condition is satisfied) to compute all higher order terms in the expansion and to obtain the corresponding cell problems.
3. Consider the Dirichlet problem (12.2.1) for a $d$–dimensional layered material, i.e.

\[ a_{ij}(y) = a_{ij}(y_1), \quad 1\text{-periodic in } y_1, \quad i, j = 1, \ldots, d. \]

Solve the corresponding cell problem and obtain formulas for the homogenized coefficients for $d \geq 3$, arbitrary.

4. Consider the Dirichlet problem (12.2.1) for a $d$–dimensional isotropic material, i.e.

\[ a_{ij}(y) = a(y)\delta_{ij}, \quad 1\text{-periodic, } i, j = 1, \ldots, d, \]

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 formulas for the homogenized coefficients (hint: use separation of variables).

5. Consider the boundary value problem (12.7.1). Assume that $A(x,y)$ 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.

6. Consider the Dirichlet problem

\[
\begin{align*}
-\nabla \cdot \left( A \left( \frac{x}{\varepsilon}, \frac{z}{\varepsilon^2} \right) \nabla u^\varepsilon \right) &= f, \quad \text{for } x \in \Omega \quad (12.8.1a) \\
u^\varepsilon(x) &= 0, \quad \text{for } x \in \partial\Omega. \quad (12.8.1b)
\end{align*}
\]

where the coefficients $A(y, z)$ are periodic in both $y$ and $z$ with period 1. Use the 3–scale expansion

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

to derive an effective homogenized equation, together with the formula for the homogenized coefficients and two cell problems.
7. 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

\[
-\nabla \cdot \left( \overline{A} \left( \frac{x}{\varepsilon} \right) \nabla u^\varepsilon \right) = f, \quad \text{for} \ x \in \Omega \tag{12.8.3a}
\]

\[
u^\varepsilon(x) = 0, \quad \text{for} \ x \in \partial \Omega, \tag{12.8.3b}
\]

where \( \overline{A}(y) \) is given by the expression derived in the preceding section of the question.

8. 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 equation (12.7.10).

9. Consider the initial boundary value problem (12.7.12). Explain why it is natural to the period of oscillations in time to be \( \varepsilon^2 \), when the period of oscillations in space is \( \varepsilon \)? Carry out the homogenization analysis based on the method of multiple scales for the cases where the coefficients are of the form \( \{ a_{ij}(x/\varepsilon, t/\varepsilon^2) \}_{i,j=1}^d \) and \( \{ a_{ij}(x/\varepsilon, t/\varepsilon^3) \}_{i,j=1}^2 \).

10. Use the method of multiple scales to derive the homogenized equation from (12.7.11).

11. Prove that the homogenized coefficient \( \overline{a} \) for equation (12.6.1) under (12.6.2) has the same upper and lower bound as \( a(y) \). Moreover, show that it is bounded from above by the average of \( a(y) \):

\[
\alpha \leq \overline{a} \leq \beta,
\]

and

\[
\overline{a} \leq \langle a(y) \rangle.
\]

\(^2\)See [17, ch.3] for further details on the derivation of the homogenized equations using the method of multiple scales.
12. Show that the equation (12.7.5) can be homogenized to obtain the effective equation (12.7.6).

13. a. (i) Consider the eigenvalue problem

\[-\Delta u^\varepsilon + \frac{1}{\varepsilon} f(x/\varepsilon) u^\varepsilon = \lambda^\varepsilon u^\varepsilon, \quad x \in \Omega\]
\[u^\varepsilon = 0, \quad x \in \partial \Omega.\]

Assume that \(f : \mathbb{T}^d \to \mathbb{R}^d\) is smooth and periodic and that
\[\int_{\mathbb{T}^d} f(y) dy = 0.\]

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 \(f \equiv 0\)?

14. a. (i) Let \(A : \mathbb{T}^d \to \mathbb{R}^{d \times d}\) be smooth and periodic and consider the eigenvalue problem

\[-\nabla \cdot \left( A(x/\varepsilon) \nabla u^\varepsilon \right) = \lambda^\varepsilon u^\varepsilon, \quad x \in \Omega\]
\[u^\varepsilon = 0, \quad x \in \partial \Omega.\]

Use a multiscale expansion to find an approximation to the eigenvalue problem in which \(\varepsilon \to 0\) is eliminated.

b. (ii) State conditions on the matrix \(A(y)\) under which the homogenized equation is a well-posed elliptic boundary value problem.
Chapter 13

Homogenization for Parabolic Equations

13.1 Introduction

In this chapter we use multi–scale 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 Chapters 11. This connection will be made more explicit at the end of the chapter.

13.2 Full Equations

We study the following initial value (Cauchy) problem

\[ \frac{\partial u(x, t)}{\partial t} = b(x) \cdot \nabla u(x, t) + D \Delta u(x, t) \quad \text{for} \; (x, t) \in \mathbb{R}^d \times \mathbb{R}^+, \quad (13.2.1a) \]

\[ u = u_0 \quad \text{for} \; (x, t) \in \mathbb{R}^d \times 0, \quad (13.2.1b) \]

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

\[ u_0(x) = f(\varepsilon x), \quad (13.2.2) \]

with \( \varepsilon \ll 1 \). Because the initial data is slowly varying it is natural to look to large length and time-scales to see the effective behavior of the PDE. If \( b \) averages to
zero in an appropriate sense then, as we will now show, the effective behavior of $u$ is that of a pure diffusion.

To see this effect we introduce the new variables $z$, $\tau$ through

$$z = \varepsilon x, \quad \tau = \varepsilon^2 t$$

and define

$$u^\varepsilon(z, \tau) = u \left( \frac{z}{\varepsilon}, \frac{\tau}{\varepsilon^2} \right).$$

The rescaled field $u^\varepsilon(z, \tau)$ satisfies the equation

$$\frac{\partial u^\varepsilon(z, \tau)}{\partial \tau} = \frac{1}{\varepsilon} b \left( \frac{z}{\varepsilon} \right) \cdot \nabla_z u^\varepsilon(z, \tau) + D \Delta_z u^\varepsilon(z, \tau).$$

Returning, for notational clarity, to the variables $(x, t)$, and using (13.2.2), we arrive at the following initial value problem:

$$\frac{\partial u^\varepsilon(x, t)}{\partial t} = \frac{1}{\varepsilon} b \left( \frac{x}{\varepsilon} \right) \cdot \nabla u^\varepsilon(x, t) + D \Delta u^\varepsilon(x, t). \quad (13.2.3a)$$

$$u^\varepsilon(x, 0) = f(x). \quad (13.2.3b)$$

This equation will be the object of our study. It can be derived directly from (13.2.1a) by setting $x \to x/\varepsilon$ and $t \to t/\varepsilon^2$.

Define the operator

$$L_0 = b(y) \cdot \nabla_y + D \Delta_y \quad (13.2.4)$$

with periodic boundary conditions and its adjoint $L^*$, also with periodic boundary conditions. Note that $L$ is the generator of an SDE. Hence it is natural to define the invariant distribution $\rho(y)$ to be the stationary solution of the adjoint equation:

$$L_0^* \rho = 0. \quad (13.2.5)$$

By Theorem 6.10 there is a unique solution to this equation, up to normalization. We assume that that the drift $b(y)$ is centered with respect to the invariant distribution of the fast process:

$$\int_{\mathbb{T}^d} b(y) \rho(y) \, dy = 0. \quad (13.2.6)$$

We will call this the centering condition. Since by Theorem 6.10 $L$ also has a one-dimensional null-space, comprising constants, it follows that the equation

$$Lv = -b$$

has a solution, unique up to constants, by the Fredholm alternative.
13.3 Simplified Equations

The field $\chi(y)$ is defined to be the solution of the cell problem

$$-L_0 \chi(y) = b(y), \ \chi \text{ is } 1 \text{-periodic, } \int_Y \chi(y) \rho(y) dy = 0. \quad (13.3.1)$$

The effective diffusion tensor (or effective diffusivity) is defined as

$$K = DI + 2D \int_{T^d} \nabla \chi(y)^T \rho(y) dy + \int_{T^d} \left(b(y) \otimes \chi(y)\right) \rho(y) dy. \quad (13.3.2)$$

**Result 13.1.** For $\varepsilon \ll 1$ and $t = \mathcal{O}(1)$ the solution of (13.2.3) $u^\varepsilon(x,t)$ is approximated by $u(x,t)$, the solution of

$$\frac{\partial u(x,t)}{\partial t} = \mathcal{K} : \nabla_x \nabla_x u(x,t),$$

$$u(x,0) = f(x). \quad (13.3.3)$$

13.4 Derivation of the Homogenized equation

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 (13.2.3) in the limit as $\varepsilon \to 0$. In particular, we want to derive Result 13.1.

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 (13.2.3) now becomes

$$\mathcal{L} = \frac{1}{\varepsilon^2} L_0 + \frac{1}{\varepsilon} L_1 + L_2.$$ 

where

$$L_0 = b(y) \cdot \nabla_y + D \Delta_y,$$

$$L_1 = b(y) \cdot \nabla_x + 2D \nabla_x \cdot \nabla_y,$$

$$L_2 = D \Delta_x.$$ 

---

1As in the elliptic case, this is where the assumption of scale separation enters: we will 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.
In terms of $x$ and $y$ equation (13.2.3a) becomes
\[
\frac{\partial u^\varepsilon}{\partial t} = \mathcal{L}u^\varepsilon.
\]
We seek a solution in the form of a multiple scales expansion
\[
u^\varepsilon(x, t) = u_0(x, y, t) + \varepsilon u_1(x, y, t) + \varepsilon^2 u_2(x, y, t) + \ldots
\]  
(13.4.1)
where $u_j(x, y, t), j = 1, 2, \ldots$, is a periodic function of $y$. We substitute (13.4.1) and equate terms of equal powers in $\varepsilon$. Note that $\mathcal{L}_0$ is a differential operator in $y$ only and recall that it is equipped with periodic boundary conditions. We obtain the following sequence of equations.

\[
\mathcal{O}(1/\varepsilon^2) \quad \mathcal{L}_0 u_0 = 0, \quad u_0 \text{ is } 1\text{-periodic},
\]  
(13.4.2a)

\[
\mathcal{O}(1/\varepsilon) \quad \mathcal{L}_0 u_1 = -\mathcal{L}_1 u_0, \quad u_1 \text{ is } 1\text{-periodic},
\]  
(13.4.2b)

\[
\mathcal{O}(1) \quad \mathcal{L}_0 u_2 = -\mathcal{L}_1 u_1 - \mathcal{L}_2 u_0 + \frac{\partial u_0}{\partial t}, \quad u_2 \text{ is } 1\text{-periodic}(13.4.2c)
\]

Note that $\mathcal{L}_0$ is a differential operator in $y$ only and recall that it is equipped with periodic boundary conditions.

Equation (13.4.2) implies that $u_0(x, y, t) = u(x, t)$ only. In order to proceed we need to study equations of the form
\[
-\mathcal{L}_0 v = f,
\]  
(13.4.3)

with periodic boundary conditions. The Fredholm alternative shows that this equation has a solution if and only if
\[
\int_{\mathbb{T}^d} f(y)\rho(y)dy = 0.
\]  
(13.4.4)

Notice that
\[
\mathcal{L}_1 u_0 = b(y_0 \cdot \nabla_x u(x, t)).
\]
The centering condition (13.2.6) shows us that (13.4.4) is satisfied so that (13.4.2) has a solution. We use separation of variables to write it 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 (up to constants), smooth solution to the cell problem.
13.5. EFFECTIVE DIFFUSIVITY

Now we proceed with the analysis of the $O(1)$ equation (13.4.2). The solvability condition 13.4.4 reads

$$
\int_{T^d} \left( \frac{\partial u}{\partial t} - L_2 u - L_1 u_1 \right) \rho \, dy = 0.
$$

The fact that $u_0 = u(x, t)$ is independent of $y$ enables us to rewrite the above equation in the form

$$
\frac{\partial u}{\partial t} = D \Delta u + \int_{T^d} (L_1 u_1) \rho \, dy.
$$

(13.4.5)

Now we have

$$
L_1 u_1 = \left( b \cdot \nabla_x (\chi \cdot \nabla_x u) + 2D \nabla_x \cdot \nabla_y (\chi \cdot \nabla_x u) \right)
$$

$$
= \left( b \otimes \chi + 2D \nabla_y \chi^T \right) : \nabla_x \nabla_x u.
$$

In view of the above calculation, equation (13.4.5) becomes

$$
\frac{\partial u(x, t)}{\partial t} = K : \nabla_x \nabla_x u(x, t),
$$

which is the homogenized equation (13.3.3a). The effective diffusivity $K$ is given by formula (13.3.2).

13.5 Properties of the Effective Diffusivity

In this section we show that the effective diffusivity is non-negative definite. This implies that the homogenized equation is well posed. To prove this we need to calculate the Dirichlet form associated with the operator $L_0$. The following is a direct consequence of Theorem 6.7 in the case of additive noise.

Lemma 13.2. Let $f(y)$ be a smooth periodic function. Then

$$
\int_{T^d} (-L_0 f(y)) f(y) \rho(y) \, dy = D \int_{T^d} |\nabla_y f(y)|^2 \rho(y) \, dy.
$$

(13.5.1)

We have the following.

Theorem 13.3. Let $\xi \in \mathbb{R}^d$ an arbitrary vector and let $\chi_\xi(y) := \chi(y) \cdot \xi$. Then

$$
\langle \xi, K \xi \rangle = \int_{T^d} |\xi + \nabla_y \chi_\xi(y)|^2 \rho(y) \, dy.
$$

In particular, the effective diffusivity in non-negative definite.
CHAPTER 13. HOMOGENIZATION FOR PARABOLIC EQUATIONS

Proof. Note that $-L_0 \chi_\xi = \xi \cdot b$. We use Lemma 13.2 to calculate

$$
\langle \xi, K \xi \rangle = D|\xi|^2 + 2D \int_{T^d} \xi \cdot \nabla_y \chi_\xi(y) \rho(y) \, dy + D \int_{T^d} (\xi \cdot b) \chi_\xi(y) \rho(y) \, dy
$$

$$
= D|\xi|^2 + 2D \int_{T^d} \xi \cdot \nabla_y \chi_\xi(y) \rho(y) \, dy + D \int_{T^d} |\nabla_y \chi_\xi(y)|^2 \rho(y) \, dy
$$

$$
= \int_{T^d} |\xi + \nabla_y \chi_\xi(y)|^2 \rho(y) \, dy.
$$

The fact that the effective diffusivity in non–negative definite follows immediately from the above equation.

It is of interest to know how the effective diffusion tensor $K$ compares with the original diffusion tensor $dI$. It turns out that the quadratic form $\langle \xi, K \xi \rangle$, for an arbitrary unit vector $\xi \in \mathbb{R}^d$ (the effective diffusivity) can be either greater or smaller than $D|\xi|^2$ (the value arising from the original diffusivity). This issue is discussed in detail in the next section.

We will see in the next section 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 $b(y)$ in (13.2.3a), 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 (13.2.6). Furthermore we will be able to compare the effective diffusivity and 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 $b$. Finally, we will study two particular cases of gradient and divergence free fields for which we can derive closed formulas for the effective diffusivity.

13.6.1 Gradient Vector Fields

We consider the case where the vector field $b(y)$ in equation (13.2.3a) is the gradient of a smooth, scalar periodic function,

$$
b(y) = -\nabla_y V(y).
$$

(13.6.1)
The function $V(y)$ is called the potential. In this case it is straightforward to derive a formula for the solution $\rho(y)$ of the stationary adjoint equation (13.2.5) with periodic boundary conditions. The distribution $\rho(y)$ defined in (13.6.3) below is called the Gibbs distribution. The normalization constant $Z$ is called the partition function.

**Lemma 13.4.** Assume that condition (13.6.1) holds and let $L_0^*$ denote the adjoint operator for the operator $L_0$ defined in (13.2.4). Then the equation

$$L_0^* \rho = 0, \quad \int_{\mathbb{T}^d} \rho(y) dy = 1,$$

(13.6.2)

subject to periodic boundary conditions on $\mathbb{T}^d$ has a unique solution given by

$$\rho(y) = \frac{1}{Z} e^{-V(y)/D}, \quad Z = \int_{\mathbb{T}^d} e^{-V(y)/D} dy.$$  

(13.6.3)

**Proof.** Equation (13.6.2), in view of equation (13.6.1), becomes

$$\nabla \cdot \left( \nabla_y V(y) \rho(y) + D \nabla_y \rho(y) \right) = 0.$$  

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 (13.6.2). Furthermore, by construction we have that

$$\int_{\mathbb{T}^d} \frac{1}{Z} e^{-V(y)/D} dy = 1,$$

and hence $\rho(y)$ is normalized. Thus we have constructed a solution. Uniqueness follows by the maximum principle. \qed

In the case of gradient flows the operator

$$L_0 = b(y) \cdot \nabla_y + D \Delta_y$$

(13.6.4)

with periodic boundary conditions becomes symmetric, in the appropriate function space. We have the following.

**Lemma 13.5.** Assume that condition (13.6.1) is satisfied and let $\rho(y)$ denote the Gibbs distribution (13.6.3). Then the operator $L_0$ given in (13.6.4) satisfies

$$\int_{\mathbb{T}^d} f(y) \left( L_0 \rho(y) \right) dy = \int_{\mathbb{T}^d} h(y) \left( L_0 \rho(y) \right) dy,$$

(13.6.5)

for all smooth periodic $f(y), h(y)$. 

Proof. Using the divergence theorem we have

\[
\int_{\mathbb{T}^d} f \mathcal{L}_0 h \rho \, dy = \frac{1}{Z} \int_{\mathbb{T}^d} f \left( - \nabla_y V \cdot \nabla_y h \right) e^{-V/D} \, dy + \frac{D}{Z} \int_{\mathbb{T}^d} f \Delta_y h e^{-V/D} \, dy \\
= \frac{D}{Z} \int_{\mathbb{T}^d} f \nabla_y h \cdot \nabla_y \left( e^{-V/D} \right) \, dy - \frac{D}{Z} \int_{\mathbb{T}^d} \nabla_y f \nabla_y h e^{-V/D} \, dy \\
- D \int_{\mathbb{T}^d} \left( \nabla_y f \cdot \nabla_y h \right) \rho (y) \, dy.
\]

The expression in the penultimate line is symmetric in \( f, h \) and hence (13.6.5) follows. \( \square \)

Remark 13.6. Actually this symmetry property arises quite naturally from the identity used in proving Theorem 6.7. Furthermore, the calculation used in the proof of the above lemma gives us the following useful formula

\[
\int_{\mathbb{T}^d} f \mathcal{L}_0 h \rho \, dy = -D \int_{\mathbb{T}^d} \left( \nabla_y f \cdot \nabla_y h \right) \rho (y) \, dy.
\] (13.6.6)

The Dirichlet form Theorem 6.7 follows from this upon setting \( f = h \).

Now we are ready to prove various properties of the effective diffusivity.

**Theorem 13.7.** Assume that \( b \) is a gradient so that (13.6.1) holds and let \( \rho (y) \) denote the Gibbs distribution (13.6.3). Then the effective diffusivity is positive semi-definite. Furthermore, the effective diffusivity is always depleted:

\[
\langle \xi, K \xi \rangle \leq D |\xi|^2 \quad \forall \xi \in \mathbb{R}^d.
\]

**Proof.** Non-negativity is proved in Theorem 13.3. To establish the comparison on
the effective diffusivity, note that we have
\[
K = DI + 2D \int_{T^d} \nabla \chi^T \rho \, dy + \int_{T^d} -\nabla_y V \otimes \chi \rho \, dy
\]
\[
= DI - 2D \int_{T^d} \nabla_y \rho \otimes \chi \, dy + \int_{T^d} -\nabla_y V \otimes \chi \rho \, dy
\]
\[
= DI - 2 \int_{T^d} -\nabla_y V \otimes \chi \rho \, dy + \int_{T^d} -\nabla_y V \otimes \chi \rho \, dy
\]
\[
= DI - \int_{T^d} -\nabla_y V \otimes \chi \rho \, dy
\]
\[
= DI - D \int_{T^d} \nabla_y \chi \otimes \nabla_y \chi \rho \, dy.
\]
This proves that the effective diffusivity is symmetric. Furthermore, for \(\chi_\xi = \chi \cdot \xi\),
\[
\langle \xi, K\xi \rangle = D|\xi|^2 - D \int_{T^d} |\nabla_y \chi_\xi|^2 \rho \, dy
\]
\[
\leq D|\xi|^2.
\]
This proves depletion. \(\blacksquare\)

### 13.6.2 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 coefficients.
\[
K = D + 2D \int_0^1 \partial_y \chi \rho \, dy + \int_0^1 -\partial_y V \chi \rho \, dy
\]
\[
= D + 2D \int_0^1 \partial_y \chi \rho \, dy + D \int_0^1 \chi \partial_y \rho \, dy
\]
\[
= D + 2D \int_0^1 \partial_y \chi \rho \, dy - D \int_0^1 \partial_y \chi \rho \, dy
\]
\[
= D \int_0^1 (1 + \partial_y \chi) \rho \, dy. \quad (13.6.7)
\]
The cell problem (13.3.1) in one dimension is
\[
D\partial_{yy} \chi - \partial_y V \partial_y \chi = \partial_y V. \quad (13.6.8)
\]
We multiply equation (13.6.8) 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 the above equation from 0 to 1 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

$$c_1 = \frac{1}{Z}, \quad \hat{Z} = \int_0^1 e^{V(y)/\sigma} \, dy.$$ 

We conclude that

$$\partial_y \chi = -1 + \frac{1}{Z} e^{V(y)/D}.$$ 

We substitute this expression into (13.6.7) to obtain

$$K = \frac{D}{\hat{Z}} \int_0^1 (1 + \partial_y \chi(y)) e^{-V(y)/D} \, dy = \frac{D}{\hat{Z}} \int_0^1 e^{V(y)/D} e^{-V(y)/D} \, dy = \frac{D}{\hat{Z}} Z \quad (13.6.9)$$

with

$$Z = \int_0^1 e^{-V(y)/D} \, dy, \quad \hat{Z} = \int_0^1 e^{V(y)/D} \, dy. \quad (13.6.10)$$

The Cauchy-Schwarz inequality shows that $Z \hat{Z} \geq 1$. 
13.6. APPLICATIONS

13.6.3 Divergence–Free Fields

In this section we consider the problem of homogenization for (13.2.3a) in the case where the vector field \( b(y) \) is divergence–free (or incompressible):

\[
\nabla \cdot b(y) = 0. 
\]  
(13.6.11)

The incompressibility of \( b(y) \) simplifies the analysis considerably because the advection operator

\[
\hat{L}_0 = b(y) \cdot \nabla_y,
\]

with periodic boundary conditions is anti–symmetric:

Lemma 13.8. Let \( b(y) \) be a smooth, periodic vector field satisfying (13.6.11). Then for all smooth, periodic functions \( f(y), h(y) \) we have

\[
\int_{\mathbb{T}^d} f(y) \left( b(y) \cdot \nabla_y h(y) \right) dy = - \int_{\mathbb{T}^d} h(y) \left( b(y) \cdot \nabla_y f(y) \right) dy.
\]

In particular,

\[
\int_{\mathbb{T}^d} f(y) \left( b(y) \cdot \nabla_y f(y) \right) dy = 0.
\]  
(13.6.12)

Proof. We use the incompressibility of \( b(y) \), together with the periodicity of \( f(y), h(y) \) and \( b(y) \) to calculate

\[
\int_{\mathbb{T}^d} f(y) \left( b(y) \cdot \nabla_y h(y) \right) dy = \int_{\mathbb{T}^d} f(y) \nabla_y \cdot \left( b(y) h(y) \right) dy
\]

\[
= - \int_{\mathbb{T}^d} \nabla_y f(y) \cdot \left( b(y) h(y) \right) dy
\]

\[
= - \int_{\mathbb{T}^d} h(y) \left( b(y) \cdot \nabla_y f(y) \right) dy.
\]

Equation (13.6.12) follows from the above calculation upon setting \( f = h \). \( \square \)

Using the previous lemma it is easy to prove that the unique invariant measure of the fast process is the Lebesgue measure.

Lemma 13.9. Let \( L_0 \) denote the operator

\[
L_0 = b(y) \cdot \nabla_y + D \Delta_y
\]
with periodic boundary conditions and with \( b(y) \) satisfying (13.6.11). Let \( \mathcal{L}_0^\ast \) denote the \( L^2(\mathbb{T}^d) \)-adjoint of \( \mathcal{L}_0 \). Then the adjoint equation

\[
\mathcal{L}_0^\ast \rho = 0, \quad \int_{\mathbb{T}^d} \rho(y) \, dy = 1,
\]

with periodic boundary conditions on \( \mathbb{T}^d \) has a unique classical solution given by

\[
\rho(y) = 1.
\]

**Proof.** Lemma 13.8 implies that the \( L^2 \)-adjoint of \( \mathcal{L}_0 \) is

\[
\mathcal{L}_0^\ast = -b(y) \cdot \nabla_y + D \Delta_y,
\]

with periodic boundary conditions. Let \( \rho(y) \) be a solution of equation (13.6.13). We multiply the equation by \( \rho(y) \), integrate over \( \mathbb{T}^d \) and use Lemma 13.8 to obtain

\[
\int_{\mathbb{T}^d} |\nabla_y \rho(y)|^2 \, dy = 0,
\]

from which we deduce that \( \rho(y) \) is a constant. Hence, the unique normalized solution of (13.6.13) is given by (13.6.14).

**Remark 13.10.** Given the solution \( \rho(y) = 1 \), uniqueness can also be proved by use of the maximum principle. A consequence of the preceding lemma is that, when

\[
\nabla_y \cdot b(y) = 0,
\]

the solvability condition (13.2.6) 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 (13.6.11). The periodicity of \( \chi(y) \), together with (13.6.14) 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

\[
\mathcal{K} := DI + \int_{\mathbb{T}^d} b(y) \otimes \chi(y) \, dy.
\]
13.6. APPLICATIONS

The effective diffusivity for gradient flows is symmetric. This is not true for divergence–free flows. However, only the symmetric part of \( K \) enters into the homogenized equation. This is because

\[
K : D = \frac{1}{2}(K + K^T) : D
\]

whenever \( D \) is symmetric. Clearly \( \nabla_x \nabla_x u \) is symmetric and so

\[
K : \nabla_x \nabla_x u = \frac{1}{2}(K + K^T) : \nabla_x \nabla_x u.
\]

For this reason we redefine the effective diffusivity to be the symmetric part of \( K \):

\[
K := DI + \frac{1}{2} \int_{T^d} (b(y) \otimes \chi(y) + \chi(y) \otimes b(y)) \, dy. \tag{13.6.16}
\]

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 \( K \).

**Theorem 13.11.** The effective diffusivity \( K \) given by (13.6.16) can be written in the form

\[
K := DI + \left( \int_{T^d} \nabla_y \chi(y) \nabla_y \chi(y)^T \, dy \right). \tag{13.6.17}
\]

**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

\[
-D \int_{T^d} \chi(y) \otimes \Delta_y \chi(y) \, dy - \int_{T^d} \chi(y) \otimes (\nabla_y \chi(y)b(y)) \, dy = \int_{T^d} \chi(y) \otimes b(y) \, dy.
\]

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

\[
D \int_{T^d} \nabla_y \chi(y) \nabla_y \chi(y)^T \, dy + \int_{T^d} (\nabla \chi(y)b(y)) \otimes \chi(y) \, dy = \int_{T^d} \chi(y) \otimes b(y) \, dy. \tag{13.6.18}
\]

We now take the outer-product with \( \chi \) to the right and use the divergence theorem only on the first integral, to obtain

\[
D \int_{T^d} \nabla_y \chi(y) \nabla_y \chi(y)^T \, dy - \int_{T^d} (\nabla \chi(y)b(y)) \otimes \chi(y) \, dy = \int_{T^d} b(y) \otimes \chi(y) \, dy. \tag{13.6.19}
\]
We add equations (13.6.18) and (13.6.19) to obtain:

$$\frac{1}{2} \int_{\mathbb{T}^d} b(y) \otimes \chi(y) + \chi(y) \otimes b(y) \, dy = D \int_{\mathbb{T}^d} \nabla_y \chi(y) \nabla_y \chi(y)^T \, dy.$$ 

Equation (13.6.17) now follows upon substituting the above expression into equation (13.6.16).

Formula (13.6.17) readily yields that the effective diffusivity is always greater than the molecular diffusivity in the following sense.

**Theorem 13.12.** For every vector $\xi \in \mathbb{R}^d$ we have

$$\langle \xi, K \xi \rangle \geq D|\xi|^2,$$

where $| \cdot |$ denotes the Euclidean norm in $\mathbb{R}^d$. Equality holds for all $\xi$ only when $\chi(y) \equiv 0$.

**Proof.** Let $\xi \in \mathbb{R}^d$ and define $\chi_\xi := \chi \cdot \xi$. From (13.6.17) we have

$$\langle \xi, K \xi \rangle := D|\xi|^2 + \int_{\mathbb{T}^d} |\nabla_y \chi_\xi(y)|^2 \, dy \geq D|\xi|^2.$$ 

Clearly equality of the two diffusivities for all $\xi$ implies that $\chi_\xi = 0$ for all $\xi$ implying that $\xi = 0$.

We remark that from the above corollary it immediately follows that the effective diffusivity is a positive definite matrix and, hence, well posedness of the homogenized equation (13.3.3), when (13.6.11) is satisfied, is ensured.

### 13.6.4 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 which can be easily solved by means of Fourier series.

We consider the problem of homogenization for (13.2.3a) in two dimensions for the following velocity field:

$$b(y) = (0, b_2(y_1)),$$ 

(13.6.20)
where $b_2(y_1)$ is a smooth, 1–periodic function. Notice that the velocity field (13.6.20) is incompressible:

$$\nabla \cdot b(y) = \frac{\partial b_1}{\partial y_1} + \frac{\partial b_2}{\partial y_2} = 0.$$ 

The two components of the cell problem satisfy

$$-D \Delta_y \chi_1(y) - b_2(y_1) \frac{\partial \chi_1(y)}{\partial y_2} = 0,$$

(13.6.21a)

$$-D \Delta_y \chi_2(y) - b_2(y_1) \frac{\partial \chi_2(y)}{\partial y_2} = b_2(y_1),$$

(13.6.21b)

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 (13.6.21a) by $\chi_1(y)$, integrate by parts over $\mathbb{T}^d$ and use Lemma 13.8, 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 condition $\langle \chi(y) \rangle = 0$. On the other hand, the fact that the right hand side of (13.6.21b) depends only on $y_1$ implies that 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 uniqueness of solutions to the cell problem implies that it is the only solution. Equation (13.6.21b) becomes:

$$-D \frac{d^2 \chi_2(y_1)}{dy_1^2} = b_2(y_1).$$

(13.6.22)

By (13.6.16) the effective diffusivity $\mathcal{K}$ is the following $2 \times 2$ matrix:

$$\mathcal{K} = \begin{pmatrix}
D + \langle b_1 \chi_1 \rangle & \langle b_2 \chi_1 + b_1 \chi_2 \rangle \\
\langle b_2 \chi_1 + b_1 \chi_2 \rangle & D + \langle b_2 \chi_2 \rangle
\end{pmatrix}$$

$$= \begin{pmatrix}
D & 0 \\
0 & D + \langle b_2(y_1) \chi_2(y_1) \rangle
\end{pmatrix}.$$ 

From this we deduce that

$$\mathcal{K}_{22} = D + \frac{1}{D} \|b_2(y_1)\|_{H^{-1}}^2.$$ 

(13.6.23)
The $H^{-1}$ norm of a real–valued, periodic function with period 1 can be expressed in terms of Fourier series:

$$\|f(y)\|_{H^{-1}}^2 = \frac{1}{2\pi^2} \sum_{k=1}^{\infty} \frac{|f_k|^2}{|k|^2}.$$ 

We will derive this result in due course. First we give a quick justification. We have that

$$\langle b_2(y_1)\chi_2(y_1) \rangle = \langle D \frac{d^2\chi_2(y_1)}{dy_1^2} \chi_2(y_1) \rangle = D \|\chi_2(y_1)\|_{H^1}^2.$$ 

From this expression and from (13.6.22) it follows that

$$\langle b_2(y_1)\chi_2(y_1) \rangle = \frac{1}{D} \|b_2(y_1)\|_{H^{-1}}^2.$$ 

Notice that remarkable fact that the effective diffusion coefficient blows up as the original diffusion coefficient tends to zero.

Hence, in order to compute the effective diffusivity we need to calculate the average of $b_2\chi_2$ over the unit cell. Since this expression depends only on $y_1$ we have:

$$\langle b_2(y_1)\chi_2(y_1) \rangle = \int_0^1 b_2(y_1)\chi_2(y_1) \, dy_1.$$ 

Now we need to solve equation (13.6.22), subject to periodic boundary conditions. We will accomplish this by using the Fourier series representation of periodic functions. In the remainder of this example we will drop all indices and subscripts: $b(y) := b_2(y_1)$, $\chi(y) := \chi_2(y_1)$.

Since the velocity field $b_2(y_1)$ is smooth and 1–periodic, it can be expressed through its Fourier series expansion

$$b(y) = \sum_{k=-\infty}^{\infty} b_k e^{2\pi iky},$$ 

(13.6.24)

where

$$b_k = \int_0^1 b(y)e^{-2\pi iky} \, dy.$$ 

(13.6.25)

It is straightforward to check that the facts that $b(y)$ has mean zero and is a real valued function imply, respectively, that $b_0 = 0$ and that $\overline{b_{-k}} = b_k$ (We will refer to
this as the reality condition. Now, the solution $\chi(y)$ of (13.6.22) is an element of $H^1(0, 1)$ and hence it admits the Fourier representation

$$\chi(y) = \sum_{k \neq 0} \chi_k e^{2\pi i ky}, \quad (13.6.26)$$

where we have already imposed the condition $\langle \chi(y) \rangle = 0$. Furthermore, since $\chi(y)$ is a real valued function, we have that $\chi_{-k} = \overline{\chi_k}$. We use (13.6.26) to calculate the second order derivative of $\chi(y)$:

$$\frac{d^2 \chi(y)}{dy^2} = -4\pi^2 \sum_{k \neq 0} |k|^2 \chi_k e^{2\pi i ky}.$$ 

We use this expression in (13.6.22), together with the Fourier expansion of $b(y)$ and the orthonormality of the Fourier basis to obtain an infinite set of equations for each Fourier coefficient of $\chi$:

$$4\pi^2 D |k|^2 \chi_k = b_k, \quad k \in \mathbb{Z}/\{0\}.$$ 

Consequently

$$\chi_k = \frac{b_k}{4\pi^2 D |k|^2}$$

and

$$\chi(y) = \sum_{k \neq 0} \frac{b_k}{4\pi^2 D |k|^2} e^{2\pi i ky}. \quad (13.6.27)$$

Formula (13.6.27) provides us with the solution of the cell problem (13.6.22). We
use this formula, together with (13.6.24) to compute:
\[
\langle b(y)\chi(y) \rangle = \int_0^1 b(y)\chi(y) \, dy \\
= \int_0^1 \left( \sum_{k \neq 0} \frac{b_k}{4\pi^2 D|k|^2} e^{2\pi i ky} \right) \left( \sum_{\ell \neq 0} b_\ell e^{2\pi i \ell y} \right) \, dy \\
= \sum_{k,\ell \neq 0} \left( \frac{b_k b_\ell}{4\pi^2 D|k|^2} \int_0^1 e^{2\pi i (k+\ell)y} \, dy \right) \\
= \sum_{k,\ell \neq 0} \left( \frac{b_k b_\ell}{4\pi^2 D|k|^2} \delta_{k,-\ell} \right) \\
= \sum_{k \neq 0} \frac{|b_k|^2}{4\pi^2 D|k|^2} \\
= \frac{1}{2\pi^2 D} \sum_{k=1}^\infty \frac{|b_k|^2}{|k|^2}.
\]
In deriving the penultimate equality we used the reality condition for \( b(y) \). We combine the above calculation with the formula for the effective diffusivity to obtain equation (13.6.23):

\[
K_{22} = D + \frac{1}{D} \| b_2(y_1) \|_{H^{-1}}^2.
\]

### 13.7 The Connection to SDE

Equation (13.2.1) is the backward Kolmogorov equation associated with the SDE
\[
\frac{dx}{dt} = b(x) + \sqrt{2D} \, dW, \quad (13.7.1)
\]
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 re-scaling, used to derive (13.2.3) from (13.2.1), to the SDE. That is we re-label according to
\[
x \to x/\varepsilon, \quad t \to t/\varepsilon^2
\]
giving the SDE
\[
\frac{dx}{dt} = \frac{1}{\varepsilon} b\left( \frac{x}{\varepsilon} \right) + \sqrt{2D} \frac{dW}{dt}. \quad (13.7.2)
\]
13.8. DISCUSSION AND BIBLIOGRAPHY

(Recall Remark 6.3 regarding the behaviour of white noise under time-rescaling).

If we introduce the variable \( y = x/\varepsilon \) then we can write this SDE in the form

\[
\frac{dx}{dt} = \frac{1}{\varepsilon} b(y) + \sqrt{2D} \frac{dW}{dt},
\]

\[
\frac{dy}{dt} = \frac{1}{\varepsilon^2} b(y) + \frac{1}{\varepsilon} \sqrt{2D} \frac{dW}{dt}.
\]

This is precisely in 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 (\( U = V \)). This has the effect of changing the operator \( L_1 \) in that chapter, so that the results derived there do not apply directly. They can, however, be extended to the study of correlated noise. Notice that the centering condition (13.2.6) is precisely the condition (11.2.2) since \( \rho \) is the stationary solution of the 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 pure Brownian motion. Interpreted in terms of the SDE we obtain the following result.

**Result 13.13.** Assume that (13.2.6) holds. For \( \varepsilon \ll 1 \) and \( t = O(1) \), \( x \) solving the SDE (13.7.3) can be approximated by \( X \) solving

\[
\frac{dX}{dt} = \sqrt{2K} \frac{dW}{dt}
\]

where the matrix \( K \) is given by (13.3.2).

### 13.8 Discussion and Bibliography

The problem of homogenization for second order parabolic PDE and its connection to the study of the long time asymptotics of solutions to SDE is studied in [17, Ch. 3]. References to the earlier literature can be found in there. See also [112]. Periodic homogenization for gradient flows is discussed in e.g. [112, 124, 152, 54]. Stochastic differential equations 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, e.g. in solid state physics and biology. See [128, Ch. 11], [126] and the references therein. Multiscale techniques were applied to this problem in [123]. On the other hand, the SDE (13.7.1) with
CHAPTER 13. HOMOGENIZATION FOR PARABOLIC EQUATIONS

divergence–free drift occur naturally in the modeling of diffusion processes in fluids. Homogenization for periodic, incompressible flows is a part of the theory of turbulent diffusion e.g. [94, Ch. 2]. In this context an interesting question concerns the dependence of the effective diffusivity on the molecular diffusion. See [94, Ch. 2] and the references therein.

It is possible to derive a homogenized equation even when the centering condition (13.2.6) is not satisfied. In this case it is necessary to use a frame co–moving with the mean flow

\[ F = \int_{T^d} b(y) \rho(y) \, dy. \]  

(13.8.1)

Then, it is possible to derive a homogenized equation of the form (13.3.3) for the rescaled field

\[ u^\varepsilon(x, t) = u \left( \frac{x}{\varepsilon} - \frac{F t}{\varepsilon^2}, \frac{t}{\varepsilon^2} \right). \]

The effective diffusivity is given by the formula

\[ K = DI + 2D \int_{T^d} \nabla_y \chi(y)^T \rho(y) \, dy + \int_{T^d} (b(y) - F) \otimes \chi(y) \rho(y) \, dy, \]  

(13.8.2)

The cell problem (13.3.1) is also modified:

\[ -L_0 \chi = b - F. \]  

(13.8.3)

It was proved in section 13.6.1 that for gradient flows the diffusion is always depleted. In fact, a 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

\[ K^\varepsilon = c_1 e^{-c_2/D}, \quad D \ll 1, \]

where

\[ K^\varepsilon := \sum_{i,j=1}^d K_{ij} \xi_i \xi_j. \]

See [26] and the references therein.

On the other hand, the effective diffusion coefficient can become arbitrarily large when a constant external force is added to the gradient drift, [127, 133]. The presence of a mean flow plays an important role in the case of divergence–free flows, in particular in connection to the small \( D \)–asymptotics of the effective diffusivity. See [95, 105, 20, 137, 18]
The fact that the effective diffusivity along the direction of the shear is inversely proportional to the molecular diffusivity, formula (13.6.23) was discovered by G.I. Taylor in [146], without any appeal to homogenization theory. This phenomenon is usually called Taylor dispersion. See also [6]. In deriving the result we used 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. In this section we will only perform formal calculations. The existence, uniqueness and regularity of solutions to the cell problem will be examined in the next section. For various properties of Fourier series we refer the reader to [61, Ch. 3].

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

$$V(y) = \begin{cases} a_1 & : y \in [0, \frac{1}{2}], \\ a_2 & : y \in (\frac{1}{2}, 1], \end{cases}$$

where $a_1, a_2$ are positive constants. Calculate the effective diffusivity. How does it depend on $D$?

3. Same as in the previous exercise for the potential

$$V(y) = \begin{cases} y & : y \in [0, \frac{1}{2}], \\ 1 - y & : y \in (\frac{1}{2}, 1], \end{cases}$$

4. Consider the problem of homogenization for the reaction–advection–diffusion equation

$$\frac{\partial u^\varepsilon}{\partial t} = \frac{1}{\varepsilon} b\left(\frac{x}{\varepsilon}\right) \cdot \nabla u^\varepsilon + \Delta u^\varepsilon + \frac{1}{\varepsilon} c\left(\frac{x}{\varepsilon}\right) u^\varepsilon, \quad (13.9.1)$$

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. Suppose that the reaction term is non-linear, i.e. 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?

5. Consider the problem of homogenization for the PDE

$$\frac{\partial u^\varepsilon}{\partial t} = \left( b_1(x) + \frac{1}{\varepsilon} b_2\left(\frac{x}{\varepsilon}\right) \right) \cdot \nabla u^\varepsilon + \Delta u^\varepsilon, \quad (13.9.2)$$

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

$$\frac{\partial u}{\partial t} = B \cdot \nabla u + \sum_{i,j=1}^{d} K_{ij} \frac{\partial^2 u}{\partial x_i \partial x_j} \quad (13.9.3)$$

and derive the cell problem(s) and the formulas for the homogenized coefficients $B$ and $K$.

6. Consider the problem of homogenization for the PDE (13.9.2) in the case where

$$b_1(x) = -\nabla V(x), \quad \text{and} \quad b_2(y) = -\nabla_y p(y),$$

where $p(y)$ is periodic.

(a) Show that in this case there exists a symmetric matrix $K$ such that

$$K_{ij} = D K_{ij}, \quad B_i = -\sum_{j=1}^{d} K_{ij} \frac{\partial V}{\partial x_j}.$$
(b) Let
\[ \mathcal{L} := B \cdot \nabla + \sum_{i,j=1}^{d} K_{ij} \frac{\partial^2}{\partial x_i \partial x_j}. \]

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
\[ \mathcal{L}^* \rho = 0. \]

7. Consider the problem of homogenization for the following PDE
\[ \frac{\partial u^\varepsilon}{\partial t} = \frac{1}{\varepsilon} b \left( \frac{x}{\varepsilon} \right) \cdot \nabla u^\varepsilon + \sum_{i,j=1}^{d} a_{ij} \left( \frac{x}{\varepsilon} \right) \frac{\partial^2 u^\varepsilon}{\partial x_i \partial x_j}, \tag{13.9.4} \]
where 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 nonnegative definite.

8. 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
\[ a_0 = 0 \]
and that
\[ a_{-k} = \overline{a_k}. \]
Chapter 14

Homogenization for Linear Transport PDE

14.1 Introduction

In this chapter we investigate the long time behavior of solutions to the linear advection equation. The techniques we employ are often referred to as homogenization techniques in the literature, and so we retain this term. However in terms of our classification in the introductory Chapter 1 the methods are actually more closely aligned to averaging methods for ODE, as discussed in Chapter 10. This connection will be made explicit in the following.

14.2 Full Equations

We study the long time behavior of solutions to the linear advection equation for steady periodic velocity fields $a$:

$$\frac{\partial u(x,t)}{\partial t} + a(x) \cdot \nabla u(x,t) = 0 \quad \text{for } (x,t) \in \mathbb{R}^d \times \mathbb{R}^+, \quad (14.2.1a)$$

$$u(x,0) = u_0(x). \quad (14.2.1b)$$

This is (13.2.1) in the case $D = 0$. Equation (14.2.1) is the backward Liouville equation corresponding to the dynamical system

$$\frac{dx}{dt} = a(x). \quad (14.2.2)$$

See (4.3.4).
As in Chapter 13 we study the case where
\[ u_0(x) = f(\varepsilon x), \]
and rescale the equation in both space and time in order to understand the behavior of solutions to equation (14.2.1), and consequently to (14.2.2), at length and time scales which are long when compared to those of the fluid velocity \( a(x) \). In this setting, the small parameter in the problem is the ratio between the characteristic length (time) scale of the velocity field – it’s period – and the largest length (time) scale of the problem – the one at which we are looking for a homogenized 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
\[
\begin{align*}
x & \rightarrow \varepsilon^{-1} x, \\
t & \rightarrow \varepsilon^{-1} t.
\end{align*}
\]
(14.2.3)
Such a transformation is natural since the transport PDE (14.2.1a) is of first order in both space and time.

The initial value problem that we wish to investigate then becomes:

\[
\begin{align*}
\frac{\partial u^\varepsilon(x,t)}{\partial t} + a\left(\frac{x}{\varepsilon}\right) \cdot \nabla u^\varepsilon(x,t) &= 0 \quad \text{for } (x,t) \in \mathbb{R}^d \times \mathbb{R}^+, \\
u^\varepsilon(x,0) &= f(x) \quad \text{for } x \in \mathbb{R}^d.
\end{align*}
\]
(14.2.4a, 14.2.4b)

\textbf{14.3 Simplified Equations}

Let
\[ L_0 = a(y) \cdot \nabla y \]
with periodic boundary conditions. We assume for the moment that there are no non-trivial functions in the null space \( \mathcal{N} \) of \( L_0 \):
\[ \mathcal{N}(L_0) = \{ \text{constants in } y \}. \]
(14.3.2)

From Chapter 4 we know that this is an ergodicity assumption on the ODE with vector field \( a \). Note that if \( a \) is divergence-free (the velocity field is incompressible) then \( L \) is skew-symmetric and so we deduce that
\[ \mathcal{N}(L_0^*) = \{ \text{constants in } y \}. \]
(14.3.3)

Under this assumption we have the following:
Result 14.1. Let $a$ be a periodic divergence-free vector field. Then, for $\varepsilon \ll 1$ and $t = O(1)$, the solution of (13.2.3) $u^\varepsilon(x, t)$ is approximated by $u(x, t)$, the solution of the homogenized equation:

$$\frac{\partial u(x, t)}{\partial t} + \tilde{a} \cdot \nabla_x u(x, t) = 0, \quad \tilde{a} := \left( \int_Y a(y) \, dy \right),$$

together with the same initial condition as for $u^\varepsilon$.

### 14.4 Derivation

We use the method of multiple scales as introduced in the two preceding chapters. To this end, we look for a solution in the form of a two-scale expansion:

$$u^\varepsilon(x, t) = u_0 \left( x, \frac{x}{\varepsilon}, t \right) + \varepsilon u_1 \left( x, \frac{x}{\varepsilon}, t \right) + \ldots . \tag{14.4.1}$$

We assume that all terms in the expansion $u_j(x, y, t), \ j = 0, 1, \ldots$ are 1–periodic in $y$ and treat $x$ and $y := \frac{x}{\varepsilon}$ as independent variables. We substitute (14.4.1) into equation (14.2.4a), use the assumed independence of $x$ and $y$ and collect equal powers of $\varepsilon$ to obtain the following set of equations

$$O(1/\varepsilon) \quad \mathcal{L}_0 u_0 = 0, \quad u_0 \text{ is } 1\text{–periodic}, \tag{14.4.2a}$$

$$O(1) \quad \mathcal{L}_0 u_1 = -\mathcal{L}_1 u_0, \quad u_1 \text{ is } 1\text{–periodic}, \tag{14.4.2b}$$

where

$$\mathcal{L}_0 = a(y) \cdot \nabla_y \quad \text{and} \quad \mathcal{L}_1 = \frac{\partial}{\partial t} + a(y) \cdot \nabla_x .$$

Under assumption (14.3.2) we can now complete the homogenization procedure. From the first equation in (14.4.2) we get 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:

$$\mathcal{L}_1 u_0 = \frac{\partial u(x, t)}{\partial t} + a(y) \cdot \nabla_x u(x, t).$$

On the other hand, an integration by parts, together with the fact that $a(y)$ is divergence–free, gives:

$$\int_Y \mathcal{L}_0 u_1 \, dy = \int_Y a(y) \cdot \nabla_y u_1 \, dy = 0.$$
This implies that a necessary condition for the existence of solutions to the equation

\[ \mathcal{L}_0 u_1 = f \]

is that the right hand side \( f \) averages to 0. By (14.3.2), this is the analogue of the Fredholm alternative for this advection equation with periodic boundary conditions. Necessity follows from the above calculation; sufficiency follows from the ergodic theorem 4.7. Applying this condition to the second equation in (14.4.2) gives

\[ 0 = \frac{\partial u(x, t)}{\partial t} + \left( \int_Y a(y) \, dy \right) \cdot \nabla_x u(x, t). \tag{14.4.3} \]

We have thus obtained the desired homogenized equation:

\[ \frac{\partial u(x, t)}{\partial t} + \bar{\alpha} \cdot \nabla_x u(x, t) = 0, \quad \bar{\alpha} := \left( \int_Y a(y) \, dy \right), \]

together with the same initial conditions as those for \( u^\varepsilon \).

### 14.5 The Connection to ODE

Recall from Chapter 4 that the solution of (14.2.4) is given by

\[ u(x, t) = f(\varphi^t(x)) \]

where \( \varphi^t(x) \) solves ODE

\[
\frac{d}{dt} \varphi^t(x) = a \left( \frac{\varphi^t(x)}{\varepsilon} \right), \\
\varphi^0(x) = x.
\]

Result 14.1 shows that, when (14.3.2) holds, this equation is well approximated by

\[ \bar{\varphi}^t(x) = \bar{\alpha}t + x, \]

the solution of

\[
\frac{d}{dt} \bar{\varphi}^t(x) = \bar{\alpha}, \\
\bar{\varphi}^0(x) = x.
\]
Another way to see this result is as follows. Let \( x = \varphi^t(x_0) \) and \( y = x/\varepsilon \). Then

\[
\begin{align*}
\frac{dx}{dt} &= a(y), \\
\frac{dy}{dt} &= \frac{1}{\varepsilon} a(y).
\end{align*}
\]

Under the ergodic hypothesis the fast process \( y \) has uniform invariant measure on the torus \( T^d \). Thus the averaging Result 10.1 gives that \( x \) is well approximated by the solution of the equation

\[
\frac{dX}{dt} = \bar{\alpha}.
\]

One can describe Result 14.1 the following way: under assumption (14.3.2) we have that

\[
\lim_{\varepsilon \to 0} \varepsilon \int_0^{t/\varepsilon} a(x(s)) \, ds = \bar{\alpha} t,
\]

where \( x(t) \) is the solution of the ODE

\[
\frac{dx}{dt} = a(x), \quad x(0) = x_0.
\]

This is another way to see that Result 14.1 is a direct consequence of the ergodicity of the flow generated by \( a(y) \) on the unit torus.

### 14.6 The One–Dimensional Case

Consider the rescaled transport equation (14.2.4a) in one dimension:

\[
\frac{\partial u}{\partial t} + a(x) \frac{\partial u}{\partial x} = 0 \quad \text{for } (x, t) \in \mathbb{R} \times \mathbb{R}^+,
\]

\[
u(x, 0) = u_0(x).
\]

We assume that \( a(y) \) is a positive, smooth, 1–periodic function. Clearly, \( a(y) \) is not divergence–free (excluding the trivial case where \( a(y) \) is a constant function) and it is not expected that Result 14.1 holds. Indeed, the stationary forward Liouville equation

\[
\mathcal{L}_0^* \rho = 0, \quad \rho > 0, \ 1–\text{periodic}
\]
together with the normalization condition
\[ \int_0^1 \rho(y) \, dy = 1, \]
no longer has constants as solution. The unique normalized solution of (14.6.2) is the probability density
\[ \rho(y) = \frac{C}{a(y)}, \quad C = \langle a(y)^{-1} \rangle^{-1}; \]
here we have used the notation \( \langle \cdot \rangle \) to denote averaging over \([0, 1]\), as in Chapter 12.

We can use the method of multiple scales to homogenize equation (14.6.1). The solvability condition (14.4.3) gives
\[ \frac{\partial u}{\partial t} + \int_0^1 L_1 u \rho(y) \, dy = 0, \]
from which we obtain the homogenized equation
\[ \frac{\partial u}{\partial t} + a^* \frac{\partial u}{\partial x} = 0, \quad (14.6.3) \]
with the same initial conditions as in (14.6.1b) and with
\[ a^* = \langle a(y)^{-1} \rangle^{-1}. \]
Notice that, in contrast to the ergodic divergence–free case, it is the harmonic average of the velocity field that appears in the homogenized equation (14.6.3).

One can also derive the homogenized equation (14.6.3) using the method of characteristics. To see this, consider the equation
\[ \frac{dx}{dt} = a \left( \frac{x}{\epsilon} \right) \]
in one dimension, and under the same assumptions as before. If we set \( y = x/\epsilon \) then it is straightforward to show that
\[ \frac{dy}{dt} = \frac{1}{\epsilon} a(y) \]
so that, if we define \( T \) by
\[ T = \int_0^1 \frac{1}{a(z)} \, dz \]
14.7. THE TWO–DIMENSIONAL CASE. KOLMOGOROV’S THEOREM

then

\[ y(n \varepsilon T) = \frac{x(0)}{\varepsilon} + n. \]

Hence

\[ x(n \varepsilon T) = x(0) + n \varepsilon. \]

It follows from continuity that \( x(t) \to X(t) \) where

\[ X(t) = x(0) + \frac{t}{T}. \]

This limiting variable satisfies the homogenized equation

\[ \frac{dX}{dt} = \frac{1}{T} = a^*. \]

14.7 The Two–Dimensional Case. Kolmogorov’s Theorem

Unfortunately, even for divergence-free fields, the ergodic hypothesis is often not satisfied. Consider the equation (14.4.2a):

\[ \mathcal{L}_0 u_0 = a(y) \cdot \nabla_y u_0 (x, y, t) = 0. \quad (14.7.1) \]

This equation rarely implies that the first term in the expansion is independent of \( y \): 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

\[ a(y) = (\sin(2\pi y_2), \sin(2\pi y_1)). \]

It is easy to check that the function

\[ f(y) = \cos(2\pi y_1) - \cos(2\pi y_2) \]

solves equation (14.7.1). Consequently, the null space of \( \mathcal{L}_0 \) depends on the velocity field \( a(y) \) and, it does not consist, in general, merely of constants in \( y \). This implies that we cannot carry out the homogenization procedure using the method of multiple scales: no useful information is contained in equations (14.4.2).

It is natural to ask whether there is a way of deciding whether a given divergence-free velocity field on \( \mathbb{T}^2 \) is ergodic or not. This is indeed possible it two dimensions. We have the following theorem, which is due to Kolmogorov.
Theorem 14.2. Let \( a(y) : \mathbb{T}^2 \rightarrow \mathbb{R}^2 \) be a smooth divergence–free velocity field with no stagnation points:

\[
a_1^2(y) + a_2^2(y) > 0, \quad \forall y \in \mathbb{T}^2.
\]

Let \( \bar{a}_i, i = 1, 2 \) denote the average of the \( i \)th component of the velocity field over \( \mathbb{T}^2 \) and define the rotation number as

\[
\gamma = \frac{\bar{a}_1}{\bar{a}_2}.
\]

Assume that \( \gamma \) is irrational and normally approximated by rationals. Then \( a(y) \) is ergodic.

Remark 14.3. A number \( a \in \mathbb{R} \) is said to be normally approximated by rationals if there exist \( C, \varepsilon \) such that for every integer \( q \)

\[
\min_p \left| a - \frac{p}{q} \right| \geq \frac{C}{q^2 + \varepsilon},
\]

where the minimum is taken over the set of all integers.

14.8 Discussion and Bibliography

The problem of homogenization for linear transport equations has been studied by many authors. See for example [36, 70, 145, 24].

We remark that the calculation of the effective velocity \( \bar{a} \) does not require the solution of a Poisson equation. This is a common characteristic of first order perturbation theory (we only needed to go up to terms of \( O(\varepsilon) \) in the expansion in order to obtain the homogenized equation). 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 method of multiple scales enabled us to obtain the homogenized equation for the linear transport equation (14.2.4a) 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 don’t really have a solvability condition which would enable us to obtain a homogenized equation. In order to study the problem for general velocity fields, not necessarily ergodic, we need to use the method of two–scale convergence. This will be done in Chapter 21.
A proof of Theorem 14.2 can be found in [135]. A similar theorem holds for velocity fields with an invariant measure other than the Lebesgue measure on $\mathbb{T}^2$. See [145].

The example studied in Section 14.6 was communicated to the authors by T. Hou. It can also be found in [38, 145].

14.9 Exercises

1. How does the dynamics in the ODE studied in Section 14.6 change if $a$ is allowed to change sign?

2. Consider the equation

$$\frac{dx}{dt} = a\left(\frac{x}{\varepsilon}\right)b\left(\frac{t}{\varepsilon^n}\right)$$

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 homogenization for (7.7.1) with a smooth periodic velocity field $a(y) : \mathbb{T}^2 \mapsto \mathbb{R}^2$ of the form

$$a(y) = (a_1(y_2), 0).$$

4. Study the problem of homogenization for (7.7.1) with a velocity field $a(y) : \mathbb{T}^2 \mapsto \mathbb{R}^2$ of the form

$$a(y) = b(y)(0, \gamma),$$

where $b(y)$ is a smooth, 1–periodic scalar function and $\gamma \in \mathbb{R}$. 
Part III

Theory
Chapter 15

Invariant Manifolds for ODE:
The Convergence Theorem

15.1 The Theorem

In this section we describe a rigorous theory substantiating the perturbation expansions for invariant manifolds in Chapter 8. We study the equations

\[ \begin{align*}
\frac{dx}{dt} &= f(x, y), \\
\frac{dy}{dt} &= \frac{1}{\varepsilon} g(x, y),
\end{align*} \tag{15.1.1} \]

for \( \varepsilon \ll 1 \) and \( x \in \mathbb{R}^l, y \in \mathbb{R}^{d-l} \). We assume that the dynamics for \( y \) with \( x \) frozen has a unique exponentially attracting fixed point, for all \( x \). Specifically we assume that there exists \( \eta : \mathbb{R}^l \to \mathbb{R}^{d-l} \) and \( \alpha > 0 \) such that

\[ g(\xi, \eta(\xi)) = 0 \quad \forall \xi \in \mathbb{R}^l \]

\[ \langle g(x, y_1) - g(x, y_2), y_1 - y_2 \rangle \leq -\alpha |y_1 - y_2|^2 \quad \forall x \in \mathbb{R}^l, \forall y_1, y_2 \in \mathbb{R}^{d-l}. \]

The dynamics with \( x \) frozen at \( \xi \) satisfies

\[ \frac{d}{dt} \varphi^t_\xi(y) = g(\xi, \varphi^t_\xi(y)), \quad \varphi^0_\xi(y) = y. \tag{15.1.2} \]

Our assumptions on \( \eta \) and \( g \) imply the following exponential convergence of \( \varphi^t_\xi(y) \) to its globally attracting fixed point \( \eta(\xi) \).

**Lemma 15.1.** For all \( y \in \mathbb{R}^{d-l} \)

\[ |\varphi^t_\xi(y) - \eta(\xi)| \leq e^{-\alpha t} |y - \eta(\xi)|. \]
Proof. Since $\eta(\xi)$ is time-independent we have
\[
\frac{d}{dt}\varphi^\xi_t(y) = g(\xi, \varphi^\xi_t(y)),
\]
\[
\frac{d}{dt}\eta(\xi) = g(\xi, \eta(\xi)).
\]
Hence
\[
\frac{1}{2} \frac{d}{dt} |\varphi^\xi_t(y) - \eta(\xi)|^2 = \langle g(\xi, \varphi^\xi_t(y)) - g(\xi, \eta(\xi)), \varphi^\xi_t(y) - \eta(\xi) \rangle 
\leq -\alpha |\varphi^\xi_t(y) - \eta(\xi)|^2.
\]
The result follows from the differential form of the Gronwall Lemma 4.3.

In essence we wish to prove a result like this when $x$ is no longer frozen at $\xi$, but rather evolves on its own time-scale of $O(1)$. We make the following standing assumptions. These simplify the analysis and make the ideas of the proof of the basic result clearer; however they can all be weakened in various different ways. The assumptions are the existence of a constant $C > 0$ such that:
\[
|f(x, y)| \leq C \quad \forall (x, y) \in \mathbb{R}^d,
\]
\[
|\nabla_x f(x, y)| \leq C \quad \forall (x, y) \in \mathbb{R}^d,
\]
\[
|\nabla_y f(x, y)| \leq C \quad \forall (x, y) \in \mathbb{R}^d,
\]
\[
|\eta(x)| \leq C \quad \forall x \in \mathbb{R}^l,
\]
\[
|\nabla \eta(x)| \leq C \quad \forall x \in \mathbb{R}^l.
\]

With these assumptions we prove that $x$ is close to $X$ solving
\[
\frac{dX}{dt} = f(X, \eta(X)). \tag{15.1.3}
\]

**Theorem 15.2.** Assume that $x(0) = X(0)$. Then there are constants $K, c > 0$ such that $x(t)$ solving (15.1.1) and $X(t)$ solving (15.1.3) satisfy
\[
|x(t) - X(t)|^2 \leq ce^{Kt}(\varepsilon |y(0) - \eta(x(0))|^2 + \varepsilon^2).
\]

Note that the error is of size $\sqrt{\varepsilon}$ for times of $O(1)$. However this can be reduced to size $\varepsilon$ if the initial deviation $y(0) - \eta(x(0))$ is of size $\sqrt{\varepsilon}$.

---

1In the first and fourth items in this list the norms are standard vector norms; in the second, third and fifth they are operator norms. All are Euclidian.
15.2 The Proof

Define \( z(t) \) by \( y(t) = \eta(x(t)) + z(t) \). Then

\[
\frac{dz}{dt} = \frac{dy}{dt} - \nabla \eta(x) \frac{dx}{dt}
\]

\[
= \frac{1}{\varepsilon} g(x, \eta(x) + z) - \nabla \eta(x) f(x, \eta(x) + z)
\]

\[
= \frac{1}{\varepsilon} \left( g(x, \eta(x) + z) - g(x, \eta(x)) \right) - \nabla \eta(x) f(x, \eta(x) + z).
\]

Now

\[
\langle g(x, \eta(x) + z) - g(x, \eta(x)), z \rangle \leq -\alpha |z|^2.
\]

Hence, choosing \( \delta^2 = \varepsilon/\alpha \),

\[
\frac{1}{2} \frac{d}{dt} |z|^2 \leq -\frac{\alpha}{\varepsilon} |z|^2 + C^2|z|
\]

\[
\leq -\frac{\alpha}{\varepsilon} |z|^2 + \frac{\delta^2}{2} C^4 + \frac{1}{2} \frac{|z|^2}{\delta^2}
\]

\[
\leq -\frac{\alpha}{2\varepsilon} |z|^2 + \frac{\varepsilon}{2\alpha} C^4.
\]

Hence

\[
\frac{d}{dt} \left( e^{\frac{\alpha}{2}\varepsilon t} |z|^2 \right) \leq \left( e^{\frac{\alpha}{2}\varepsilon t} \right) \frac{\varepsilon}{\alpha} C^4.
\]

Integrating gives

\[
|z(t)|^2 \leq e^{-\frac{\alpha}{2}\varepsilon t} |z(0)|^2 + \left( 1 - e^{-\frac{\alpha}{2}\varepsilon t} \right) \frac{\varepsilon^2 C^4}{\alpha^2}.
\]

(15.2.1)

Now

\[
\frac{dX}{dt} = f(X, \eta(X))
\]

\[
\frac{dx}{dt} = f(x, \eta(x) + z).
\]

Hence

\[
\frac{d}{dt} (x - X) = f(x, \eta(x) + z) - f(X, \eta(X))
\]

\[
= f(x, \eta(x) + z) - f(x, \eta(X)) + f(x, \eta(X)) - f(X, \eta(X)).
\]
Thus
\[ \frac{1}{2} \frac{d}{dt} |x - X|^2 \leq C|\eta(x) - \eta(X) + z||x - X| + C|x - X|^2 \]
\[ \leq \left( C^2 + C \right) |x - X|^2 + C|z||x - X|. \]

It follows that
\[ \frac{d}{dt} |x - X|^2 \leq \left( 3C^2 + 2C \right) |x - X|^2 + |z|^2. \]

Letting \( K = 3C^2 + 2C \), and using the bound (15.2.1) for \( |z(t)|^2 \) we obtain
\[ \frac{d}{dt} \left( e^{-Kt} |x - X|^2 \right) \leq e^{-\frac{K}{2}t} |z(0)|^2 + e^{-Kt} \frac{\varepsilon^2}{\alpha^2} C^4. \]

Integrating and using \( x(0) = X(0) \) gives the desired result. \( \square \)

15.3 Discussion and Bibliography

Theorem 15.2 shows that \( x \) from the full equations (15.1.1) remains close to \( X \) solving the reduced equations (15.1.3) over time-scales which are of the order \( \ln(\varepsilon^{-1}) \). On longer time-scales the individual solutions can diverge, because of the exponential separation of trajectories which may be present in any dynamical system. Notice, however, that (15.2.1) shows that
\[ \limsup_{t \to \infty} |y(t) - \eta(x(t))| \leq \frac{C^4}{\alpha^2} \varepsilon^2 \]
suggesting that \( y(t) \) is approximately slaved to \( x(t) \), via \( y = \eta(x) \), for arbitrary time-intervals. There are results concerning the approximation of \( x \) over arbitrarily long times. These long-time approximation results are built on making rigorous the construction of an invariant manifold as described in Chapter 8. The idea is as follows. Consider the equations
\[ \frac{dx}{dt} = f(x, \eta(x) + z), \]
\[ \frac{dz}{dt} = \frac{1}{\varepsilon} g(x, \eta(x) + z) - \nabla \eta(x)f(x, \eta(x) + z). \]

Notice that \( y = \eta(x) + z \). Using the fact that \( \nabla_y g(x, \eta(x)) \) is negative-definite it is possible to prove the existence of an invariant manifold for \( z \) with the form
\[ z = \varepsilon \eta_1(x; \varepsilon) \]
with \( \eta_1 \) bounded uniformly in \( \varepsilon \). To be precise, the equations for \( x \) and \( z \) started with initial conditions \( z(0) = \varepsilon \eta_1(x(0); \varepsilon) \) will satisfy \( z(t) = \varepsilon \eta_1(x(t); \varepsilon) \) for all positive time. Furthermore the manifold is attracting so that \(|z(t) - \varepsilon \eta_1(x(t); \varepsilon)| \to 0 \) as \( t \to \infty \). Thus we have an attractive invariant manifold for \( y \) with the form

\[
 y = \eta(x) + \varepsilon \eta_1(x; \varepsilon).
\]

The existence and uniqueness of invariant manifolds can be proved by a variety of techniques, predominantly the Lyapunov-Perron approach (Hale [63], Temam [148]) and the Hadamard graph transform (Wells [153]). Important work in this area is due to Fenichel [46, 47] who sets up a rather general construction of normally hyperbolic invariant manifolds. The book of Carr [27] has a clear introduction to the Lyapunov-Perron approach to proving existence of invariant manifolds.

15.4 Exercises

1. Show that, under the assumptions on \( g \) stated at the beginning of the chapter, \( \varphi^t_{\xi} : \mathbb{R}^{d-l} \to \mathbb{R}^{d-l} \) is a contraction mapping for any \( t > 0 \). What is its fixed point?

2. Prove a result similar to Theorem 15.2 but removing the assumption that \( \eta \) and \( f \) are globally bounded; use instead linear growth assumptions on \( \eta \) and \( f \).

3. Consider the equations

\[
\begin{align*}
\frac{dx}{dt} &= Ax + \varepsilon f_0(x, y), \\
\frac{dy}{dt} &= -\frac{1}{\varepsilon} By + g_0(x, y),
\end{align*}
\]

(15.4.1)

for \( \varepsilon \ll 1 \) and \( x \in \mathbb{R}^l, y \in \mathbb{R}^{d-l} \). Assume that \( B \) is symmetric positive-definite. Let \( z(t, x_0; \eta) \) solve the equation

\[
\frac{dz}{dt} = Az + f_0(z, \eta(z)), \quad z(0, x_0; \eta) = x_0.
\]

Define \( T\eta \) by

\[
(T\eta)(x_0) = \int_{-\infty}^{0} e^{-Bs/\varepsilon} g_0(z(s, x_0; \eta), \eta(z(s, x_0; \eta)))ds.
\]
CHAPTER 15. INVARIANT MANIFOLDS FOR ODE: THE CONVERGENCE THEOREM

Show that if $\eta$ is a fixed point of $T$ then $y = \eta(x)$ is an invariant manifold for the equations (15.4.1). (This is known as the Lyapunov-Perron approach to the construction of invariant manifolds).

4. Assume that $f_0, g_0$ and all derivatives are uniformly bounded. Prove that $T$ from the previous question has a fixed point. To do this apply a contraction mapping argument in a space of Lipschitz graphs $\eta$ which are sufficiently small and have sufficiently small Lipschitz constant.
Chapter 16

Averaging for Markov Chains: The Convergence Theorem

16.1 The Theorem

In this chapter we prove a result concerning averaging for Markov chains. The set-up is as in Chapter 9. To make the proofs simpler we concentrate on the finite state-space case. Let $\mathcal{I}_x, \mathcal{I}_y \subseteq \{1, 2, \ldots \}$ be finite sets. Consider a Markov chain $z^\varepsilon(t) = \left( \begin{array}{c} x^\varepsilon(t) \\ y^\varepsilon(t) \end{array} \right)$ on $\mathcal{I}_x \times \mathcal{I}_y$. We assume that the backward equation has form

$$\frac{dv}{dt} = \frac{1}{\varepsilon}Q_0v + Q_1v$$

(16.1.1)

where $Q_0, Q_1$ are given by (9.2.4). Let $x^0(t)$ be a Markov chain on $\mathcal{I}_x$ with backward equation

$$\frac{dv_0}{dt} = \bar{Q}_1v_0,$$

(16.1.2)

and with $\bar{Q}_1$ given by (9.3.1). We are interested in approximating $x^\varepsilon(t)$ by $x^0(t)$. Note that the formula for the approximate process implied by the Kolmogorov equation is exactly that derived in Chapter 9 by means of formal asymptotics.

Note that $x^\varepsilon(t)$ is not itself Markovian; only the pair $(x^\varepsilon(t), y^\varepsilon(t))$ is. Thus we are approximating a non-Markov stochastic process by a Markovian one. To be precise we prove that, at any fixed time, the statistics of $x^\varepsilon(t)$ are close to those of $x(t)$. That is, we prove weak convergence of $x^\varepsilon(t)$ to $x^0(t)$ at any fixed time $t$. 

213
Theorem 16.1. For any $t > 0$, $x^\varepsilon(t) \Rightarrow x^0(t)$, as $\varepsilon \to 0$.

16.2 The Proof

Let $v_0$ be defined as in (16.1.2). We then have

$$v_0 \in \text{Null}(Q_0), \quad \frac{dv_0}{dt} - Q_1 v_0 \perp \text{Null}(Q_0^*) .$$

This follows from the construction of $v_0$ in Chapter 9. Hence there exists $v_1$ so that

$$Q_0 v_0 = 0, \quad Q_0 v_1 = \frac{dv_0}{dt} - Q_1 v_0 .$$

We can make $v_1$ unique by insisting that it has zero average against the null-space of $Q_0^*$, although this particular choice is not necessary. We simply ask that a solution is chosen which is $\varepsilon$-independent. Since the equation itself is $\varepsilon$-independent, such a solution can always be chosen simply by insisting that the part of the solution $v_1$ in the null-space of $Q_0$ is independent of $\varepsilon$. For any such $v_1$, and for $v_0$ given by (16.1.2), define

$$r = v - v_0 - \varepsilon v_1 .$$

Substituting $v = v_0 + \varepsilon v_1 + r$ into (16.1.1) and using the properties of $v_0, v$, we obtain

$$\frac{dv_0}{dt} + \varepsilon \frac{dv_1}{dt} + \frac{dr}{dt} = \frac{1}{\varepsilon} Q_0 v_0 + Q_0 v_1 + \frac{1}{\varepsilon} Q_0 r + Q_1 v_0 + \varepsilon Q_1 v_1 + Q_1 r .$$

Hence

$$\frac{dr}{dt} = \left( \frac{1}{\varepsilon} Q_0 + Q_1 \right) r + \varepsilon q , \quad q = Q_1 v_1 - \frac{dv_1}{dt} .$$

Now $Q = \frac{1}{\varepsilon} Q_0 + Q_1$ is the generator of a Markov chain. Hence using $\| \cdot \|_\infty$ to denote the supremum norm on vectors over the finite set $I_x \times I_y$, as well as the induced operator norm, we have

$$\| e^{Qt} \|_\infty = 1 . \quad (16.2.1)$$

This follows from (5.1.2) because $e^{Qt}$ is a stochastic matrix.
16.3. DISCUSSION AND BIBLIOGRAPHY

By the variation of constants formula we have

\[ r(t) = e^{Qt}r(0) + \varepsilon \int_0^t e^{Q(t-s)}q(s)ds, \quad (16.2.2) \]

viewing \( r(t), q(t) \) as vectors on \( I_x \times I_y \), for each \( t \). We assume that \( v(i, j, 0) = \phi(i), v_0(i, 0) = \phi(i) \) and then

\[
\begin{align*}
    v(i, j, t) &= \mathbb{E}\phi(x^\varepsilon(t)|x^\varepsilon(0) = i, y^\varepsilon(0) = j) \quad (16.2.3a) \\
    v_0(i, t) &= \mathbb{E}\phi(x^0(t)|x(0) = i). \quad (16.2.3b)
\end{align*}
\]

Weak convergence of \( x^\varepsilon(t) \) to \( x^0(t) \), for fixed \( t \), is proved if

\[ v(i, j, t) \to v_0(i, t), \quad \text{as} \quad \varepsilon \to 0, \]

for any \( \phi : I_x \mathbb{R} \). Equations (16.2.3) imply that \( r(0) = -\varepsilon v_1(0) \). Using (16.2.1) we have, from (16.2.2),

\[
\begin{align*}
    ||r(t)||_\infty &\leq \varepsilon ||e^{Qt}||_{\infty} ||v_1(0)||_\infty + \varepsilon \int_0^t ||e^{Q(t-s)}||_\infty ||q(s)||_\infty ds \\
    &\leq \varepsilon ||v_1(0)||_\infty + \varepsilon \int_0^t ||q(s)||_\infty ds \\
    &\leq \varepsilon (||v_1(0)||_\infty + t \sup_{0 \leq s \leq t} ||q(s)||_\infty).
\end{align*}
\]

Hence, for any fixed \( t > 0 \), \( r(t) \to 0 \) as \( \varepsilon \to 0 \) and it follows that \( v \to v_0 \) as \( \varepsilon \to 0 \).

Remark 16.2. The proof actually gives a convergence rate because it shows that

\[ ||v(t) - v_0(t)||_\infty = O(\varepsilon), \quad \text{as} \quad \varepsilon \to 0. \]

16.3 Discussion and Bibliography

The result we prove only shows that the random variable \( x^\varepsilon(t) \) converges to \( x^0(t) \) for each fixed \( t \). It is also possible to prove the more interesting result that the convergence actually occurs pathwise in the Skorokhod topology (defined in [55]). See [136].
16.4 Exercises

1. Consider the two state continuous time Markov chain $y$ with generator

$$L = \frac{1}{\varepsilon} \begin{pmatrix} -a & a \\ b & -b \end{pmatrix}$$

and state-space $\mathcal{I} = \{-1, +1\}$. Consider the ODE on $\mathbb{T}^d$ given by

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

where $f: \mathbb{T}^d \times \mathcal{I} \to \mathbb{R}^d$.

a. Write down the generator for this process.

b. Using multiscale analysis, show that the averaged SDE is

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

where

$$F(x) = \lambda f(x, +1) + (1 - \lambda) f(x, -1)$$

and $\lambda \in (0, 1)$ should be specified.

2. Prove the assertions made in the preceding exercise.

3. Conjecture what the behaviour of $x$ is in the case where $F(x) = 0$ and large times are considered.
Chapter 17

Averaging for SDE: The Convergence Theorem

17.1 The Theorem

The goal of this chapter is to develop a rigorous theory based on the averaging principle for SDE that we developed in Chapter 10. To allow for a simplified proof we study the problem on the torus. Consider the SDE on \( \mathbb{T}^d \) given by

\[
\begin{align*}
\frac{dx}{dt} &= f(x, y), \quad (17.1.1a) \\
\frac{dy}{dt} &= \frac{1}{\varepsilon} g(x, y) + \frac{1}{\sqrt{\varepsilon}} \beta(x, y) \, dV \\
(17.1.1b)
\end{align*}
\]

where \( V \) is a standard Brownian motion on \( \mathbb{R}^{d-l} \), \( f : \mathbb{T}^l \times \mathbb{T}^{d-l} \to \mathbb{R}^l \), \( g : \mathbb{T}^l \times \mathbb{T}^{d-l} \to \mathbb{R}^{d-l} \), \( \beta : \mathbb{T}^l \times \mathbb{T}^{d-l} \to \mathbb{R}^{(d-l) \times (d-l)} \) are smooth periodic functions.

Assume also that, writing \( z = (x^T, y^T)^T \),

\[
\exists \beta > 0 : \langle \xi, \beta(x, y) \xi \rangle \geq \beta \xi^2, \quad \forall \xi \in \mathbb{R}^{d-l}, \quad z \in \mathbb{T}^d.
\]

Recall that, under this assumption, the process found by freezing \( x = \xi \) in (17.1.1b), is ergodic (Theorem 6.10). Thus an effective equation for the evolution of \( x \) can be found by averaging \( f \) over the invariant measure of this ergodic process. We now make this idea precise.

The process \( \varphi^\xi_t \) given by (10.2.2) is ergodic and has a smooth invariant density \( \rho^\infty(y; \xi) \). This invariant density spans the null-space of \( \mathcal{L}_0 \), found as the adjoint of \( \mathcal{L} \) given by

\[
\mathcal{L}_0 = g(x, y) \cdot \nabla y + \frac{1}{2} B(x, y) : \nabla y \nabla y
\]

(17.1.2)
where \( B(x, y) = \beta(x, y)\beta(x, y)^T \); both \( \mathcal{L}_0 \) and \( \mathcal{L}_0^* \) have periodic boundary conditions. The averaged equations are then

\[
\frac{dX}{dt} = F(X), \quad (17.1.3a)
\]

\[
F(\xi) = \int_{T^{d-1}} f(\xi, y)\rho^\infty(y; \xi)dy. \quad (17.1.3b)
\]

Note that \( F : T^l \rightarrow \mathbb{R}^l \) is periodic by construction. The resulting formulae are exactly those given in Chapter 10, specialized to the particular drift and diffusion coefficients on the torus studied in this chapter.

**Theorem 17.1.** Let \( p > 1 \). Let \( x(0) = X(0) \). Then the function \( x(t) \) solving (17.1.1) converges to \( X(t) \) solving (17.1.3) in \( L^p(\Omega, C([0, T], T^l)) \); if \( x(0) = X(0) \) then, for any \( T > 0 \), there is \( C = C(T) \) such that

\[
\mathbb{E} \sup_{0 \leq t \leq T} |x(t) - X(t)|^p \leq C\varepsilon^{p/2}.
\]

### 17.2 The Proof

Recall that \( \mathcal{L}_0 \) is the generator for \( \varphi_x^t(y) \), with \( x \) viewed as a fixed parameter, given by (17.1.2). Thus \( \mathcal{L}_0 \) is a differential operator in \( y \) only; \( x \) appears as a parameter. Now let \( \phi(x, y) \) solve the elliptic boundary value problem

\[
\mathcal{L}_0\phi(x, y) = f(x, y) - F(x),
\]

\[
\int_{T^{d-1}} \phi(x, y)dy = 0,
\]

\[
\phi(x, \cdot) \quad \text{is periodic on } T^{d-1}.
\]

By construction

\[
\int_{T^{d-1}} (f(x, y) - F(x))\rho^\infty(y; x)dy = 0
\]

and \( \rho^\infty \) spans \( \text{Null}(\mathcal{L}_0^*) \). By the Fredholm alternative, \( \phi \) has a unique solution.

**Lemma 17.2.** The functions \( f, \phi, \nabla_x\phi, \nabla_y\phi, \beta \) are smooth and bounded.

**Proof.** The properties of \( f, \beta \) follow from the fact that they are defined on the torus and have derivatives of all orders by assumption. Since the invariant density \( \rho^\infty \) is the solution of an elliptic eigenvalue problem on the torus, it too is smooth and
periodic. Hence $F$ is smooth and periodic. Consequently $f - F$ is smooth and periodic. Hence $\phi$ and all its derivatives are smooth and periodic. \qed

**Proof of Theorem** Notice that the generator for (17.1.1) is

$$\mathcal{L} = \frac{1}{\varepsilon} \mathcal{L}_0 + \mathcal{L}_1.$$ Here $\mathcal{L}_0$ is given by (17.1.2) and

$$\mathcal{L}_1 = f(x, y) \cdot \nabla_x.$$ Now we apply the Itô formula (Lemma 6.4) to $\phi$ to obtain the following informal expression, with precise interpretation found by integrating in time:

$$\frac{d\phi}{dt}(x, y) = \frac{1}{\varepsilon} (\mathcal{L}_0 \phi)(x, y) + f(x, y) \cdot \nabla_x \phi(x, y) + \frac{1}{\sqrt{\varepsilon}} \nabla_y \phi(x, y) \beta(x, y) \frac{dV}{dt}.$$ Since $\mathcal{L}_0 \phi = f - F$ we obtain

$$\frac{dx}{dt} = F(x) + (\mathcal{L}_0 \phi)(x, y)$$

$$= F(x) + \varepsilon \frac{d\phi}{dt} - \varepsilon f(x, y) \cdot \nabla_x \phi(x, y)$$

$$- \sqrt{\varepsilon} \nabla_y \phi(x, y) \beta(x, y) \frac{dV}{dt}. \quad (17.2.1)$$

(Again a formal expression made rigorous by time-integration). The functions $f, \phi, \nabla_x \phi$ are all smooth and bounded by Lemma 17.2 so that

$$\theta(t) := (\phi(x(t), y(t)) - \phi(x(0), y(0))) - \int_0^t f(x(s), y(s)) \cdot \nabla_x \phi(x(s), y(s)) ds$$ satisfies

$$\sup_{0 \leq \tau \leq t} |\theta(\tau)| \leq C_1.$$ Now define

$$M(t) := - \int_0^t \nabla_y \phi(x(s), y(s)) \beta(x(s), y(s)) dV(s).$$ Since $\nabla_y \phi, \beta$ are smooth and bounded, by Lemma 17.2, the Itô isometry gives

$$\mathbb{E}[M(t)]^2 = \int_0^t \mathbb{E}[\nabla_y \phi(x(s), y(s)) \beta(x(s), y(s))]^2 ds$$

$$\leq C_2 t. \quad (17.2.2)$$
Now the precise interpretation of (17.2.1) is
\[ x(t) = x(0) + \int_0^t F(x(s)) \, ds + \varepsilon \theta(s) + \sqrt{\varepsilon} M(t). \]

Also, from (17.1.3),
\[ X(t) = X(0) + \int_0^t F(X(s)) \, ds. \]

Let \( e(t) = x(t) - X(t) \) so that, using \( e(0) = 0 \),
\[ e(t) = \int_0^t (F(x(s)) - F(X(s))) \, ds + \varepsilon \theta(t) + \sqrt{\varepsilon} M(t). \]

Since \( F \) is Lipschitz on \( \mathbb{T}^l \) we obtain, for \( t \in [0, T] \),
\[ |e(t)| \leq \int_0^t L |e(s)| \, ds + \varepsilon C_1 + \sqrt{\varepsilon} |M(t)|. \]

Hence, by (17.2.2) and the Burkholder-Davis-Gundy inequality Theorem 3.15, we obtain
\[
\mathbb{E} \sup_{0 \leq t \leq T} |e(t)|^p \leq C \left( \varepsilon^p + \varepsilon^{p/2} \left( \mathbb{E} \sup_{0 \leq t \leq T} |M(t)|^p \right) + L^p T^{p-1} \int_0^T \mathbb{E} |e(s)|^p \, ds \right)^{1/p} \\
\leq C \left( \varepsilon^p + \varepsilon^{p/2} \left( \mathbb{E} |M(T)|^2 \right)^{p/2} + L^p T^{p-1} \int_0^T \mathbb{E} |e(s)|^p \, ds \right)^{1/p} \\
\leq C \left( \varepsilon^{p/2} + \int_0^T \mathbb{E} \sup_{0 \leq \tau \leq s} |e(\tau)|^p \, ds \right)^{1/p}.
\]

By the integrated version of the Gronwall inequality in Lemma 4.3 we deduce that
\[
\mathbb{E} \sup_{0 \leq t \leq T} |e(t)|^p \leq C \varepsilon^{p/2}.
\]

Note that the constants \( C \) grows with \( T \).

### 17.3 Discussion and Bibliography

Note that our convergence result proves strong convergence: we compare each path of the SDE for \( (x, y) \) with the approximating ODE for \( X \). This strong convergence result is possible primarily because the limiting approximation is in this case deterministic. When the approximation is itself stochastic, as arises for (10.6.1), then it
is more natural to study weak convergence type results (but see Exercise 1 below). For results of the latter type see [41].

Averaging results for ODE are proved by different techniques. Much of the original motivation for this work comes from averaging of perturbed integrable Hamiltonian systems, expressed in action-angle variables. See [62] for references.

17.4 Exercises

1. Consider equation (10.6.1) in the case where $\alpha(x, y) \equiv 1$. Write down the averaged dynamics for $X$ and modify the techniques of this chapter to prove strong convergence of $x$ to $X$.

2. Consider equation (10.6.1) in the case where $d = 2, l = 1$ and $\alpha(x, y) \equiv y^2$. If $y$ is a scalar OU process (6.4.2), independent of $x$, write down the averaged dynamics for $X$ and use the properties of the OU process to prove weak convergence.
Chapter 18

Homogenization for SDE: The Convergence Theorem

18.1 The Theorem

In this chapter we develop a rigorous theory based on the homogenization principle derived in Chapter 11. We consider a simple setting where the entire problem is posed on the torus, in order to elucidate the principle ideas. Furthermore, we work in the skew-product setting where the fast process is independent of the slow process, and where the fluctuating term in the slow process depends only on the fast process; this allows for a simplified proof which nonetheless contains the essence of the main ideas.

Consider the SDE on $\mathbb{T}^d$ given by

$$
\begin{align*}
\frac{dx}{dt} &= \frac{1}{\varepsilon} f_0(y) + f_1(x, y) + \alpha \frac{dU}{dt} \\
\frac{dy}{dt} &= \frac{1}{\varepsilon^2} g(y) + \frac{1}{\varepsilon} \beta(y) \frac{dV}{dt}
\end{align*}
$$

(18.1.1)

where $U$ (resp. $V$) is a standard Brownian motion on $\mathbb{R}^l$ (resp. $\mathbb{R}^{d-l}$) and $\alpha \in \mathbb{R}^{l \times l}$. The functions $f_0 : \mathbb{T}^{d-l} \mapsto \mathbb{R}^l$, $f_1 : \mathbb{T}^l \times \mathbb{T}^{d-l} \mapsto \mathbb{R}^l$, $g : \mathbb{T}^{d-l} \mapsto \mathbb{R}^{d-l}$, $\beta : \mathbb{T}^{d-l} \mapsto \mathbb{R}^{(d-l) \times (d-l)}$ are smooth and periodic. We also assume that

$$
\exists \beta > 0 : \langle \xi, \beta(y) \xi \rangle \geq \beta |\xi|^2, \quad \forall \xi \in \mathbb{R}^{d-l}, \quad y \in \mathbb{T}^{d-l}.
$$

Recall that, under this assumption, the process $y$ in (18.1.1b) is ergodic (see Result 6.10.) Thus an effective equation for the evolution of $x$ can be found by averag-
ing \( f_1 \) over the invariant measure of this ergodic process, and by examining the fluctuations induced by \( f_0 \). The aim of this chapter is to quantify this idea.

Define the generator \( \mathcal{L}_0 \) by (17.1.2):

\[
\mathcal{L}_0 v = g(y) \cdot \nabla_y v + \frac{1}{2} B(y) : \nabla_y \nabla_y v
\]

with periodic boundary conditions, and where

\[
B(y) = \beta(y)\beta(y)^T.
\]

The precise statement about ergodicity of the \( y \) process is that \( \varphi^t(y) \) given by (10.2.2) is ergodic and has a smooth invariant density \( \rho^\infty(y) \); this function spans the null-space of \( \mathcal{L}_0^* \). Note that \( \varphi^t(\cdot) \) and \( \rho^\infty(\cdot) \) are independent of \( \xi \) here, because \( g, \beta \) depend only on \( y \). We assume that

\[
\int_{\mathbb{T}^{d-l}} f_0(y) \rho^\infty(y) dy = 0.
\]

Under this assumption the calculations in Chapter 11 we may homogenize the SDE (18.1.1) to obtain

\[
\frac{dx}{dt} = F(x) + A \frac{dW}{dt}
\]

(18.1.2)

where

\[
F(\xi) = \int_{\mathbb{T}^{d-l}} f_1(\xi, y) \rho^\infty(y) dy
\]

and

\[
AA^T = \alpha \alpha^T + \int_{\mathbb{T}^{d-l}} \left( f_0(y) \otimes \Phi(y) + \Phi(y) \otimes f_0(y) \right) \rho^\infty(y) dy.
\]

Here \( \Phi(y) \) solves the cell problem

\[
\mathcal{L}_0 \Phi(y) = -f_0(y),
\]

(18.1.3)

\[
\int_{\mathbb{T}^{d-l}} \Phi(y) \rho^\infty(y) = 0,
\]

\[
\Phi(y) \quad \text{periodic on } \mathbb{T}^{d-l}.
\]

This has a unique solution, by the Fredholm alternative. The resulting formulae are exactly those given in Chapter 11, specialized to the particular drift and diffusion coefficients studied in this chapter.
Notice that, for $\xi \in \mathbb{R}^l$, $\phi = \Phi \cdot \xi$, the proof of Theorem 11.5 shows that $AA^T$ is positive-definite

$$\langle \xi, AA^T \xi^T \rangle = |\alpha^T \xi|^2 + \int_{T_{d-1}} |\beta^T(y) \nabla \phi(y)|^2 \rho^\infty(y)dy.$$ 

The Remark 11.6 shows that

$$AA^T = \alpha \alpha^T + \int_{T_{d-1}} \rho^\infty(y) \left( \nabla \Phi(y) \beta \otimes \nabla \Phi(y) \beta \right) dy.$$ 

Hence the SDE for $X$ is well-defined.

**Theorem 18.1.** Let $x(t)$ solve (17.2.1) and $X(t)$ solve (18.1.2) and set $x(0) = X(0)$. Then, for any $T > 0$, $x \Rightarrow X$ in $C([0, T], \mathbb{T}^l)$.

### 18.2 The Proof

We will use the following lemma.

**Lemma 18.2.** Let $w \in C([0, T], \mathbb{R}^r)$; let $F \in C^1(\mathbb{T}^l; \mathbb{T}^l)$ and let $A \in \mathbb{R}^{l \times r}$. If $u$ satisfies the integral equation

$$u(t) = u(0) + \int_0^t F(u(s)) ds + Aw(t)$$

then $u \in C([0, T], \mathbb{T}^l)$ and the mapping $w \mapsto u$ is a continuous mapping from $(C([0, T], \mathbb{R}^r))$ into $C([0, T], \mathbb{T}^l)$.

**Proof.** Existence and uniqueness follows by a standard contraction mapping argument, based on the iteration

$$u^{(n+1)}(t) = u(0) + \int_0^t F(u^n(s)) ds + Aw(t),$$

and using the fact that $F$ is globally Lipschitz, with constant $L$. For continuity consider the equations

$$u^{(i)}(t) = u(0) + \int_0^t F(u^{(i)}(s)) ds + Aw^{(i)}(t)$$

for $i = 1, 2$. Subtracting and letting $e = u^1 - u^2$, $\delta = w^1 - w^2$, we get

$$e(t) = \int_0^t \left( F(u^1(s)) - F(u^2(s)) \right) ds + A\delta(s).$$
Hence
\[ |e(t)| \leq \int_0^t L|e(s)| ds + |A||\delta(t)|. \]

From the integrated form of the Gronwall inequality in Lemma 4.3 it follows that
\[ \sup_{0 \leq t \leq T} |e(t)| \leq C \sup_{0 \leq t \leq T} |\delta(t)|. \]

This establishes continuity.

Let \( \chi(x, y) \) solve the cell problem
\[
\begin{align*}
L_0 \chi(x, y) &= f_1(x, y) - F(x), \\
\int_{T^{d-l}} \chi(x, y) \rho^\infty(y) dy &= 0, \\
\chi(x, y) &\text{ periodic on } T^{d-l}.
\end{align*}
\]

This has a unique solution, by the Fredholm alternative, since \( f_1 - F \) averages to zero over \( T^{d-l} \). The following lemma has proof which is very similar to the proof of Lemma 17.2; hence we omit it.

**Lemma 18.3.** The functions \( f_0, f_1, \Phi, \chi \) and all their derivatives are smooth and bounded.

**Proof of Theorem**

Notice that, by the Itô formula Lemma 6.4, we have that
\[
\frac{d\chi}{dt} = \frac{1}{\varepsilon^2} L_0 \chi + \frac{1}{\varepsilon} L_1 \chi + L_2 \chi + \nabla_x \chi \frac{dU}{dt} + \frac{1}{\varepsilon} \nabla_y \chi \frac{dV}{dt};
\]

the rigorous interpretation is, as usual, the integrated form. Here the operators \( L_1, L_2 \) are given by
\[
\begin{align*}
L_1 &= f_0(x, y) \cdot \nabla_x, \\
L_2 &= f_1(x, y) \cdot \nabla_x + \frac{1}{2}\alpha^T : \nabla_x \nabla_x.
\end{align*}
\]

Now define
\[
\theta(t) = \varepsilon^2 \left( \chi(x(t), y(t)) - \chi(x(0), y(0)) \right)
- \varepsilon \int_0^t \left( L_1 \chi \right)(x(s), y(s)) ds
- \varepsilon^2 \int_0^t \left( L_2 \chi \right)(x(s), y(s)) ds.
\]
and

\[ M_1(t) = \varepsilon^2 \int_0^t \nabla_x \chi(x(s), y(s)) \alpha dU(s) \]
\[ + \varepsilon \int_0^t \nabla_y \chi(x(s), y(s)) \beta(y(s)) dV(s). \]  

(18.2.4)

Since \( \chi, f_0 \) and \( f_1 \) and all their derivatives are bounded, we have

\[ \mathbb{E} \sup_{0 \leq t \leq T} |\theta(t)|^p \leq C \varepsilon^p. \]  

(18.2.5)

Furthermore

\[ \mathbb{E}[M_1(t)]^2 = \varepsilon^4 \int_0^t |\nabla_x \chi(x(s), y(s)) \alpha(x(s), y(s))|^2 \, ds \]
\[ + \varepsilon^2 \int_0^t |\nabla_y \chi(x(s), y(s)) \beta(y(s))|^2 \, ds. \]

By the Burkholder-Davis-Gundy inequality of Theorem 3.15 we deduce that

\[ \mathbb{E} \sup_{0 \leq t \leq T} |M_1(t)|^p \leq C \varepsilon^p. \]  

(18.2.6)

Hence, by (18.2.2) and (18.2.1), we deduce that

\[ \int_0^t (f_1(x(s), y(s)) - F(x(s))) \, ds = r(t) := \theta(t) + M_1(t) \]  

(18.2.7)

where we have shown that

\[ r(t) = O(\varepsilon^p) \quad \text{in} \quad L^p(\Omega, C([0, T], \mathbb{R}^l)). \]  

(18.2.8)

Now apply the Itô formula to \( \Phi \) solving (18.1.3) to obtain, since \( \Phi \) is independent of \( x \),

\[ \frac{d\Phi}{dt} = \mathcal{L}_0 \Phi + \frac{1}{\varepsilon} \nabla_y \Phi \beta \frac{dV}{dt}. \]

(As usual the rigorous interpretation is in integrated form). Hence

\[ \frac{1}{\varepsilon} \int_0^t f_0(y(s)) \, ds = \varepsilon(\Phi(y(0)) - \Phi(y(s))) + \int_0^t \nabla_y \Phi(y(s)) \beta(y(s)) \, dV. \]  

(18.2.9)

Since \( \Phi \) is bounded on \( \mathbb{T}^{d-l} \) we deduce that, as \( \varepsilon \to 0 \),

\[ \varepsilon(\Phi(y(t)) - \Phi(y(0))) = O(\varepsilon^p) \quad \text{in} \quad L^p(\Omega, C([0, T], \mathbb{R}^l)). \]  

(18.2.10)
If we set
\[ M_2(t) = \int_0^t (\nabla_y \Phi)(y(s)) \beta(y(s)) dV(s), \]
then the Martingale Central Limit Theorem 3.31 implies that, as \( \varepsilon \to 0 \),
\[ M_2(t) \Rightarrow \alpha_2 W(t), \quad (18.2.11) \]
where \( W(t) \) is standard Brownian motion on \( \mathbb{R}^l \) and
\[ \alpha_2 \alpha_2^T = \lim_{t \to \infty} \frac{1}{t} \int_0^t (\nabla_y \Phi)(y(s)) \beta(y(s)) \otimes (\nabla_y \Phi)(y(s)) \beta(y(s)) ds \]
\[ = \int_{\mathbb{T}^d} \rho^{\infty}(y) (\nabla_y \Phi)(y) \beta(y) \otimes (\nabla_y \Phi)(y) \beta(y) du. \]
Combining (18.2.7), (18.2.9) in (18.1.1a) we obtain
\[ x(t) = x(0) + \int_0^t F(x(s)) ds + \alpha U(t) + M_2(t) + \eta(t) \]
where
\[ \eta(t) = r(t) + \varepsilon (\Phi(y(0)) - \Phi(y(t))). \]
By (18.2.8), (18.2.10) we have that
\[ \eta(t) \to 0, \quad \text{in} \quad L^p(\Omega, C([0, T], \mathbb{R}^l)) \]
and hence by (18.2.11),
\[ (\eta(t), M_2(t)) \Rightarrow (0, \alpha_2 W(t)), \quad \text{in} \quad C([0, T], \mathbb{R}^{2l}). \]
Because \( U(t) \) is independent of \( \eta(t), M(t) \) we deduce that
\[ (U(t), \eta(t), M_2(t)) \Rightarrow (U(t), 0, \alpha_2 W(t)), \quad \text{in} \quad C([0, T], \mathbb{R}^{3l}). \]

The mapping \( (U, \eta, M) \to x \) is continuous from \( C([0, T], \mathbb{R}^{3l}) \) into \( C([0, T], \mathbb{T}^d) \),
by Lemma 18.2. Hence, because weak convergence is preserved under continuous mappings, we deduce that \( x(t) \Rightarrow X(t) \) in \( C([0, T], \mathbb{T}^d) \), where \( X \) solves
\[ X(t) = x(0) + \int_0^t F(X(s)) ds + \alpha U(t) + \alpha_2 W(t). \]
18.3 Discussion and Bibliography

The proofs presented here have been simplified by the assumption that the limiting process for $X$ has additive noise. Thus we were able to use the Martingale Central Limit Theorem in a very straightforward way. In the general case where the limiting process has general state-dependent noise proofs of convergence are more complicated. See [136, 17, 41].

18.4 Exercises

1. Consider the coupled pair of scalar SDEs

$$\frac{dx}{dt} = f(x, y) - \frac{1}{\epsilon} y + \left(y^2 + a(x)\right) \frac{dU}{dt} \quad (18.4.1a)$$

$$\frac{dy}{dt} = -\frac{1}{\epsilon^2} y + \frac{1}{\epsilon} \sqrt{2} \frac{dV}{dt} \quad (18.4.1b)$$

Write down the homogenized equations.

2. Assume that $f, a$ and all derivatives are bounded. Using the exact solution of the OU process prove a convergence theorem related to your conjectured homogenized equation in the previous exercise.
Chapter 19

Homogenization for Elliptic PDE: The Convergence Theorem

19.1 The Theorems

In this chapter we prove the homogenization theorem for second order uniformly elliptic PDE with periodic coefficients and Dirichlet boundary conditions:

**Theorem 19.1.** Let \( u^\varepsilon \) be the solution of

\[
\begin{align*}
-\nabla \cdot (A^\varepsilon(x) \nabla u^\varepsilon) &= f, \text{ for } x \in \Omega, \\
u^\varepsilon(x) &= 0, \text{ for } x \in \partial \Omega.
\end{align*}
\]

with \( f \in H^{-1}(\Omega) \) and \( A^\varepsilon(x) = A\left(\frac{x}{\varepsilon}\right) \), \( A(y) \in M_{\text{per}}(\alpha, \beta, \mathcal{Y}) \). Furthermore, let \( u \) be the solution of the homogenized problem

\[
\begin{align*}
-\overline{A} : \nabla \nabla u &= f, \text{ for } x \in \Omega, \\
u(x) &= 0, \text{ for } x \in \partial \Omega,
\end{align*}
\]

with \( \overline{A} \) given by

\[
\overline{A} = \int_Y \left( A(y) + A(y) \nabla_y \chi(y) \right) dy.
\]

and where the vector field \( \chi(y) \) satisfies

\[
-\nabla_y \cdot (A(y) \nabla_y \chi(y)) = \nabla_y \cdot A(y)T, \quad \chi(y) \text{ is } 1\text{-periodic}.
\]

Then

\[
u^\varepsilon \rightharpoonup u \text{ weakly in } H^1_0(\Omega).
\]
In addition to the basic homogenization theorem, we will also prove a result which shows that retaining extra terms in the multiscale expansion does indeed give improved approximations. The following result says that we can get strong convergence in $H^1(\Omega)$ provided that we take the first order corrector field into account.

Theorem 19.2. Consider $u^\varepsilon(x)$ and $u(x)$ as in Theorem 19.1. Then

$$\lim_{\varepsilon \to 0} \| u^\varepsilon(x) - \left( u(x) + \varepsilon \chi \left( \frac{x}{\varepsilon} \right) \cdot \nabla u(x) \right) \|_{H^1(\Omega)} = 0. \quad (19.1.5)$$

19.2 Proof of the Homogenization Theorem

In this section we prove the homogenization theorem, Theorem 19.1 using the method of two–scale convergence. Before starting with the proof let us make some remarks on our approach. The first step in our analysis is to use the energy estimates from Chapter 7 to deduce that $u^\varepsilon$ as well as $\nabla u^\varepsilon$ have two–scale convergent subsequences. The second step is to use a test function of the form

$$\phi^\varepsilon(x) = \phi_0(x) + \varepsilon \phi_1 \left( x, \frac{x}{\varepsilon} \right), \quad (19.2.1)$$

in order to pass to the two scale limit. In this way we obtain a coupled system of equations for the first two terms in the expansion $\{ u, u_1 \}$, the two–scale system, see equation (19.2.2) below. Well–posedness of this system is proved using the Lax–Milgram lemma, Theorem 2.41. The final step is to decouple this system of equations using separation of variables, thereby obtaining the homogenized equation for $u$.

Let us remark that the basic compactness theorems of two–scale convergence, Theorems 2.33 and 2.36, enable us to extract a two–scale convergent subsequence but they do not allow us to conclude that the whole sequence converges to a limit. In the case of Theorem 19.1 the limit $u$ is unique, since the homogenized equation (19.1.2) has a unique solution. This implies that the whole sequence converges, not just a subsequence. In the calculations that follow we will use this result without any further reference.

We will break the proof of Theorem 19.1 into three parts, as described above. The following lemma provides us with the first two terms of the two–scale expansion, together with the coupled system of equations that they satisfy.
Lemma 19.3. Let \( u^\varepsilon(x) \) be the solution of (19.1.1) with the assumptions of Theorem 19.1. Then there exist functions \( \{ u(x), u_1(x, y) \} \in \{ H_0^1(\Omega) \times L^2(\Omega; H_{\text{per}}^1(\mathcal{Y})/\mathbb{R}) \} \) such that \( u^\varepsilon \) and \( \nabla u^\varepsilon \) two–scale converge to \( u(x) \) and \( \nabla_x u + \nabla_y u_1 \). Furthermore, \( \{ u, u_1 \} \) satisfy the two–scale system

\[
\begin{align*}
-\nabla_y \cdot \left( A(\mathcal{Y}) \left( \nabla_x u + \nabla_y u_1 \right) \right) &= 0 \quad \text{in } \Omega \times \mathcal{Y}, \\
-\nabla_x \cdot \left( \int_{\mathcal{Y}} A(y) \left( \nabla_x u + \nabla_y u_1 \right) \, dy \right) &= f \quad \text{in } \Omega, \\
u(x) &= 0 \quad \text{on } \partial \Omega, \quad u_1(x, y) \text{ is periodic in } y.
\end{align*}
\]  

(19.2.2a, 19.2.2b, 19.2.2c)

Proof. 1. We have that \( \| u^\varepsilon \|_{H_0^1(\Omega)} \leq C \) which implies the first part of the lemma: there exist functions \( u(x) \in H_0^1(\Omega), u_1(x, y) \in L^2(\Omega; H_{\text{per}}^1(Y)/\mathbb{R}) \) such that

\[
\begin{align*}
u^\varepsilon(x) &\overset{2}{\rightharpoonup} u(x), \\
\nabla u^\varepsilon(x) &\overset{2}{\rightharpoonup} \nabla_x u(x) + \nabla_y u_1(x, y).
\end{align*}
\]  

(19.2.3a, 19.2.3b)

Furthermore, the two–scale limit of \( u^\varepsilon(x) \) is also the weak \( H_0^1(\Omega) \)–limit of this sequence.

2. The weak formulation of (19.1.1) is

\[
\int_{\Omega} \langle \nabla^\varepsilon(x) \nabla u^\varepsilon, \nabla \phi^\varepsilon \rangle \, dx = \langle f, \phi^\varepsilon \rangle \quad \forall \phi^\varepsilon \in H_0^1(\Omega).
\]  

(19.2.4)

We use a test function of the form (19.2.1)

\[
\phi^\varepsilon(x) = \phi_0(x) + \varepsilon \phi_1 \left( x, \frac{x}{\varepsilon} \right), \quad \phi_0 \in C_0^\infty(\Omega), \phi_1 \in C_0^\infty(\Omega; C_0^\infty(\mathcal{Y})).
\]

We clearly have that \( \phi^\varepsilon \in H_0^1(\Omega) \). Upon using this test function in (19.2.4) and rearranging terms we obtain:

\[
\begin{align*}
\int_{\Omega} \langle \nabla^\varepsilon, (A^\varepsilon(x))^T \cdot \nabla \phi^\varepsilon \rangle \, dx &= \int_{\Omega} \langle \nabla^\varepsilon, A^\varepsilon(x) \cdot \left( \nabla_x \phi_0(x) + \nabla_y \phi_1 \left( x, \frac{x}{\varepsilon} \right) \right) \rangle \, dx \\
&\quad + \varepsilon \int_{\Omega} \langle \nabla^\varepsilon, (A^\varepsilon(x))^T \cdot \nabla_x \phi_1 \left( x, \frac{x}{\varepsilon} \right) \rangle \, dx \\
&=: I_1 + \varepsilon I_2 \\
&= \langle f, \phi_0 + \varepsilon \phi_1 \rangle.
\end{align*}
\]
Now, the function \((A_\epsilon(x))^T (\nabla_x \phi_0(x) + \nabla_y \phi_1(x, \frac{x}{\epsilon}))\) is of the form \(\phi_1(y)\phi_2(x, y)\) with \(\phi_1(y) \in L^\infty(Y)\) and \(\phi_2(x, y) \in L^2(\Omega; C^\text{per}(Y))\). Hence, it is an admissible test function. We can thus pass to the two scale limit to obtain:

\[
I_1 \rightarrow \int_{\Omega} \int_Y \langle A(y) \cdot (\nabla_x u + \nabla_y u_1), (\nabla_x \phi_0 + \nabla_y \phi_1) \rangle \, dydx.
\]

The function \(A_\epsilon(x) \cdot \nabla_x \phi_1(x, \frac{x}{\epsilon})\) is also an admissible test function. Passing to the limit in \(I_2\) we obtain:

\[
I_2 \rightarrow 0.
\]

Moreover, we have that \(\phi_0 + \epsilon \phi_1 \rightharpoonup \phi_0\) weakly in \(H^1_0(\Omega)\) which implies that

\[
\langle f, \phi_0 + \epsilon \phi_1 \rangle \rightarrow \langle f, \phi_0 \rangle.
\]

Putting the above considerations together we obtain the limiting equation

\[
\int_{\Omega} \int_Y \langle A(y) \cdot (\nabla_x u + \nabla_y u_1), \nabla_x \phi_0 + \nabla_y \phi_1 \rangle \, dydx = \langle f, \phi_0 \rangle. \tag{19.2.5}
\]

In deriving (19.2.5) we assumed that the test functions \(\phi_0, \phi_1\) are smooth, a density argument however enables us to conclude that (19.2.5) holds for every \(\phi_0 \in L^2(\Omega), \phi_1 \in L^2(\Omega; H^1_\text{per}(Y)/\mathbb{R})\).

3. Now, (19.2.5) is the weak formulation of the two–scale system (19.2.2). To see this, we set \(\phi_0 = 0\) to obtain:

\[
\int_{\Omega} \int_Y \langle A(y) \cdot (\nabla_x u + \nabla_y u_1), \nabla_y \phi_1 \rangle \, dydx = 0.
\]

which is precisely the weak formulation of (19.2.2a). Setting now \(\phi_1 = 0\) in (19.2.5) we get

\[
\int_{\Omega} \int_Y \langle A(y) \cdot (\nabla_x u + \nabla_y u_1), \nabla_x \phi_0 \rangle \, dydx = \langle f, \phi_0 \rangle.
\]

which is the weak formulation of (19.2.2b). The boundary conditions (19.2.2c) follow from the fact that \(u(x) \in H^1_0(\Omega)\) and \(u_1(x, y) \in L^2(\Omega; H^1_\text{per}(Y)/\mathbb{R})\).

Now we need to prove that the two–scale system is well posed. This is the content of the next lemma.

**Lemma 19.4.** Under the assumptions of Theorem 19.1 the two–scale system (19.2.2) has a unique solution \(\{u(x), u_1(x, y)\} \in \{H^1_0(\Omega) \times L^2(\Omega; H^1_\text{per}(Y)/\mathbb{R})\}\).
19.2. PROOF OF THE HOMOGENIZATION THEOREM

Proof. We will use the Lax–Milgram Lemma. The weak formulation of the two-scale system is given by equation (19.2.5) with \((\phi_0, \phi_1) \in H^1_0(\Omega) \times L^2(\Omega; H^1_{per}(Y)/\mathbb{R})\).

We will denote this product space by \(X\). This is a Hilbert space with inner product

\[
(U, V)_X = (\nabla u, \nabla v)_{L^2(\Omega)} + (\nabla_y u_1, \nabla_y v_1)_{L^2(\Omega \times Y)}
\]

for all \(U = (u, u_1), V = (v, v_1)\) and induced norm

\[
\|U\|_X^2 = \|\nabla u\|_{L^2(\Omega)} + \|\nabla_y u_1\|_{L^2(\Omega \times Y)}.
\]

Let us define now the bilinear form

\[
a[U, \Phi] = \int_{\Omega} \int_Y \langle A(y) \cdot (\nabla_x u + \nabla_y u_1), \nabla_x \phi_0 + \nabla_y \phi_1 \rangle \, dy \, dx,
\]

with \(\Phi := (\phi_0, \phi_1)\). We have to check that this bilinear form is continuous and coercive. Let us start with continuity. We use the \(L^\infty\) bound on \(A(y)\), together with the Cauchy–Schwartz inequality to obtain:

\[
a[U, \Phi] \leq \beta \int_{\Omega} \int_Y \langle \nabla_x u + \nabla_y u_1, \nabla_x \phi_0 + \nabla_y \phi_1 \rangle \, dy \, dx \leq C \|U\|_X \|\Phi\|_X.
\]

We proceed with coercivity. We use the fact that the integral of the derivative of a periodic function over the unit cell vanishes to obtain:

\[
a[U, U] = \int_{\Omega} \int_Y \langle A(y) \cdot (\nabla_x u + \nabla_y u_1), \nabla_x u + \nabla_y u_1 \rangle \, dy \, dx \\
\geq \alpha \int_{\Omega} \int_Y |\nabla_x u + \nabla_y u_1|^2 \, dy \, dx \\
= \alpha \left( \int_{\Omega} \int_Y |\nabla_x u|^2 \, dy \, dx + 2 \int_{\Omega} \int_Y \nabla_x u \cdot \nabla_y u_1 \, dy \, dx + \int_{\Omega} \int_Y |\nabla_y u_1|^2 \, dy \, dx \right) \\
= \alpha \left( \|u\|_X^2 + \|\nabla_y u_1\|_{L^2(\Omega \times Y)}^2 \right) = \alpha \|U\|_X^2
\]

and consequently

\[
a[U, U] \geq \alpha \|U\|_X^2.
\]

Hence, the bilinear form \(a[U, \Phi]\) continuous and coercive and the Lax–Milgram Lemma applies. This proves existence and uniqueness of solutions of the two-scale system in \(X\).

Now we can conclude the proof of Theorem 19.1. \(\square\)
Lemma 19.5. Consider the unique solution \((u, u_1) \in H_0^1(\Omega) \times L^2(\Omega; H^1_{\text{per}}(Y)/\mathbb{R})\) of the two–scale system (19.2.2). Then \(u\) is the unique solution of the homogenized equation (19.1.2) and \(u_1(x, y)\) is of the form
\[
 u_1(x, y) = \chi(y) \cdot \nabla u(x),
\] (19.2.6)
where \(\chi(y)\) is the solution of the cell problem (19.1.4).

Proof. We substitute (19.2.6) into (19.2.2a) to obtain
\[
 -\nabla_y \cdot (A(y) \nabla_y \chi(y)) \nabla_x u = \nabla_y A(y) \cdot \nabla_x u.
\]
This equation is satisfied provided that \(\chi(y) \in (H^1_{\text{per}}(Y)/\mathbb{R})^d\) is the unique solution of the cell problem. Now equation (19.2.2b) becomes
\[
 -\nabla_x \left( \int_Y A(y) \left( \nabla_x u - (\nabla_y \chi) \nabla_x u \right) \, dy \right) = f.
\]
This is precisely the homogenized equation with the homogenized coefficients given by (19.1.3).

The above three lemmas provide us with the proof of Theorem 19.1.

Remark 19.6. The fact that the choice (19.2.6) for \(u_1\) enables us to solve the two–scale system, provided that \(u_0\) satisfies the homogenized equation, implies that this is the only possible set of functions \(\{u, u_1\}\) which solves the two–scale system, since we have already proved uniqueness of solutions.

19.3 Proof of the Corrector Result

Theorem 19.1 implies that \(u^\varepsilon\) converges to \(u(x)\) strongly in \(L^2(\Omega)\). Thus, in order to prove Theorem 19.2, it is enough to prove that
\[
 \lim_{\varepsilon \to 0} \left\| \nabla u^\varepsilon(x) - \nabla \left( u(x) + \varepsilon u_1 \left( x, \frac{x}{\varepsilon} \right) \right) \right\|_{L^2(\Omega; \mathbb{R}^d)} = 0,
\]
or, equivalently,
\[
 \lim_{\varepsilon \to 0} \left\| \nabla u^\varepsilon(x) - \left( \nabla u(x) + \varepsilon \nabla u_1 \left( x, \frac{x}{\varepsilon} \right) + \nabla_y u_1 \left( x, \frac{x}{\varepsilon} \right) \right) \right\|_{L^2(\Omega; \mathbb{R}^d)} = 0.
\]
Assuming now enough regularity on \(u_1\) so that it can be considered to be an admissible test function we have that \(\|\varepsilon \nabla u_1 \left( x, \frac{x}{\varepsilon} \right)\|_{L^2(\Omega; \mathbb{R}^d)} \to 0\). Hence, it is enough to prove that
\[
 \lim_{\varepsilon \to 0} \left\| \nabla u^\varepsilon(x) - \left( \nabla u(x) + \nabla_y u_1 \left( x, \frac{x}{\varepsilon} \right) \right) \right\|_{L^2(\Omega; \mathbb{R}^d)} = 0.
\]
The uniform ellipticity of $A$ now implies:

$$\alpha \| \nabla u^\varepsilon (x) - \left( \nabla u(x) + \nabla_y u_1 \left( x, \frac{x}{\varepsilon} \right) \right) \|_{L^2(\Omega; \mathbb{R}^d)}$$

$$= \alpha \int_{\Omega} \left| \nabla u^\varepsilon (x) - \left( \nabla u(x) + \nabla_y u_1 \left( x, \frac{x}{\varepsilon} \right) \right) \right|^2 dx$$

$$\leq \int_{\Omega} \langle A^\varepsilon(x) \cdot (\nabla_x u^\varepsilon - \nabla_x u - \nabla_y u_1), \nabla_x u^\varepsilon - \nabla_x u - \nabla_y u_1 \rangle dx$$

$$\leq \int_{\Omega} \langle A^\varepsilon(x) \cdot \nabla_x u^\varepsilon(x), \nabla_x u^\varepsilon(x) \rangle dx$$

$$+ \int_{\Omega} \langle A^\varepsilon(x) \cdot (\nabla_x u + \nabla_y u_1), \nabla_x u + \nabla_y u_1 \rangle dx dx$$

$$- \int_{\Omega} \langle (A^\varepsilon(x) + (A^\varepsilon(x))^T) \cdot \nabla_x u^\varepsilon, \nabla_x u + \nabla_y u_1 \rangle dx$$

$$\to \langle f, u \rangle + \int_{\Omega} \int_{\mathcal{Y}} \langle (A(y) \cdot (\nabla_x u(x) + \nabla_y u_1(x, y)), \nabla_x u(x) + \nabla_y u_1(x, y) \rangle dy dx$$

$$+ \int_{\Omega} \int_{\mathcal{Y}} \langle (A(y) + (A(y))^T) \cdot (\nabla_x u(x) + \nabla_y u_1(x, y)), \nabla_x u(x) + \nabla_y u_1(x, y) \rangle dy dx$$


Consequently

$$\lim_{\varepsilon \to 0} \left\| \nabla u^\varepsilon (x) - \left( \nabla u(x) + \nabla_y u_1 \left( x, \frac{x}{\varepsilon} \right) \right) \right\|_{L^2(\Omega; \mathbb{R}^d)} = 0$$

and the theorem is proved \qed

### 19.4 Discussion and Bibliography

The proofs of Theorems 19.1 and 19.2 are taken from [4]. The idea of using appropriate test functions for characterizing the limit of a sequence of functions is very common in the theory of PDEs. See [42] and the references therein. In the context of homogenization test functions of the form (19.2.1) have been used by Kurtz in [85]. See also the perturbed test function approach of Evans, [44, 43].

Tartar’s method of oscillating test functions is based on constructing appropriate test functions using the cell problem. See [143, 144], [30, Ch. 8] and the references therein.

It is not always possible to decouple the two–scale system and to obtain a closed equation for the first term in the two–scale expansion, the homogenized
238CHAPTER 19. HOMOGENIZATION FOR ELLIPTIC PDE: THE CONVERGENCE THEOREM

equation. In order to do this we need an $H^1$–estimate on the solution $u^\varepsilon(x)$ of the unhomogenized PDE. This enables us to conclude that the two–scale limit is independent of the microscale and consequently a homogenized equation actually exists. This is not always the case. See for example the case of linear transport PDE studied in Chapter 21. Other examples can be found in [4, 3].

19.5 Exercises

1. Let $a(y) : \mathbb{T}^d \rightarrow \mathbb{R}^d$ be a smooth, mean zero, 1–periodic, divergence–free field and consider the problem of homogenization for the steady–state advection–diffusion equation

\[
-\Delta u^\varepsilon + \frac{1}{\varepsilon} a \left( \frac{x}{\varepsilon} \right) \cdot \nabla u^\varepsilon = f, \quad \text{for } x \in \Omega, \tag{19.5.1a}
\]

\[
u^\varepsilon(x) = 0, \quad \text{for } x \in \partial \Omega. \tag{19.5.1b}
\]

Use the method of two–scale convergence to prove the homogenization theorem for this PDE.

2. Same as in the previous exercise for the PDE

\[
-\Delta u^\varepsilon + \frac{1}{\varepsilon} a \left( \frac{x}{\varepsilon} \right) \cdot \nabla u^\varepsilon + \frac{1}{\varepsilon} V \left( \frac{x}{\varepsilon} \right) u^\varepsilon = f, \quad \text{for } x \in \Omega,
\]

\[
u^\varepsilon(x) = 0, \quad \text{for } x \in \partial \Omega,
\]

where $V(y)$ is smooth, 1–periodic, mean zero with $\inf_{y \in \mathbb{T}^d} V(y) > 0$.

3. Same as in the previous exercise for the PDE (19.5.1) with

\[
a(y) = -\nabla_y V(y),
\]

where $V(y)$ is smooth and 1–periodic.

4. Use two–scale convergence to prove the homogenization theorem for the Neumann problem.
Chapter 20

Homogenization for Parabolic PDE: The Convergence Theorem

20.1 Introduction

In this chapter we prove the homogenization theorem for second order parabolic PDE of the form studied in Chapter 13. The method of proof is structurally very similar to that used in Chapter 16 where we prove an averaging theorem for Markov chains. To be precise, we obtain an equation for the error in a multiscale expansion, and directly estimate the error from this equation. The crucial estimate used in Chapter 16 follows from the fact that $Q$ in that chapter is a stochastic matrix; the analogous estimate in this chapter follows from the maximum principle for parabolic PDE.

20.2 The Theorem

Theorem 20.1. Let $u^\varepsilon(x, t)$ be the solution of

\[
\frac{\partial u^\varepsilon}{\partial t} = \frac{1}{\varepsilon} b\left(\frac{x}{\varepsilon}\right) \cdot \nabla u^\varepsilon + D \Delta u^\varepsilon \quad (x, t) \in \mathbb{R}^d \times (0, T),
\]

(20.2.1a)

\[
u^\varepsilon = f(x) \quad (x, t) \in \mathbb{R}^d \times \{0\},
\]

(20.2.1b)

with $b(y) \in C^\infty_{\text{per}}(\mathbb{T}^d)$ and $f(x) \in C^\infty_{b}(\mathbb{R}^d)$. Let $u(x, t)$ be the solution of the homogenized equation

\[
\frac{\partial u}{\partial t} = K : \nabla \nabla u \quad (x, t) \in \mathbb{R}^d \times (0, T),
\]

(20.2.2a)
\[
\begin{align*}
\text{CHAPTER 20. HOMOGENIZATION FOR PARABOLIC PDE: THE CONVERGENCE THEOREM}
\end{align*}
\]

\[
\begin{align*}
u &= f(x) \quad (x, t) \in \mathbb{R}^d \times \{0\}, \quad (20.2.2b) \\
\text{where } K \text{ is given by } (13.3.1). \text{ Then}
\| u^\varepsilon (x, t) - u(x, t) \|_{L^\infty (\mathbb{R}^d \times (0, T))} \leq C \varepsilon. \quad (20.2.3)
\end{align*}
\]

Thus \( u^\varepsilon \to u \) in \( L^\infty (\mathbb{R}^d \times (0, T)) \).

20.3 The Proof

In Chapter 13 we derived the two–scale expansion

\[
\begin{align*}
u^\varepsilon (x, t) &\approx u(x, t) + \varepsilon u_1 \left( x, \frac{x}{\varepsilon}, t \right) + \varepsilon^2 u_2 \left( x, \frac{x}{\varepsilon}, t \right),
\end{align*}
\]

where

\[
\begin{align*}
u_1 \left( x, \frac{x}{\varepsilon}, t \right) &= \chi \left( \frac{x}{\varepsilon} \right) \cdot \nabla x u(x, t) . \quad (20.3.1)
\end{align*}
\]

An analysis similar to the one presented in Chapter 12 enables us to obtain the expressions for \( u_2 \) (see Exercise 1, Chapter 13) and

\[
\begin{align*}
u_2 \left( x, \frac{x}{\varepsilon}, t \right) &= \Theta \left( \frac{x}{\varepsilon} \right) : \nabla x \nabla x u(x, t) . \quad (20.3.2)
\end{align*}
\]

The vector field \( \chi(y) \) and the matrix \( \Theta(y) \) solve the Poisson equations

\[
\begin{align*}
- \mathcal{L}_0 \chi(y) &= b(y) \\
- \mathcal{L}_0 \Theta &= b(y) \otimes \chi(y) + 2D \nabla y \chi(y) \\
&\quad - \int_{\mathcal{Y}} \left( b(y) \otimes \chi(y) + 2D \nabla y \chi(y) \right) \rho(y) \, dy , \quad (20.3.3)
\end{align*}
\]

with periodic boundary conditions. Here \( \rho(y) \) is the invariant distribution given by (13.2.5). The corrector fields \( \chi(y) \), \( \Theta(y) \) solve uniformly elliptic PDE with smooth coefficients and periodic boundary conditions. Consequently, both \( \chi(y) \) and \( \Theta(y) \) together with all their derivatives are bounded:

\[
\begin{align*}
\| \chi(y) \|_{C^k(\mathbb{R}^d; \mathbb{R}^d)} &\leq C, \quad \| \Theta(y) \|_{C^k(\mathbb{R}^d; \mathbb{R}^{d \times d})} \leq C, \quad (20.3.4)
\end{align*}
\]

for every integer \( k > 0 \). Furthermore, our assumptions on the initial conditions \( f(x) \) imply that \( u(x, t) \), which solves a PDE with constant coefficients, is bounded.
in \( L^\infty(\mathbb{R}^d) \), together with all of its derivatives with respect to space and time. This, together with estimates (20.3.4) provide us with the bounds

\[
\|u_1^\varepsilon(x,t)\|_{L^\infty((0,T)\times\mathbb{R}^d)} \leq C, \quad \|u_2^\varepsilon(x,t)\|_{L^\infty((0,T)\times\mathbb{R}^d)} \leq C, \quad (20.3.5)
\]

with the constant \( C \) being independent of \( \varepsilon \). In writing the above we use the notation \( u_i^\varepsilon(x,t) := u_i(x,x/\varepsilon,t), \ i = 1, 2. \)

Let \( R^\varepsilon(x,t) \) denote the remainder defined through the equation

\[
u^\varepsilon(x,t) = u(x,t) + \varepsilon u_1 \left( x, \frac{x}{\varepsilon}, t \right) + \varepsilon^2 u_2 \left( x, \frac{x}{\varepsilon}, t \right) + R^\varepsilon(x,t). \quad (20.3.6)
\]

Let

\[
\mathcal{L}^\varepsilon = \frac{1}{\varepsilon} \left( b(y) \cdot \nabla y + D \Delta y \right) + \frac{1}{\varepsilon} \left( b(y) \cdot \nabla x + 2D \nabla x \cdot \nabla y \right) + D \Delta x
\]

with \( y = \frac{x}{\varepsilon} \) and where

\[
\mathcal{L}_0 = b(y) \cdot \nabla y + D \Delta y,
\]

\[
\mathcal{L}_1 = b(y) \cdot \nabla x + 2D \nabla x \cdot \nabla y,
\]

\[
\mathcal{L}_2 = D \Delta x.
\]

Recall that \( u_0, u_1 \) and \( u_2 \) are constructed so that

\[
\mathcal{O}(\frac{1}{\varepsilon^2}) \quad \mathcal{L}_0 u_0 = 0,
\]

\[
\mathcal{O}(\frac{1}{\varepsilon}) \quad \mathcal{L}_0 u_1 = -\mathcal{L}_1 u_0,
\]

\[
\mathcal{O}(1) \quad \mathcal{L}_0 u_2 = -\mathcal{L}_1 u_1 - \mathcal{L}_2 u_0 + \frac{\partial u_0}{\partial t}.
\]

We apply \( \mathcal{L}^\varepsilon \) to the expansion (20.3.6) to obtain

\[
\mathcal{L}^\varepsilon u^\varepsilon = \mathcal{L}^\varepsilon \left( u + \varepsilon u_1 + \varepsilon^2 u_2 \right)
\]

\[
= \frac{1}{\varepsilon^2} \mathcal{L}_0 u + \frac{1}{\varepsilon} \left( \mathcal{L}_0 u + \mathcal{L}_1 u \right) + \left( \mathcal{L}_0 u_2 + \mathcal{L}_1 u_1 + \mathcal{L}_2 u \right)
\]

\[
+ \varepsilon \left( \mathcal{L}_1 u_2 + \mathcal{L}_2 u_1 \right) + \varepsilon^2 \mathcal{L}_2 u_2 + \mathcal{L}^\varepsilon R^\varepsilon
\]

\[
= \frac{\partial u}{\partial t} + \varepsilon \left( \mathcal{L}_1 u_2 + \mathcal{L}_2 u_1 \right) + \varepsilon^2 \mathcal{L}_2 u_2 + \mathcal{L}^\varepsilon R^\varepsilon.
\]
On the other hand
\[
\frac{\partial u^\varepsilon}{\partial t} = \frac{\partial u}{\partial t} + \varepsilon \frac{\partial u_1}{\partial t} + \varepsilon^2 \frac{\partial u_2}{\partial t} + \frac{\partial R^\varepsilon}{\partial t}.
\]
We combine the above two equations together with the unhomogenized equation to obtain
\[
\frac{\partial R^\varepsilon}{\partial t} = \mathcal{L} R^\varepsilon + \varepsilon F^\varepsilon(x, t),
\]
with
\[
F^\varepsilon(x, t) = \mathcal{L}_1 u_2 + \mathcal{L}_2 u_1 - \frac{\partial u_1}{\partial t} + \varepsilon \left( \mathcal{L}_2 u_2 - \frac{\partial u_2}{\partial t} \right).
\] (20.3.7)
Furthermore
\[
u^\varepsilon(x, 0) = f(x) = u(x, 0) + \varepsilon u_1 \left( x, \frac{x}{\varepsilon}, 0 \right) + \varepsilon^2 u_2 \left( x, \frac{x}{\varepsilon}, 0 \right) + R^\varepsilon(x, 0),
\]
and consequently, on account of (20.2.1b) we have
\[
R^\varepsilon(x, 0) = \varepsilon h^\varepsilon(x)
\]
with
\[
h^\varepsilon(x, 0) = u_1 \left( x, \frac{x}{\varepsilon}, 0 \right) + \varepsilon u_2 \left( x, \frac{x}{\varepsilon}, 0 \right). \quad (20.3.8)
\]
Putting the above calculations together we obtain the following Cauchy problem for the remainder \( R^\varepsilon(x, t) \)
\[
\frac{\partial R^\varepsilon}{\partial t} = \mathcal{L} R^\varepsilon + \varepsilon F^\varepsilon(x, t) \quad (x, t) \in \mathbb{R}^d \times (0, T),
\]
(20.3.9a)
\[
R^\varepsilon = \varepsilon h^\varepsilon(x) \quad (x, t) \in \mathbb{R}^d \times \{0\},
\]
(20.3.9b)
with \( F^\varepsilon(x, t) \) and \( h^\varepsilon(x) \) given by (20.3.7) and (20.3.8), respectively.

In order to prove Theorem 20.1 we will need estimates on \( F^\varepsilon \) and \( h^\varepsilon \).

**Lemma 20.2.** Under assumptions \( F^\varepsilon(x, t) \) and \( h^\varepsilon(x, t) \) satisfy
\[
\| F^\varepsilon(x, t) \|_{L^\infty(\mathbb{R}^d \times (0, T))} \leq C \quad (20.3.10)
\]
and
\[
\| h^\varepsilon(x) \|_{L^\infty(\mathbb{R}^d)} \leq C,
\]
(20.3.11)
respectively, where the constant \( C \) is independent of \( \varepsilon \).
Proof. We have

\[
F^\varepsilon(x, t) = L_1 \left( \Theta \left( \frac{x}{\varepsilon} \right) : \nabla_x \nabla_x u(x, t) \right) + L_2 \left( \chi \left( \frac{x}{\varepsilon} \right) \cdot \nabla_x u(x, t) \right) \\
- \frac{\partial}{\partial t} \left( \chi \left( \frac{x}{\varepsilon} \right) \cdot \nabla_x u(x, t) \right) \\
+ \varepsilon \left( D \Delta_x \Theta \left( \frac{x}{\varepsilon} \right) : \nabla_x \nabla_x u(x, t) - \frac{\partial}{\partial t} \left( \Theta \left( \frac{x}{\varepsilon} \right) : \nabla_x \nabla_x u(x, t) \right) \right) \\
= C^\varepsilon(x) : D^3_x u(x, t) + D^\varepsilon(x) : D^4_x u(x, t) \\
+ E^\varepsilon(x) : D^2_x \frac{\partial u}{\partial t}(x, t) + H^\varepsilon(x) \cdot \nabla_x \frac{\partial u}{\partial t}(x, t),
\]

where the functions \( C^\varepsilon(x), D^\varepsilon(x) \) and \( E^\varepsilon(x) \) involve \( \chi \left( \frac{x}{\varepsilon} \right), \Theta \left( \frac{x}{\varepsilon} \right) \) and their derivatives. Estimates combined with (20.3.12), imply (20.3.10).

Furthermore

\[
h^\varepsilon(x) = \chi \left( \frac{x}{\varepsilon} \right) \cdot \nabla_x u(x, 0) + \varepsilon \Theta \left( \frac{x}{\varepsilon} \right) : \nabla_x \nabla_x u(x, 0).
\]

The uniform estimates on \( \chi(y) \) and \( \Theta(y) \), together with our assumptions on the initial conditions of (20.2.1) lead to estimate (20.3.11).

From equation (20.3.9a), the estimate (7.6.4) from Chapter 7 gives

\[
\|R^\varepsilon\|_{L^\infty(\mathbb{R}^d \times (0, T))} \leq \varepsilon \|h^\varepsilon\|_{L^\infty(\mathbb{R}^d)} + \varepsilon \int_0^T \|F^\varepsilon(\cdot, s)\|_{L^\infty(\mathbb{R}^d)} \, ds \\
\leq \varepsilon C + \varepsilon CT \\
\leq C \varepsilon.
\]

We combine (20.3.6) with (20.3.13) and (20.3.5) and use the triangle inequality to obtain

\[
\|u^\varepsilon - u\|_{L^\infty(\mathbb{R}^d \times (0, T))} = \|\varepsilon u_1 + \varepsilon^2 u_2 + R^\varepsilon\|_{L^\infty(\mathbb{R}^d \times (0, T))} \\
\leq \varepsilon \|u_1\|_{L^\infty(\mathbb{R}^d \times (0, T))} + \varepsilon^2 \|u_2\|_{L^\infty(\mathbb{R}^d \times (0, T))} + \|R^\varepsilon\|_{L^\infty(\mathbb{R}^d \times (0, T))} \\
\leq C \varepsilon,
\]

from which (20.2.3) follows.

20.4 Discussion and Bibliography

Our proof of the homogenization theorem has relied on the maximum principle. This is the simplest approach, provided that one is willing to assume sufficient
regularity on the coefficients of the PDE, and it is very well suited for PDE in unbounded domains. However, other techniques may be used to study homogenization for parabolic equations. These include probabilistic methods [121], energy methods [17, Ch. 2] and the method of two-scale convergence [3].

The method employed in this chapter, namely the derivation of a PDE for the error term and the use a priori estimates to control can be applied to various problems in the theory of singular perturbations. Some other examples can be found in [115, 120]. The method is usually called bootstrapping.

20.5 Exercises

1. Consider the case where $\nabla \cdot b(x) = 0$. Assume that the solution of the Cauchy problem is smooth, bounded and decays sufficiently fast at infinity.

   i. Consider the inhomogeneous Cauchy problem

   $\frac{\partial R}{\partial t} = \frac{1}{\varepsilon} b \left( \frac{x}{\varepsilon} \right) \cdot \nabla R + D \Delta R + F(x, t) \quad (x, t) \in \mathbb{R}^d \times (0, T),$

   $R(x, 0) = f(x),$

   where $f(x) \in L^2(\mathbb{R}^d)$ and $F(x, t) \in L^2((0, T) \times \mathbb{R}^d)$. Prove the estimate

   $\|R\|_{L^2((0, T) \times \mathbb{R}^d)}^2 + C_1 \|\nabla R\|_{L^2((0, T) \times \mathbb{R}^d)}^2 \leq C_2 \|f\|_{L^2(\mathbb{R}^d)}^2 + C_3 \|F\|_{L^2((0, T) \times \mathbb{R}^d)}^2.$

   ii. Use this to prove convergence in $L^2((0, T) \times \mathbb{R}^d)$.

   iii. What is the maximum time interval $(0, T)$ over which the homogenization theorem holds?

2. Similarly for the Cauchy problem

   $\frac{\partial u^\varepsilon}{\partial t} = \nabla \cdot \left( A \left( \frac{x}{\varepsilon} \right) \nabla u^\varepsilon \right) \quad (x, t) \in \mathbb{R}^d \times (0, T),$

   $u^\varepsilon(x, 0) = u_{\text{ini}}(x),$

   where the matrix $A(y)$ satisfies the standard periodicity, smoothness and uniform ellipticity assumptions.
3. Consider the initial boundary value problem

\[
\frac{\partial u^\varepsilon}{\partial t} = \nabla \cdot \left( A \left( \frac{x}{\varepsilon} \right) \nabla u^\varepsilon \right) \quad (x, t) \in \Omega \times (0, T),
\]

\[
u^\varepsilon(x, 0) = u_{in}(x), \quad x \Omega
\]

\[
u^\varepsilon(x, t) = 0, \quad (x, t) \in \partial \Omega \times [0, T],
\]

where \( \Omega \) is a bounded domain in \( \mathbb{R}^d \) with smooth boundary and \( A(y) \) satisfies the standard assumptions.

i. Prove the estimate

\[
\| u^\varepsilon \|^2_{L^\infty((0,T)\times\Omega)} + C \| u^\varepsilon \|^2_{L^2((0,T);H^1_0(\Omega))} \leq \| u_{in}(x) \|^2_{L^2(\Omega)}.
\]

ii. Use the above estimate to prove the homogenization theorem using the method of two-scale convergence.

iii. Can you apply the bootstrapping method to this problem?
Chapter 21

Homogenization for Transport PDE: The Convergence Theorem

21.1 Introduction

In this chapter we prove the homogenization theorem for linear transport equations with periodic, divergence–free velocity fields. We show that, when the velocity field does not generate an ergodic flow on the unit torus, the homogenized limit leads to a coupled system of equations, the two–scale system. We use the method of two–scale convergence to prove the main theorem.

21.2 The Theorem

Theorem 21.1. Let \( u^\varepsilon(x, t) \) be the weak solution of

\[
\frac{\partial u^\varepsilon(x, t)}{\partial t} + a \left( \frac{x}{\varepsilon} \right) \cdot \nabla u^\varepsilon(x, t) = 0 \quad \text{for} \quad (x, t) \in \mathbb{R}^d \times \mathbb{R}^+, \tag{21.2.1a}
\]

\[
u^\varepsilon(x, 0) = u_{in}(x) \quad \text{for} \quad x \in \mathbb{R}^d. \tag{21.2.1b}
\]

and assume that \( u_{in}(x) \in L^2(\mathbb{R}^d) \) and that \( a(y) \) is smooth, divergence–free and 1–periodic. Then \( u^\varepsilon(x, t) \) two–scale converges to \( u_0(x, y, t) \in L^2(\mathbb{R}^+ \times \mathbb{R}^d; L^2_{\text{per}}(Y)) \) which satisfies

\[
\int_{\mathbb{R}^+} \int_{\mathbb{R}^d} \int_Y a(y) \cdot \nabla_y \phi(x, y, t) u_0(x, y, t) \, dy \, dx \, dt = 0, \tag{21.2.2a}
\]

\[
a(y) \cdot \nabla_y \phi(x, y, t) = 0, \quad \phi(x, y, t) \in C_0^\infty(\mathbb{R}^+ \times \mathbb{R}^d; C_{\text{per}}^\infty(Y)), \tag{21.2.2b}
\]

247
\[ \int_{\mathbb{R}^+} \int_{\mathbb{R}^d} \int_{\mathbb{Y}} \left( \frac{\partial \phi(x, y, t)}{\partial t} + a(y) \cdot \nabla_x \phi(x, y, t) \right) u_0(x, y, t) \, dy \, dx \, dt + \int_{\mathbb{R}^d} u_{\text{in}}(x) \left( \int_{\mathbb{Y}} \phi(x, y, 0) \, dy \right) \, dx = 0. \] (21.2.2c)

We will refer to equations (21.2.2) as the two-scale system. Notice that the two-scale system involves an auxiliary function \( \phi(x, y, t) \in C_0^\infty(\mathbb{R}^+ \times \mathbb{R}^d; C^\infty_{\text{per}}(\mathbb{Y})) \).

### 21.3 The Proof

**Proof.** We will need appropriate energy estimates on the solution of (21.2.1). Assuming that the solution is smooth and compactly supported, we multiply equation (21.2.1a) by \( u^\varepsilon \), integrate over \( \mathbb{R}^d \) and use the incompressibility of \( a(y) \) to obtain:

\[ \frac{d}{dt} \int_{\mathbb{R}^d} |u^\varepsilon(x, t)|^2 \, dx = 0. \]

We integrate now with respect to time over the interval \([0, t]\) and then take the supremum over \( t \) to obtain:

\[ \|u^\varepsilon(x, t)\|_{L^\infty(\mathbb{R}^+; L^2(\mathbb{R}^d))} \leq \|u_{\text{in}}(x)\|_{L^2(\mathbb{R}^d)} \leq C. \] (21.3.1)

Estimate (21.3.1) implies, by Theorem 2.39, that there exists a subsequence, still denoted by \( u^\varepsilon \), which two-scale converges to a function \( u_0(x, y, t) \in L^2(\mathbb{R}^+ \times \mathbb{R}^d \times \mathbb{Y}) \).

The weak formulation of (21.2.1) is

\[ \int_{\mathbb{R}^+} \int_{\mathbb{R}^d} \left( \frac{\partial \phi^\varepsilon(x, t)}{\partial t} + a\left( \frac{x}{\varepsilon} \right) \cdot \nabla_x \phi^\varepsilon(x, t) \right) u^\varepsilon(x, t) \, dx \, dt + \int_{\mathbb{R}^d} u_{\text{in}}(x) \phi^\varepsilon(x, 0) \, dx = 0, \] (21.3.2)

for every \( \phi^\varepsilon(x, t) \in C_0^\infty(\mathbb{R}^+ \times \mathbb{R}^d) \). We choose an admissible test function of the form

\[ \phi^\varepsilon = \varepsilon \phi_1 \left( x, \frac{x}{\varepsilon}, t \right), \quad \phi_1(x, y, t) \in C_0^\infty(\mathbb{R}^+ \times \mathbb{R}^d; C^\infty_{\text{per}}(\mathbb{Y})). \]

Inserting \( \phi^\varepsilon \) in (21.3.2), using the chain rule (\( \phi^\varepsilon \) is a function of both \( x \) and \( \frac{x}{\varepsilon} \)) and passing to the limit as \( \varepsilon \) tends to 0 we obtain:

\[ \int_{\mathbb{R}^+} \int_{\mathbb{R}^d} \int_{\mathbb{Y}} a(y) \cdot \nabla_y \phi(x, y, t) u_0(x, y, t) \, dy \, dx \, dt = 0. \]
This is equation (21.2.2a). We choose now an admissible test function \( \phi^\varepsilon = \phi \left( x, \frac{x}{\varepsilon}, t \right), \quad \phi(x, y, t) \in C_0^\infty(\mathbb{R}^+ \times \mathbb{R}^d; C^\infty_{\text{per}}(Y)) \) such that (21.2.2b) is satisfied. We insert this test function in (7.7.2) and pass to the limit as \( \varepsilon \) tends to 0 to obtain

\[
\int_{\mathbb{R}^+} \int_{\mathbb{R}^d} \int_Y \left( \frac{\partial \phi(x, y, t)}{\partial t} + a(y) \cdot \nabla_x \phi(x, y, t) \right) u_0(x, y, t) \, dy \, dx \, dt \\
+ \int_{\mathbb{R}^d} u_{in}(x) \left( \int_Y \phi(x, y, 0) \, dy \right) \, dx = 0,
\]

which is precisely (21.2.2c). 

**Remark 21.2.** In the case where \( a(y) \) is ergodic, divergence–free vector field, equation (21.2.2a) is satisfied if and only if \( u_0 \) is independent of \( y \). Then, we can obtain the (weak formulation of) the homogenized equation

\[
\frac{\partial u}{\partial t} + \left( \int_Y a(y) \, dy \right) \cdot \nabla u = 0, \quad u(x, 0) = u_{in}(x)
\]

by choosing a test function which is independent of \( y \).

Of course, in order for the homogenized systems of equations to be of any interest, we need to prove that it is well posed. This is the content of the following theorem.

**Theorem 21.3.** There exists a unique solution \( u_0(x, t, y) \in L^2(\mathbb{R}^+ \times \mathbb{R}^d \times Y) \) of equations (21.2.2a), (21.2.2b) and (21.2.2c).

**Proof.** The existence of a solution follows from the existence of a two–scale limit for \( u^\varepsilon \). Let us proceed with uniqueness. Let \( u_1(x, y, t), \ u_2(x, y, t) \) be two solutions of the homogenized system with the same initial conditions. We form the difference

\[
U(x, y, t) = u_1(x, y, t) - u_2(x, y, t).
\]

This function satisfies, by linearity, the same system of equations (21.2.2a), (21.2.2b) and (21.2.2c), with zero initial conditions. We have:

\[
\int_0^T \int_{\mathbb{R}^d} \int_Y \left( \frac{\partial \phi(x, y, t)}{\partial t} + a(y) \cdot \nabla_x \phi(x, y, t) \right) U(x, y, t) \, dy \, dx \, dt = 0,
\]
where $T > 0$ arbitrary but fixed. Now, $U(x, y, t)$ satisfies equation (21.2.2a) and thus, assuming that it is smooth enough, we can use it as a test function in the above equation to obtain:

$$\int_0^T \int_{\mathbb{R}^d} \int_Y \left( \frac{\partial U(x, y, t)}{\partial t} + a(y) \cdot \nabla_x U(x, y, t) \right) U(x, y, t) \, dy \, dx \, dt = 0.$$ 

Assuming that $U(x, y, t)$ has compact support we can integrate by parts the second term in the above equation to obtain:

$$\int_{\mathbb{R}^d} \int_Y a(y) \cdot \nabla_x U(x, y, t) U(x, y, t) \, dy \, dx \, dt = - \int_{\mathbb{R}^d} \int_Y a(y) U(x, y, t) \cdot \nabla_x U(x, y, t) \, dy \, dx \, dt,$$

from which we deduce that this term vanishes and, thus, we are left with

$$\frac{1}{2} \int_{\mathbb{R}^d} \frac{d}{dt} \int_{\mathbb{R}^d} \int_Y |U(x, y, t)|^2 \, dy \, dx \, dt = 0.$$

We use now the fact that, by assumption, $U(x, y, t)$ is compactly supported to deduce that

$$\int_{\mathbb{R}^d} \int_Y |U(x, y, t)|^2 \, dy \, dx = 0.$$

Consequently, $U(x, y, t) \equiv 0$ and, thus, the solution $u_0(x, y, t)$ is unique. □

21.4 Discussion and Bibliography

In our proof we have followed [36]. A complete (in the $L^2$ sense) set of test functions was used in [70] in order to characterize the homogenized limit in the general two–dimensional case. It was shown there that the homogenized limit was an infinite symmetric set of linear hyperbolic equations. The method of characteristics was used in order to prove the homogenization theorem for two dimensional flows in [145].

21.5 Exercises

1. Consider the following Cauchy problem

$$\frac{\partial u^\varepsilon(x, t)}{\partial t} + a \left( x, \frac{x}{\varepsilon} \right) \cdot \nabla u^\varepsilon(x, t) = 0 \quad \text{for} \ (x, t) \in \mathbb{R}^d \times \mathbb{R}^+,$$
21.5. EXERCISES

\[ u^\varepsilon(x, 0) = u_{in}(x) \quad \text{for } x \in \mathbb{R}^d, \]

where \( a(x, y) \in C^\infty_b(\mathbb{R}^d, C^\infty(\mathcal{Y}); \mathbb{R}^d) \) with \( \nabla \cdot a(x, x/\varepsilon) = 0 \). Use the method of two-scale convergence to prove the homogenization theorem.

2. Similarly for the forced transport PDE

\[
\frac{\partial u^\varepsilon(x, t)}{\partial t} + a(x) \cdot \nabla u^\varepsilon(x, t) = f \left( x, \frac{x}{\varepsilon} \right) \quad \text{for } (x, t) \in \mathbb{R}^d \times \mathbb{R}^+, 
\]

\[ u^\varepsilon(x, 0) = u_{in}(x) \quad \text{for } x \in \mathbb{R}^d, \]

with \( f(x, y) \) being smooth, periodic in its second argument and \( f(x, x/\varepsilon) \) is bounded in \( L^2(\mathbb{R}^d) \).

3. Similarly for the transport equation (21.2.1a) with oscillating initial data

\[ u^\varepsilon(x, 0) = u_{in} \left( x, \frac{x}{\varepsilon} \right), \]

where \( u_{in}(x, y) \) being smooth, periodic in its second argument and \( u_{in}(x, x/\varepsilon) \) is bounded in \( L^2(\mathbb{R}^d) \).

4. Combine the above exercises to prove the homogenization theorem for the transport PDE

\[
\frac{\partial u^\varepsilon(x, t)}{\partial t} + a \left( x, \frac{x}{\varepsilon} \right) \cdot \nabla u^\varepsilon(x, t) = f \left( x, \frac{x}{\varepsilon} \right) \quad \text{for } (x, t) \in \mathbb{R}^d \times \mathbb{R}^+, 
\]

\[ u^\varepsilon(x, 0) = u_{in} \left( x, \frac{x}{\varepsilon} \right) \quad \text{for } x \in \mathbb{R}^d, \]
Bibliography


BIBLIOGRAPHY


