

Texts in Applied Mathematics 60

Grigorios A. Pavliotis

Stochastic Processes and Applications

Diffusion Processes, the Fokker-Planck
and Langevin Equations

 Springer

Texts in Applied Mathematics

Volume 60

Editors-in-Chief:

Stuart Antman, College Park, MD, USA
Leslie Greengard, New York, NY, USA
Philip Holmes, Princeton, NJ, USA

Series Editors:

James Keener, Salt Lake City UT, USA
Robert Kohn, New York NY, USA
Paul Newton, Los Angeles CA, USA
Charles Peskin, New York NY, USA
Robert Pego, Pittsburgh PA, USA
Lenya Ryzhik, Stanford CA, USA
Amit Singer, Princeton NJ, USA
Angela Stevens, Münster, Germany
Andrew Stuart, Coventry, UK
Thomas Witelski, Durham NC, USA

More information about this series at <http://www.springer.com/series/1214>

Grigorios A. Pavliotis

Stochastic Processes and Applications

Diffusion Processes, the Fokker-Planck
and Langevin Equations



Springer

Grigorios A. Pavliotis
Department of Mathematics
Imperial College London
London, UK

ISSN 0939-2475 ISSN 2196-9949 (electronic)
ISBN 978-1-4939-1322-0 ISBN 978-1-4939-1323-7 (eBook)
DOI 10.1007/978-1-4939-1323-7
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2014949241

Mathematics Subject Classification: 60H10, 60H35, 60J60, 35Q84, 65C30, 82C05, 82C31

© Springer Science+Business Media New York 2014

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Για την *Julia*, τον *Adrian*-Αργύρη και τον
Lucas-Γιώργο

Γρηγόρης

Preface

The purpose of this book is to present various results and techniques from the theory of stochastic processes and in particular diffusion processes that are useful in the study of stochastic problems in the natural sciences. The focus is mostly on analytical methods, although numerical and statistical techniques for studying stochastic differential equations are also presented. The goal is the development of methods that are applicable to a wide variety of stochastic models that appear in physics, chemistry, and other natural sciences. Some specific applications, such as stochastic resonance and Brownian motors, are also discussed.

Probabilistic and stochastic techniques have become important aspects of mathematical modeling in the natural as well as the social sciences. Stochastic models are used in areas as diverse as climate and atmosphere/ocean science, molecular biology, finance, and the social sciences. It is now recognized that probability theory and stochastic processes are essential for the training of students in applied mathematics, physics, and engineering.

The book is intended for beginning graduate students in these disciplines. Prerequisites include ordinary and partial differential equations, linear algebra, elementary functional analysis, and techniques from applied mathematics such as perturbation theory. Basic knowledge of scientific computing and of a programming language would be helpful. Some familiarity with probability theory and stochastic processes in discrete time would also be helpful, but it is not essential. Some background in statistical physics would also be useful, in particular in the last two chapters.

The first three chapters of the book contain introductory material on stochastic processes, diffusion processes, and stochastic differential equations. Most of this material is quite standard and can be found in many of the standard textbooks on stochastic processes and stochastic differential equations. Even though the material covered in these chapters is quite standard, an effort has been made to introduce techniques for simulating stochastic processes early on, such as the algorithm for simulating Gaussian processes and the Karhunen–Loëve expansion, and to comment on the connection between stochastic processes and statistical mechanics, for example the Green–Kubo formula.

In Chap. 4, the Fokker–Planck equation is studied in detail. Standard techniques for obtaining solutions to the Fokker–Planck equation such as eigenfunction expansions are introduced. More advanced topics such as the study of convergence to equilibrium for diffusion processes using functional inequalities, reversible diffusions, and Markov chain Monte Carlo are also covered. Techniques from operator theory and functional analysis, similar to those used in quantum mechanics, play a prominent role.

Chapter 5 introduces several techniques that are useful in modeling with stochastic differential equations. Colored noise and the Stratonovich stochastic differential equation are discussed, as are numerical methods and statistical inference for stochastic differential equations. An elementary introduction to noise-induced transitions is also presented.

In Chap. 6, we study the Langevin equation and the associated Fokker–Planck equation in phase space. Advanced techniques from analysis such as the theory of hypoellipticity and hypocoercivity are introduced to study this equation, in particular to prove exponentially fast convergence to equilibrium. Such techniques have proved to be very useful in recent years in the study of the problem of convergence to equilibrium for several models that appear in nonequilibrium statistical mechanics. We also study the Langevin equation in harmonic and in periodic potentials, expansion of the solution to the Fokker–Planck equation in Hermite polynomials, and the overdamped and underdamped limits. The analysis of the overdamped and in particular, the underdamped limit, for the Langevin equation is kept at a formal level.

Exit-time problems for diffusion processes are introduced in Chap. 7. Applications such as the calculation of the rate of escape from a potential well, stochastic resonance, and Brownian motors are discussed. These topics enable us to develop analytical techniques that are useful in the study of several stochastic problems that appear in the natural sciences. The emphasis in this chapter is on formal analytical calculations, rather than on the rigorous analysis of exit-time problems.

Chapters 8 and 9 discuss topics at the interface between stochastic processes and nonequilibrium (or, rather, time-dependent) statistical mechanics. The generalized Langevin equation is derived in Chap. 8, starting from a simple model for an open system. This derivation justifies, to a certain extent, the study of the Langevin and overdamped Langevin (Smoluchowski) equations that is presented in earlier chapters. Linear response theory and the Green–Kubo formalism are studied in Chap. 9. The models and techniques presented in these two chapters are useful in both mathematical and computational statistical mechanics.

The appendices include some comments on the notation used in the book and some elementary material on probability theory. This material is very standard and can be found in any standard textbook on probability theory.

All chapters include a section called “Discussion and Bibliography,” in which additional material and references to the literature can be found. Exercises are also included at the end of every chapter. These exercises range from filling in parts of derivations and performing routine calculations to small computational and analytical projects. I hope that the format of the book will make it appropriate for use both

as a textbook and for self-study. Naturally, the choice of the topics covered and the approach taken are influenced by my own research interests and background.

Parts of this book were used for several years as the basis for a course on applied stochastic processes offered to fourth-year undergraduates and Master's students in applied mathematics at the Department of Mathematics, Imperial College, London. They were also used for a course on nonequilibrium statistical mechanics that I taught at the Department of Mathematics, Free University Berlin, in the summer semester of 2012. I thank all the students who attended those courses.

Acknowledgments

Sebastian Krumscheid, Michela Ottobre, Andrew Duncan, and Marc Pradas read earlier drafts of the book and provided me with many useful comments and suggestions. Sebastian Krumscheid also prepared many of the figures in Sect. 5.3. Section 6.2 was written in collaboration with Michela Ottobre. Marc Pradas helped me with editing some of the figures in Chap. 7. Han Cheng Lie, Ben Goddard, Mathieu Majorel, Nikolas Nüsken, Sebastian Reich, Urbain Vaes and Kostas Zygalakis read parts of the book and made many useful remarks. I am grateful to all of them.

Parts of the book was written during the academic year 2011–2012 while I was on sabbatical in Paris and Berlin, particularly in the period March–August 2012 in Berlin. I am thankful to CERMICS, École des Ponts, Paris, France, and the Bio-computing Group, Department of Mathematics, FU Berlin, Germany, for their hospitality and financial support. I am also grateful to the EPSRC for financial support under grants EP/H034587/1 and EP/J009636/1.

I encourage readers to send me comments, suggestions, or lists of typos and errors at e-mail: g.pavliotis@imperial.ac.uk.

London, UK
July 2014

Grigorios A. Pavliotis

Contents

1	Introduction to Stochastic Processes	1
1.1	Definition of a Stochastic Process	1
1.2	Stationary Processes	3
1.3	Brownian Motion	11
1.4	Examples of Stochastic Processes	14
1.5	The Karhunen–Loève Expansion	17
1.6	Discussion and Bibliography	22
1.7	Exercises	23
2	Diffusion Processes	29
2.1	Examples of Markov Processes	29
2.2	Markov Processes and the Chapman–Kolmogorov Equation	33
2.3	The Generator of a Markov Process	38
2.4	Ergodic Markov Processes	40
2.5	Diffusion Processes and the Forward and Backward Kolmogorov Equations	43
2.6	Discussion and Bibliography	51
2.7	Exercises	53
3	Introduction to Stochastic Differential Equations	55
3.1	Introduction	55
3.2	The Itô and Stratonovich Stochastic Integrals	58
3.3	Solutions of Stochastic Differential Equations	64
3.4	Itô’s Formula	66
3.5	Examples of SDEs	73
3.6	The Lamperti Transformation and Girsanov’s Theorem	77
3.7	Linear Stochastic Differential Equations	79
3.8	Discussion and Bibliography	82
3.9	Exercises	84

4	The Fokker–Planck Equation	87
4.1	Basic Properties of the Fokker–Planck Equation	88
4.2	Examples of Diffusion Processes and Their Fokker–Planck Equation	91
4.2.1	Brownian Motion with Absorbing Boundary Conditions	92
4.2.2	Brownian Motion with Reflecting Boundary Condition	93
4.3	Diffusion Processes in One Dimension	99
4.4	The Ornstein–Uhlenbeck Process and Hermite Polynomials	103
4.5	The Smoluchowski Equation	109
4.6	Reversible Diffusions	116
4.7	Eigenfunction Expansions for Reversible Diffusions	120
4.8	Markov Chain Monte Carlo	122
4.9	Reduction to a Schrödinger Operator	125
4.10	Discussion and Bibliography	130
4.11	Exercises	135
5	Modeling with Stochastic Differential Equations	139
5.1	Colored Noise and Stratonovich SDEs	139
5.2	Numerical Solution of Stochastic Differential Equations	145
5.2.1	Numerical Stability and Implicit Schemes	154
5.3	Parameter Estimation for Stochastic Differential Equations	156
5.4	Noise-Induced Transitions	166
5.5	Discussion and Bibliography	172
5.6	Exercises	174
6	The Langevin Equation	181
6.1	The Fokker–Planck Equation in Phase Space	181
6.2	Hypoellipticity and Hypocoercivity	185
6.2.1	Hypoellipticity	186
6.2.2	Hypocoercivity	189
6.3	The Langevin Equation in a Harmonic Potential	195
6.4	Expansion in Hermite Polynomials	202
6.5	The Overdamped and Underdamped Limits for the Langevin Equation	204
6.5.1	The Overdamped Limit	205
6.5.2	The Underdamped Limit	210
6.6	The Langevin Equation in a Periodic Potential	215
6.6.1	The Overdamped and Underdamped Limits of the Diffusion Coefficient	219
6.6.2	Brownian Motion in a Tilted Periodic Potential	226
6.7	Discussion and Bibliography	228
6.8	Exercises	231

7	Exit Problems for Diffusion Processes and Applications	235
7.1	Brownian Motion in a Double-Well Potential	235
7.2	The Mean Exit Time	237
7.2.1	Examples	240
7.3	Escape from a Potential Well	242
7.4	Stochastic Resonance	245
7.5	Brownian Motors	252
7.6	Discussion and Bibliography	261
7.7	Exercises	265
8	Derivation of the Langevin Equation	267
8.1	Open Classical Systems	267
8.2	The Markovian Approximation	273
8.3	Derivation of the Langevin Equation	278
8.4	Discussion and Bibliography	280
8.5	Exercises	281
9	Linear Response Theory for Diffusion Processes	283
9.1	Linear Response Theory	283
9.2	The Fluctuation–Dissipation Theorem	289
9.3	The Green–Kubo Formula	290
9.4	Discussion and Bibliography	294
9.5	Exercises	295
Appendices		297
Appendix A Frequently Used Notation		299
Appendix B Elements of Probability Theory		303
B.1	Basic Definitions from Probability Theory	303
B.2	Random Variables	305
B.3	Conditional Expectation	309
B.4	The Characteristic Function	311
B.5	Gaussian Random Variables	312
B.5.1	Gaussian Measures in Hilbert Spaces	314
B.6	Types of Convergence and Limit Theorems	317
B.7	Discussion and Bibliography	319
References		321
Index		335

Chapter 1

Introduction to Stochastic Processes

In this chapter, we present some basic results from the theory of stochastic processes and investigate the properties of some standard continuous-time stochastic processes. In Sect. 1.1, we give the definition of a stochastic process. In Sect. 1.2, we present some properties of stationary stochastic processes. In Sect. 1.3, we introduce Brownian motion and study some of its properties. Various examples of stochastic processes in continuous time are presented in Sect. 1.4. The Karhunen–Loëve expansion, one of the most useful tools for representing stochastic processes and random fields, is presented in Sect. 1.5. Further discussion and bibliographical comments are presented in Sect. 1.6. Section 1.7 contains exercises.

1.1 Definition of a Stochastic Process

Stochastic processes describe dynamical systems whose time evolution is of a probabilistic nature. The precise definition is given below.¹

Definition 1.1 (Stochastic process). Let T be an ordered set, $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space, and (E, \mathcal{G}) a measurable space. A stochastic process is a collection of random variables $X = \{X_t; t \in T\}$ such that for each fixed $t \in T$, X_t is a random variable from $(\Omega, \mathcal{F}, \mathbb{P})$ to (E, \mathcal{G}) . The set Ω is known as the sample space, where E is the state space of the stochastic process X_t .

The set T can be either discrete, for example the set of positive integers \mathbb{Z}_+ , or continuous, $T = \mathbb{R}_+$. The state space E will usually be \mathbb{R}^d equipped with the σ -algebra of Borel sets.

¹ The notation and basic definitions from probability theory that we will use can be found in Appendix B.

A stochastic process X may be viewed as a function of both $t \in T$ and $\omega \in \Omega$. We will sometimes write $X(t), X(t, \omega)$ or $X_t(\omega)$ instead of X_t . For a fixed sample point $\omega \in \Omega$, the function $X_t(\omega) : T \mapsto E$ is called a sample path (realization, trajectory) of the process X .

Definition 1.2 (finite-dimensional distributions). The finite-dimensional distributions (FDDs) of a stochastic process are the distributions of the E^k -valued random variables $(X(t_1), X(t_2), \dots, X(t_k))$ for an arbitrary positive integer k and arbitrary times $t_i \in T, i \in \{1, \dots, k\}$:

$$F(\mathbf{x}) = \mathbb{P}(X(t_i) \leq x_i, i = 1, \dots, k),$$

with $\mathbf{x} = (x_1, \dots, x_k)$.

From experiments or numerical simulations, we can obtain information only about the FDDs of a process. A natural question arises: are the FDDs of a stochastic process sufficient to determine a stochastic process uniquely? This answer to this question is affirmative for processes with continuous paths, which is the class of stochastic processes that we will study in these notes.²

Definition 1.3. We say that two processes X_t and Y_t are equivalent if they have same FDDs.

Gaussian Stochastic Processes

A very important class of continuous-time processes is that of Gaussian processes, which arise in many applications.

Definition 1.4. A one-dimensional continuous-time Gaussian process is a stochastic process for which $E = \mathbb{R}$ and all the FDDs are Gaussian, i.e., every finite-dimensional vector $(X_{t_1}, X_{t_2}, \dots, X_{t_k})$ is an $\mathcal{N}(\mu_k, K_k)$ random variable for some vector μ_k and a symmetric nonnegative definite matrix K_k for all $k = 1, 2, \dots$ and for all t_1, t_2, \dots, t_k .

From the above definition, we conclude that the FDDs of a Gaussian continuous-time stochastic process are Gaussian with probability distribution function

$$\gamma_{\mu_k, K_k}(\mathbf{x}) = (2\pi)^{-n/2} (\det K_k)^{-1/2} \exp \left[-\frac{1}{2} \langle K_k^{-1}(\mathbf{x} - \mu_k), \mathbf{x} - \mu_k \rangle \right],$$

where $\mathbf{x} = (x_1, x_2, \dots, x_k)$.

² In fact, all we need is for the stochastic process to be *separable*. See the discussion in Sect. 1.6.

It is straightforward to extend the above definition to arbitrary dimensions. A Gaussian process $x(t)$ is characterized by its mean

$$m(t) := \mathbb{E}x(t)$$

and the covariance (or autocorrelation) matrix

$$C(t, s) = \mathbb{E} \left((x(t) - m(t)) \otimes (x(s) - m(s)) \right).$$

Thus, the first two moments of a Gaussian process are sufficient for a complete characterization of the process.

It is not difficult to simulate Gaussian stochastic processes on a computer. Given a random-number generator that generates $\mathcal{N}(0, 1)$ (pseudo)random numbers, we can sample from a Gaussian stochastic process by calculating the square root of the covariance. A simple algorithm for constructing a skeleton of a continuous-time Gaussian process is the following:

- Fix Δt and define $t_j = (j - 1)\Delta t$, $j = 1, \dots, N$.
- Set $X_j := X(t_j)$, and define the Gaussian random vector $X^N = \left\{ X_j^N \right\}_{j=1}^N$. Then $X^N \sim \mathcal{N}(\mu^N, \Gamma^N)$ with $\mu^N = (\mu(t_1), \dots, \mu(t_N))$ and $\Gamma_{ij}^N = C(t_i, t_j)$.
- Then $X^N = \mu^N + \Lambda \mathcal{N}(0, I)$ with $\Gamma^N = \Lambda \Lambda^T$.

We can calculate the square root of the covariance matrix C using either the Cholesky factorization, via the spectral decomposition of C , or the singular value decomposition (SVD).

1.2 Stationary Processes

In many stochastic processes that appear in applications, their statistics remain invariant under time translations. Such stochastic processes are called *stationary*. It is possible to develop a quite general theory of stochastic processes that enjoy this symmetry property. It is useful to distinguish between stochastic processes for which all FDDs are translation-invariant (strictly stationary processes) and processes for which this translation invariance holds only for the first two moments (weakly stationary processes).

Strictly Stationary Processes

Definition 1.5. A stochastic process is called (strictly) stationary if all FDDs are invariant under time translation: for every integer k and for all times $t_i \in T$, the distribution of $(X(t_1), X(t_2), \dots, X(t_k))$ is equal to that of $(X(s+t_1), X(s+t_2), \dots, X(s+t_k))$ for every s such that $s+t_i \in T$ for all $i \in \{1, \dots, k\}$. In other words,

$$\mathbb{P}(X_{t_1+s} \in A_1, X_{t_2+s} \in A_2 \dots X_{t_k+s} \in A_k) = \mathbb{P}(X_{t_1} \in A_1, X_{t_2} \in A_2 \dots X_{t_k} \in A_k), \quad \forall s \in T.$$

Example 1.1. Let Y_0, Y_1, \dots be a sequence of independent identically distributed random variables and consider the stochastic process $X_n = Y_n$. Then X_n is a strictly stationary process (see Exercise 1). Assume, furthermore, that $\mathbb{E}Y_0 = \mu < +\infty$. Then by the strong law of large numbers, Eq. (B.26), we have that

$$\frac{1}{N} \sum_{j=0}^{N-1} X_j = \frac{1}{N} \sum_{j=0}^{N-1} Y_j \rightarrow \mathbb{E}Y_0 = \mu,$$

almost surely. In fact, the *Birkhoff ergodic theorem* states that for every function f such that $\mathbb{E}f(Y_0) < +\infty$, we have

$$\lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{j=0}^{N-1} f(X_j) = \mathbb{E}f(Y_0), \quad (1.1)$$

almost surely. The sequence of iid random variables is an example of an ergodic strictly stationary process.

We will say that a stationary stochastic process that satisfies (1.1) is *ergodic*. For such processes, we can calculate expectation values of observable $\mathbb{E}f(X_t)$ using a single sample path, provided that it is long enough ($N \gg 1$).

Example 1.2. Let Z be a random variable and define the stochastic process $X_n = Z$, $n = 0, 1, 2, \dots$. Then X_n is a strictly stationary process (see Exercise 2). We can calculate the long-time average of this stochastic process:

$$\frac{1}{N} \sum_{j=0}^{N-1} X_j = \frac{1}{N} \sum_{j=0}^{N-1} Z = Z,$$

which is independent of N and does not converge to the mean of the stochastic processes $\mathbb{E}X_n = \mathbb{E}Z$ (assuming that it is finite) or any other deterministic number. This is an example of a nonergodic process.

Second-Order Stationary Processes

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let X_t , $t \in T$ (with $T = \mathbb{R}$ or \mathbb{Z}), be a real-valued random process on this probability space with finite second moment $\mathbb{E}|X_t|^2 < +\infty$ (i.e., $X_t \in L^2(\Omega, \mathbb{P})$ for all $t \in T$). Assume that it is strictly stationary. Then

$$\mathbb{E}(X_{t+s}) = \mathbb{E}X_t, \quad s \in T, \quad (1.2)$$

from which we conclude that $\mathbb{E}X_t$ is constant, and

$$\mathbb{E}((X_{t_1+s} - \mu)(X_{t_2+s} - \mu)) = \mathbb{E}((X_{t_1} - \mu)(X_{t_2} - \mu)), \quad s \in T, \quad (1.3)$$

implies that the *covariance function* depends on the difference between the two times t and s :

$$C(t, s) = C(t - s).$$

This motivates the following definition.

Definition 1.6. A stochastic process $X_t \in L^2$ is called second-order stationary, wide-sense stationary, or weakly stationary if the first moment $\mathbb{E}X_t$ is a constant and the covariance function $\mathbb{E}(X_t - \mu)(X_s - \mu)$ depends only on the difference $t - s$:

$$\mathbb{E}X_t = \mu, \quad \mathbb{E}((X_t - \mu)(X_s - \mu)) = C(t - s).$$

The constant μ is the expectation of the process X_t . Without loss of generality, we can set $\mu = 0$, since if $\mathbb{E}X_t = \mu$, then the process $Y_t = X_t - \mu$ has mean zero. A mean zero process is called a centered process. The function $C(t)$ is the *covariance* (sometimes also called autocovariance) or the *autocorrelation function* of the X_t . Notice that $C(t) = \mathbb{E}(X_t X_0)$, whereas $C(0) = \mathbb{E}X_t^2$, which is finite, by assumption. Since we have assumed that X_t is a real valued process, we have that $C(t) = C(-t)$, $t \in \mathbb{R}$.

Let now X_t be a strictly stationary stochastic process with finite second moment. The definition of strict stationarity implies that $\mathbb{E}X_t = \mu$, a constant, and $\mathbb{E}((X_t - \mu)(X_s - \mu)) = C(t - s)$. Hence, a strictly stationary process with finite second moment is also stationary in the wide sense. The converse is not true, in general. It is true, however, for Gaussian processes: since the first two moments of a Gaussian process are sufficient for a complete characterization of the process, a Gaussian stochastic process is strictly stationary if and only if it is weakly stationary.

Example 1.3. Let Y_0, Y_1, \dots be a sequence of independent identically distributed random variables and consider the stochastic process $X_n = Y_n$. From Example 1.1, we know that this is a strictly stationary process, irrespective of whether Y_0 is such that $\mathbb{E}Y_0^2 < +\infty$. Assume now that $\mathbb{E}Y_0 = 0$ and $\mathbb{E}Y_0^2 = \sigma^2 < +\infty$. Then X_n is a second-order stationary process with mean zero and correlation function $R(k) = \sigma^2 \delta_{k0}$. Notice that in this case, we have no correlation among the values of the stochastic process at different times n and k .

Example 1.4. Let Z be a single random variable and consider the stochastic process $X_n = Z$, $n = 0, 1, 2, \dots$. From Example 1.2, we know that this is a strictly stationary process irrespective of whether $\mathbb{E}|Z|^2 < +\infty$. Assume now that $\mathbb{E}Z = 0$, $\mathbb{E}Z^2 = \sigma^2$. Then X_n becomes a second-order stationary process with $R(k) = \sigma^2$. Note that in this case, the values of our stochastic process at different times are strongly correlated.

We will see later in this chapter that for second-order stationary processes, ergodicity is related to fast decay of correlations. In the first of the examples above, there was no correlation between our stochastic processes at different times, and the stochastic process is ergodic. In contrast, in our second example, there is very strong correlation between the stochastic process at different times, and this process is not ergodic.

Continuity properties of the covariance function are equivalent to continuity properties of the paths of X_t in the L^2 sense, i.e.,

$$\lim_{h \rightarrow 0} \mathbb{E}|X_{t+h} - X_t|^2 = 0.$$

Lemma 1.1. *Assume that the covariance function $C(t)$ of a second-order stationary process is continuous at $t = 0$. Then it is continuous for all $t \in \mathbb{R}$. Furthermore, the continuity of $C(t)$ is equivalent to the continuity of the process X_t in the L^2 sense.*

Proof. Fix $t \in \mathbb{R}$, and (without loss of generality) set $\mathbb{E}X_t = 0$. We calculate

$$\begin{aligned} |C(t+h) - C(t)|^2 &= |\mathbb{E}(X_{t+h}X_0) - \mathbb{E}(X_tX_0)|^2 = \mathbb{E}|((X_{t+h} - X_t)X_0)|^2 \\ &\leq \mathbb{E}(X_0)^2 \mathbb{E}(X_{t+h} - X_t)^2 \\ &= C(0)(\mathbb{E}X_{t+h}^2 + \mathbb{E}X_t^2 - 2\mathbb{E}(X_tX_{t+h})) \\ &= 2C(0)(C(0) - C(h)) \rightarrow 0, \end{aligned}$$

as $h \rightarrow 0$. Thus, continuity of $C(\cdot)$ at 0 implies continuity for all t .

Assume now that $C(t)$ is continuous. From the above calculation, we have

$$\mathbb{E}|X_{t+h} - X_t|^2 = 2(C(0) - C(h)), \quad (1.4)$$

which converges to 0 as $h \rightarrow 0$. Conversely, assume that X_t is L^2 -continuous. Then from the above equation, we get $\lim_{h \rightarrow 0} C(h) = C(0)$. \square

Notice that from (1.4), we immediately conclude that $C(0) > C(h)$, $h \in \mathbb{R}$.

The Fourier transform of the covariance function of a second-order stationary process always exists. This enables us to study second-order stationary processes using tools from Fourier analysis. To make the link between second-order stationary processes and Fourier analysis, we will use Bochner's theorem, which applies to all nonnegative functions.

Definition 1.7. A function $f(x) : \mathbb{R} \mapsto \mathbb{R}$ is called nonnegative definite if

$$\sum_{i,j=1}^n f(t_i - t_j) c_i \bar{c}_j \geq 0 \quad (1.5)$$

for all $n \in \mathbb{N}$, $t_1, \dots, t_n \in \mathbb{R}$, $c_1, \dots, c_n \in \mathbb{C}$.

Lemma 1.2. *The covariance function of a second-order stationary process is a non-negative definite function.*

Proof. Without loss of generality we can assume that the process is mean zero. We will use the notation $X_t^c := \sum_{i=1}^n X_{t_i} c_i$. We have

$$\begin{aligned} \sum_{i,j=1}^n C(t_i - t_j) c_i \bar{c}_j &= \sum_{i,j=1}^n \mathbb{E} X_{t_i} X_{t_j} c_i \bar{c}_j \\ &= \mathbb{E} \left(\sum_{i=1}^n X_{t_i} c_i \sum_{j=1}^n X_{t_j} \bar{c}_j \right) = \mathbb{E} (X_t^c \bar{X}_t^c) \\ &= \mathbb{E} |X_t^c|^2 \geq 0. \end{aligned}$$

□

Theorem 1.1 (Bochner). *Let $C(t)$ be a continuous positive definite function. Then there exists a unique nonnegative measure ρ on \mathbb{R} such that $\rho(\mathbb{R}) = C(0)$ and*

$$C(t) = \int_{\mathbb{R}} e^{i\omega t} \rho(d\omega) \quad \forall t \in \mathbb{R}. \quad (1.6)$$

Let X_t be a second-order stationary process with autocorrelation function $C(t)$ whose Fourier transform is the measure $\rho(d\omega)$. The measure $\rho(d\omega)$ is called the *spectral measure* of the process X_t . In the following, we will assume that the spectral measure is absolutely continuous with respect to the Lebesgue measure on \mathbb{R} with density $S(\omega)$, i.e., $\rho(d\omega) = S(\omega)d\omega$. The Fourier transform $S(\omega)$ of the covariance function is called the *spectral density* of the process:

$$S(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-it\omega} C(t) dt. \quad (1.7)$$

From (1.6), it follows that the autocorrelation function of a mean-zero second-order stationary process is given by the inverse Fourier transform of the spectral density:

$$C(t) = \int_{-\infty}^{\infty} e^{it\omega} S(\omega) d\omega. \quad (1.8)$$

The autocorrelation function of a second-order stationary process enables us to associate a timescale to X_t , the *correlation time* τ_{cor} :

$$\tau_{cor} = \frac{1}{C(0)} \int_0^{\infty} C(\tau) d\tau = \frac{1}{\mathbb{E}(X_0^2)} \int_0^{\infty} \mathbb{E}(X_\tau X_0) d\tau.$$

The slower the decay of the correlation function, the larger the correlation time. Note that when the correlations do not decay sufficiently fast, so that $C(t)$ is not integrable, then the correlation time will be infinite.

Example 1.5. Consider a mean-zero second-order stationary process with correlation function

$$C(t) = C(0)e^{-\alpha|t|}, \quad (1.9)$$

where $\alpha > 0$. We will write $C(0) = \frac{D}{\alpha}$, where $D > 0$. The spectral density of this process is

$$\begin{aligned} S(\omega) &= \frac{1}{2\pi} \frac{D}{\alpha} \int_{-\infty}^{+\infty} e^{-i\omega t} e^{-\alpha|t|} dt \\ &= \frac{1}{2\pi} \frac{D}{\alpha} \left(\int_{-\infty}^0 e^{-i\omega t} e^{\alpha t} dt + \int_0^{+\infty} e^{-i\omega t} e^{-\alpha t} dt \right) \\ &= \frac{1}{2\pi} \frac{D}{\alpha} \left(\frac{1}{-i\omega + \alpha} + \frac{1}{i\omega + \alpha} \right) \\ &= \frac{D}{\pi} \frac{1}{\omega^2 + \alpha^2}. \end{aligned}$$

This function is called the *Cauchy* or *Lorentz* distribution. The correlation time is (we have that $C(0) = D/\alpha$)

$$\tau_{cor} = \int_0^{\infty} e^{-\alpha t} dt = \alpha^{-1}.$$

A real-valued Gaussian stationary process defined on \mathbb{R} with correlation function given by (1.9) is called a stationary *Ornstein–Uhlenbeck process*. We will study such stochastic processes in detail in later chapters. The Ornstein–Uhlenbeck process X_t can be used as a model for the velocity of a Brownian particle. It is of interest to calculate the statistics of the position of the Brownian particle, i.e., of the integral (we assume that the Brownian particle starts at 0)

$$Z_t = \int_0^t Y_s ds. \quad (1.10)$$

The particle position Z_t is a mean-zero Gaussian process. Set $\alpha = D = 1$. The covariance function of Z_t is

$$\mathbb{E}(Z_t Z_s) = 2 \min(t, s) + e^{-\min(t, s)} + e^{-\max(t, s)} - e^{-|t-s|} - 1. \quad (1.11)$$

Ergodic Properties of Second-Order Stationary Processes

Second-order stationary processes have nice ergodic properties, provided that the correlation between values of the process at different times decays sufficiently fast. In this case, it is possible to show that we can calculate expectations by calculating time averages. An example of such a result is the following.

Proposition 1.3. *Let $\{X_t\}_{t \geq 0}$ be a second-order stationary process on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with mean μ and covariance $C(t)$, and assume that $C(t) \in L^1(0, +\infty)$. Then*

$$\lim_{T \rightarrow +\infty} \mathbb{E} \left| \frac{1}{T} \int_0^T X_s ds - \mu \right|^2 = 0. \quad (1.12)$$

For a proof of this result, we will first need the following result, which is a property of symmetric functions.

Lemma 1.4. *Let $C(t)$ be an integrable symmetric function. Then*

$$\int_0^T \int_0^T C(t-s) dt ds = 2 \int_0^T (T-s) C(s) ds. \quad (1.13)$$

Proof. We make the change of variables $u = t - s$, $v = t + s$. The domain of integration in variables t, s is $[0, T] \times [0, T]$. In the variables u, v , it becomes $[-T, T] \times [|u|, 2T - |u|]$. The Jacobian of the transformation is

$$J = \frac{\partial(t, s)}{\partial(u, v)} = \frac{1}{2}.$$

The integral becomes

$$\begin{aligned} \int_0^T \int_0^T C(t-s) dt ds &= \int_{-T}^T \int_{|u|}^{2T-|u|} C(u) J dv du \\ &= \int_{-T}^T (T - |u|) C(u) du \\ &= 2 \int_0^T (T - u) C(u) du, \end{aligned}$$

where the symmetry of the function $C(u)$ was used in the last step. \square

Proof of Theorem 1.3. We use Lemma (1.4) to calculate:

$$\begin{aligned} \mathbb{E} \left| \frac{1}{T} \int_0^T X_s ds - \mu \right|^2 &= \frac{1}{T^2} \mathbb{E} \left| \int_0^T (X_s - \mu) ds \right|^2 \\ &= \frac{1}{T^2} \mathbb{E} \int_0^T \int_0^T (X_t - \mu)(X_s - \mu) dt ds \\ &= \frac{1}{T^2} \int_0^T \int_0^T C(t-s) dt ds \\ &= \frac{2}{T^2} \int_0^T (T - u) C(u) du \\ &\leq \frac{2}{T} \int_0^T \left| \left(1 - \frac{u}{T}\right) C(u) \right| du \leq \frac{2}{T} \int_0^{+\infty} C(u) du \rightarrow 0, \end{aligned}$$

using the dominated convergence theorem and the assumption $C(\cdot) \in L^1(0, +\infty)$. \square

Assume that $\mu = 0$, and define

$$D = \int_0^{+\infty} C(t) dt, \quad (1.14)$$

which, from our assumption on $C(t)$, is a finite quantity.³ The above calculation suggests that for $t \gg 1$, we have

$$\mathbb{E} \left(\int_0^t X(s) ds \right)^2 \approx 2Dt.$$

This implies that at sufficiently long times, the mean square displacement of the integral of the ergodic second-order stationary process X_t scales linearly in time, with proportionality coefficient $2D$. Let now X_t be the velocity of a (Brownian) particle. The particle position Z_t is given by (1.10). From our calculation above, we conclude that

$$\mathbb{E}Z_t^2 = 2Dt,$$

where

$$D = \int_0^\infty C(t) dt = \int_0^\infty \mathbb{E}(X_t X_0) dt \quad (1.15)$$

is the *diffusion coefficient*. Thus, one expects that at sufficiently long times and under appropriate assumptions on the correlation function, the time integral of a stationary process will approximate a Brownian motion with diffusion coefficient D . The diffusion coefficient is an example of a transport coefficient, and (1.15) is an example of the Green–Kubo formula: a transport coefficient can be calculated in terms of the time integral of an appropriate autocorrelation function. In the case of the diffusion coefficient, we need to calculate the integral of the velocity autocorrelation function. We will explore this topic in more detail in Chap. 9.

Example 1.6. Consider the stochastic process with an exponential correlation function from Example 1.5, and assume that this stochastic process describes the velocity of a Brownian particle. Since $C(t) \in L^1(0, +\infty)$, Proposition 1.3 applies. Furthermore, the diffusion coefficient of the Brownian particle is given by

$$\int_0^{+\infty} C(t) dt = C(0)\tau_c = \frac{D}{\alpha^2}.$$

Remark 1.1. Let X_t be a strictly stationary process, and let f be such that $\mathbb{E}(f(X_0))^2 < +\infty$. A calculation similar to the one that we did in the proof of Proposition 1.3 enables us to conclude that

$$\lim_{T \rightarrow +\infty} \frac{1}{T} \int_0^T f(X_s) ds = \mathbb{E}f(X_0), \quad (1.16)$$

in $L^2(\Omega)$. In this case, the autocorrelation function of X_t is replaced by

$$C_f(t) = \mathbb{E}[(f(X_t) - \mathbb{E}f(X_0))(f(X_0) - \mathbb{E}f(X_0))].$$

³ Note, however, that we do not know whether it is nonzero. This requires a separate argument.

1.3 Brownian Motion

The most important continuous-time stochastic process is Brownian motion. Brownian motion is a process with almost surely continuous paths and independent Gaussian increments. A process X_t has independent increments if for every sequence $t_0 < t_1 < \dots < t_n$, the random variables

$$X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}$$

are independent. If furthermore, for every $t_1, t_2, s \in T$ and Borel set $B \subset \mathbb{R}$, we have

$$\mathbb{P}(X_{t_2+s} - X_{t_1+s} \in B) = \mathbb{P}(X_{t_2} - X_{t_1} \in B),$$

then the process X_t has stationary independent increments.

Definition 1.8. A one-dimensional standard *Brownian motion* $W(t) : \mathbb{R}^+ \rightarrow \mathbb{R}$ is a real-valued stochastic process with almost surely (a.s.) continuous paths such that $W(0) = 0$, it has independent increments, and for every $t > s \geq 0$, the increment $W(t) - W(s)$ has a Gaussian distribution with mean 0 and variance $t - s$, i.e., the density of the random variable $W(t) - W(s)$ is

$$g(x; t, s) = \left(2\pi(t-s)\right)^{-\frac{1}{2}} \exp\left(-\frac{x^2}{2(t-s)}\right). \quad (1.17)$$

A standard d -dimensional standard Brownian motion $W(t) : \mathbb{R}^+ \rightarrow \mathbb{R}^d$ is a vector of d independent one-dimensional Brownian motions:

$$W(t) = (W_1(t), \dots, W_d(t)),$$

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

$$g(\mathbf{x}; t, s) = \left(2\pi(t-s)\right)^{-d/2} \exp\left(-\frac{\|\mathbf{x}\|^2}{2(t-s)}\right).$$

Brownian motion is also referred to as a *Wiener process*. In Fig. 1.1, we plot a few sample paths of Brownian motion.

As we have already mentioned, Brownian motion has almost surely continuous paths. More precisely, it has a continuous modification: Consider two stochastic processes X_t and Y_t , $t \in T$, that are defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The process Y_t is said to be a modification of X_t if $\mathbb{P}(X_t = Y_t) = 1$ for all $t \in T$. The fact that there is a continuous modification of Brownian motion follows from the following result, which is due to Kolmogorov.

Theorem 1.2 (Kolmogorov). *Let X_t , $t \in [0, \infty)$, be a stochastic process on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Suppose that there are positive constants α and β and that for each $T \geq 0$, there is a constant $C(T)$ such that*

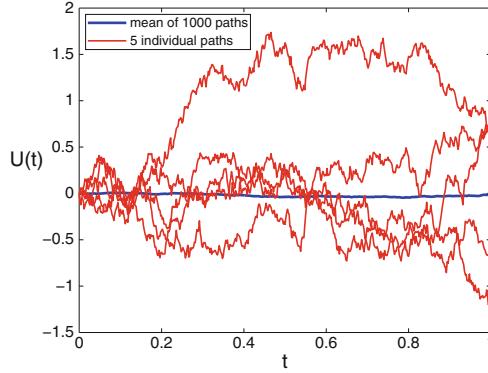


Fig. 1.1 Brownian sample paths

$$\mathbb{E}|X_t - X_s|^\alpha \leq C(T)|t - s|^{1+\beta}, \quad 0 \leq s, t \leq T. \quad (1.18)$$

Then there exists a continuous modification Y_t of the process X_t .

We can check that (1.18) holds for Brownian motion with $\alpha = 4$ and $\beta = 1$ using (1.17). It is possible to prove rigorously the existence of a Wiener process (Brownian motion):

Theorem 1.3 (Wiener). *There exists an almost surely continuous process W_t with independent increments such that $W_0 = 0$ and for each $t \geq 0$, the random variable W_t is $\mathcal{N}(0, t)$. Furthermore, W_t is almost surely locally Hölder continuous with exponent α for every $\alpha \in (0, \frac{1}{2})$.*

Note that Brownian paths are not differentiable.

We can construct Brownian motion through the limit of an appropriately rescaled random walk: Let X_1, X_2, \dots be iid random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with mean 0 and variance 1. Define the discrete-time stochastic process S_n with $S_0 = 0$, $S_n = \sum_{j=1}^n X_j$, $n \geq 1$. Define now a continuous-time stochastic process with continuous paths as the linearly interpolated appropriately rescaled random walk

$$W_t^n = \frac{1}{\sqrt{n}} S_{[nt]} + (nt - [nt]) \frac{1}{\sqrt{n}} X_{[nt]+1},$$

where $[.]$ denotes the integer part of a number. Then W_t^n converges weakly as $n \rightarrow +\infty$ to a one-dimensional standard Brownian motion. See Fig. 1.2.

An alternative definition of the one-dimensional standard Brownian motion is that of a Gaussian stochastic process on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with continuous paths for almost all $\omega \in \Omega$, and FDDs with zero mean and covariance $\mathbb{E}(W_t W_j) = \min(t, t_j)$. One can then show that Definition 1.8 follows from the above definition.

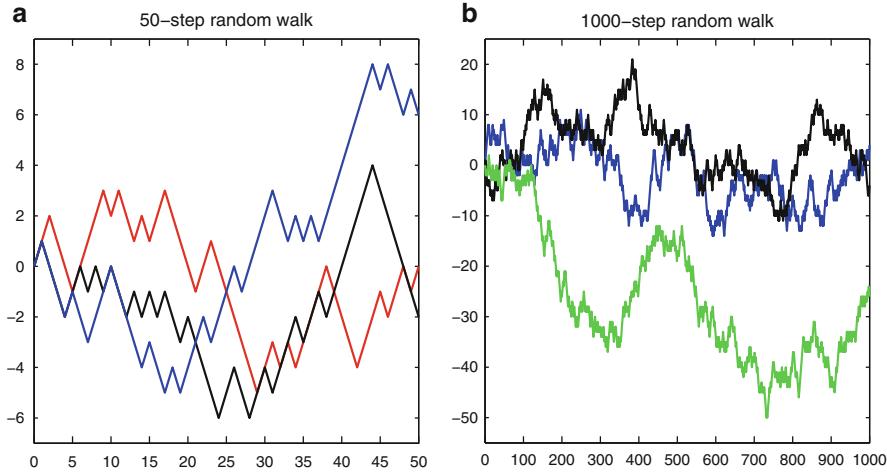


Fig. 1.2 Sample paths of a random walk of length (a) $n = 50$ and (b) $n = 1000$

For the d -dimensional Brownian motion, we have [see (B.7) and (B.8)]

$$\mathbb{E}W(t) = 0 \quad \forall t \geq 0$$

and

$$\mathbb{E}\left((W(t) - W(s)) \otimes (W(t) - W(s))\right) = (t - s)I, \quad (1.19)$$

where I denotes the identity matrix. Moreover,

$$\mathbb{E}\left(W(t) \otimes W(s)\right) = \min(t, s)I. \quad (1.20)$$

Although Brownian motion has stationary increments, it is not itself a stationary process. The probability density of the one-dimensional Brownian motion is

$$g(x, t) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t}.$$

We can easily calculate all moments:

$$\begin{aligned} \mathbb{E}(W(t)^n) &= \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{+\infty} x^n e^{-x^2/2t} dx \\ &= \begin{cases} 1.3 \dots (n-1)t^{n/2}, & n \text{ even}, \\ 0, & n \text{ odd}. \end{cases} \end{aligned}$$

In particular, the mean square displacement of Brownian motion grows linearly in time.

Brownian motion is invariant under various transformations in time.

Proposition 1.5. Let W_t denote a standard Brownian motion in \mathbb{R} . Then W_t has the following properties:

- (i) (Rescaling). For each $c > 0$, define $X_t = \frac{1}{\sqrt{c}}W(ct)$. Then $(X_t, t \geq 0) = (W_t, t \geq 0)$ in law.
- (ii) (Shifting). For each $c > 0$, $W_{c+t} - W_c$, $t \geq 0$, is a Brownian motion that is independent of W_u , $u \in [0, c]$.
- (iii) (Time reversal). Define $X_t = W_{1-t} - W_1$, $t \in [0, 1]$. Then $(X_t, t \in [0, 1]) = (W_t, t \in [0, 1])$ in law.
- (iv) (Inversion). Let X_t , $t \geq 0$, be defined by $X_0 = 0$, $X_t = tW(1/t)$. Then $(X_t, t \geq 0) = (W_t, t \geq 0)$ in law.

The equivalence in the above result holds in law and not in a pathwise sense. The proof of this proposition is left as an exercise.

We can also add a drift and change the diffusion coefficient of the Brownian motion: We will define a Brownian motion with drift μ and variance σ^2 as the process

$$X_t = \mu t + \sigma W_t.$$

The mean and variance of X_t are

$$\mathbb{E}X_t = \mu t, \quad \mathbb{E}(X_t - \mathbb{E}X_t)^2 = \sigma^2 t.$$

Notice that X_t satisfies the equation

$$dX_t = \mu dt + \sigma dW_t.$$

This is an example of a *stochastic differential equation*. We will study stochastic differential equations in Chaps. 3 and 5.

1.4 Examples of Stochastic Processes

We present now a few examples of stochastic processes that appear frequently in applications.

The Ornstein–Uhlenbeck Process

The stationary Ornstein–Uhlenbeck process that was introduced earlier in this chapter can be defined through the Brownian motion via a time change.

Lemma 1.6. *Let $W(t)$ be a standard Brownian motion and consider the process*

$$V(t) = e^{-t}W(e^{2t}).$$

Then $V(t)$ is a Gaussian stationary process with mean 0 and correlation function

$$R(t) = e^{-|t|}. \quad (1.21)$$

For the proof of this result, we first need to show that time-changed Gaussian processes are also Gaussian.

Lemma 1.7. *Let $X(t)$ be a Gaussian stochastic process, and let $Y(t) = X(f(t))$, where $f(t)$ is a strictly increasing function. Then $Y(t)$ is also a Gaussian process.*

Proof. We need to show that for all positive integers N and all sequences of times $\{t_1, t_2, \dots, t_N\}$, the random vector

$$\{Y(t_1), Y(t_2), \dots, Y(t_N)\} \quad (1.22)$$

is a multivariate Gaussian random variable. Since $f(t)$ is strictly increasing, it is invertible, and hence there exist s_i , $i = 1, \dots, N$, such that $s_i = f^{-1}(t_i)$. Thus, the random vector (1.22) can be rewritten as

$$\{X(s_1), X(s_2), \dots, X(s_N)\},$$

which is Gaussian for all N and all choices of times s_1, s_2, \dots, s_N . Hence $Y(t)$ is also Gaussian. \square

Proof of Lemma 1.6. The fact that $V(t)$ is a mean-zero process follows immediately from the fact that $W(t)$ is of mean zero. To show that the correlation function of $V(t)$ is given by (1.21), we calculate

$$\begin{aligned} \mathbb{E}(V(t)V(s)) &= e^{-t-s}\mathbb{E}(W(e^{2t})W(e^{2s})) = e^{-t-s}\min(e^{2t}, e^{2s}) \\ &= e^{-|t-s|}. \end{aligned}$$

That the process $V(t)$ is Gaussian follows from Lemma 1.7 (observe that the transformation that gives $V(t)$ in terms of $W(t)$ is invertible, and we can write $W(s) = s^{1/2}V(\frac{1}{2}\ln(s))$). \square

Brownian Bridge

We can modify Brownian motion so that the resulting process is fixed at both ends. Let $W(t)$ be a standard one-dimensional Brownian motion. We define the Brownian bridge (from 0 to 0) to be the process

$$B_t = W_t - tW_1, \quad t \in [0, 1]. \quad (1.23)$$

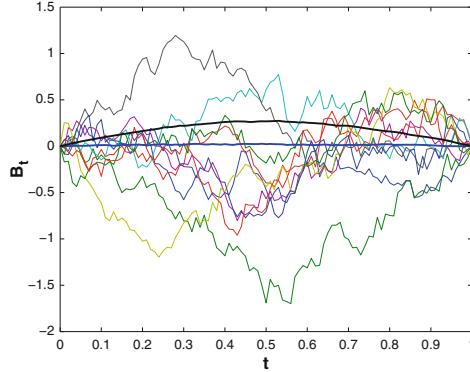


Fig. 1.3 Sample paths and first (blue curve) and second (black curve) moment of the Brownian bridge

Observe that $B_0 = B_1 = 0$. Equivalently, we can define the Brownian bridge to be the continuous Gaussian process $\{B_t : 0 \leq t \leq 1\}$ such that

$$\mathbb{E}B_t = 0, \quad \mathbb{E}(B_t B_s) = \min(s, t), \quad s, t \in [0, 1]. \quad (1.24)$$

An equivalent definition of the Brownian bridge is through an appropriate time change of the Brownian motion:

$$B_t = (1-t)W\left(\frac{t}{1-t}\right), \quad t \in [0, 1]. \quad (1.25)$$

Conversely, we can write the Brownian motion as a time change of the Brownian bridge:

$$W_t = (t+1)B\left(\frac{t}{1+t}\right), \quad t \geq 0.$$

We can use the algorithm for simulating Gaussian processes to generate paths of the Brownian bridge process and to calculate moments. In Fig. 1.3, we plot a few sample paths and the first and second moments of the Brownian bridge.

Fractional Brownian Motion

A fractional Brownian motion is a one-parameter family of Gaussian processes whose increments are correlated.

Definition 1.9. A (normalized) fractional Brownian motion W_t^H , $t \geq 0$, with Hurst parameter $H \in (0, 1)$ is a centered Gaussian process with continuous sample paths whose covariance is given by

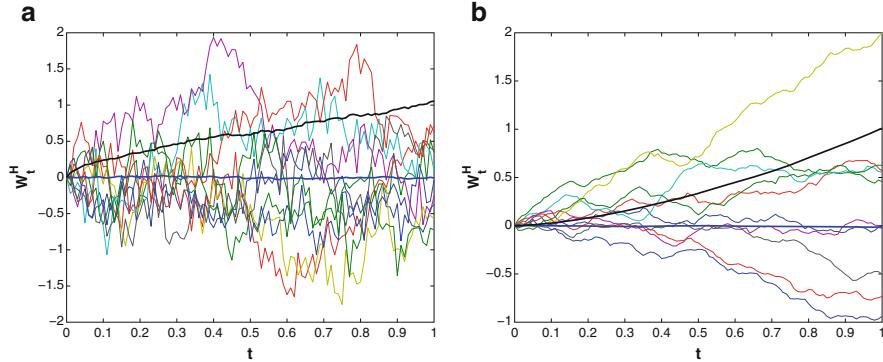


Fig. 1.4 Sample paths of fractional Brownian motion for Hurst exponent **(a)** $H = 0.3$ and **(b)** $H = 0.8$ and first (blue curve) and second (black curve) moment

$$\mathbb{E}(W_t^H W_s^H) = \frac{1}{2} (s^{2H} + t^{2H} - |t-s|^{2H}). \quad (1.26)$$

The Hurst exponent controls the correlations between the increments of a fractional Brownian motion as well as the regularity of the paths: they become smoother as H increases.

Some of the basic properties of fractional Brownian motion are summarized in the following proposition.

Proposition 1.8. *A fractional Brownian motion has the following properties:*

- (i) *When $H = \frac{1}{2}$, then $W_t^{\frac{1}{2}}$ becomes standard Brownian motion.*
- (ii) $W_0^H = 0$, $\mathbb{E}W_t^H = 0$, $\mathbb{E}(W_t^H)^2 = |t|^{2H}$, $t \geq 0$.
- (iii) *It has stationary increments, and $\mathbb{E}(W_t^H - W_s^H)^2 = |t-s|^{2H}$.*
- (iv) *It has the following self-similarity property:*

$$(W_{\alpha t}^H, t \geq 0) = (\alpha^H W_t^H, t \geq 0), \quad \alpha > 0, \quad (1.27)$$

where the equivalence is in law.

The proof of these properties is left as an exercise. In Fig. 1.4, we present sample plots and the first two moments of the fractional Brownian motion for $H = 0.3$ and $H = 0.8$. As expected, for larger values of the Hurst exponent, the sample paths are more regular.

1.5 The Karhunen–Loève Expansion

Let $f \in L^2(\mathcal{D})$, where \mathcal{D} is a subset of \mathbb{R}^d , and let $\{e_n\}_{n=1}^\infty$ be an orthonormal basis in $L^2(\mathcal{D})$. Then it is well known that f can be written as a series expansion

$$f = \sum_{n=1}^{\infty} f_n e_n,$$

where

$$f_n = \int_{\Omega} f(x) e_n(x) dx.$$

The convergence is in $L^2(\mathcal{D})$:

$$\lim_{N \rightarrow \infty} \left\| f(x) - \sum_{n=1}^N f_n e_n(x) \right\|_{L^2(\mathcal{D})} = 0.$$

It turns out that we can obtain a similar expansion for an L^2 mean-zero process that is continuous in the L^2 sense:

$$\mathbb{E}X_t^2 < +\infty, \quad \mathbb{E}X_t = 0, \quad \lim_{h \rightarrow 0} \mathbb{E}|X_{t+h} - X_t|^2 = 0. \quad (1.28)$$

For simplicity, we will take $T = [0, 1]$. Let $R(t, s) = \mathbb{E}(X_t X_s)$ be the autocorrelation function. Notice that from (1.28), it follows that $R(t, s)$ is continuous in both t and s ; see Exercise 19.

Let us assume an expansion of the form

$$X_t(\omega) = \sum_{n=1}^{\infty} \xi_n(\omega) e_n(t), \quad t \in [0, 1], \quad (1.29)$$

where $\{e_n\}_{n=1}^{\infty}$ is an orthonormal basis in $L^2(0, 1)$. The random variables ξ_n are calculated as

$$\int_0^1 X_t e_k(t) dt = \int_0^1 \sum_{n=1}^{\infty} \xi_n e_n(t) e_k(t) dt = \sum_{n=1}^{\infty} \xi_n \delta_{nk} = \xi_k,$$

where we have assumed that we can interchange the summation and integration. We will assume that these random variables are orthogonal:

$$\mathbb{E}(\xi_n \xi_m) = \lambda_n \delta_{nm},$$

where $\{\lambda_n\}_{n=1}^{\infty}$ are positive numbers that will be determined later.

Assuming that an expansion of the form (1.29) exists, we can calculate

$$\begin{aligned} R(t, s) &= \mathbb{E}(X_t X_s) = \mathbb{E} \left(\sum_{k=1}^{\infty} \sum_{\ell=1}^{\infty} \xi_k e_k(t) \xi_{\ell} e_{\ell}(s) \right) \\ &= \sum_{k=1}^{\infty} \sum_{\ell=1}^{\infty} \mathbb{E}(\xi_k \xi_{\ell}) e_k(t) e_{\ell}(s) \\ &= \sum_{k=1}^{\infty} \lambda_k e_k(t) e_k(s). \end{aligned}$$

Consequently, in order for the expansion (1.29) to be valid, we need

$$R(t, s) = \sum_{k=1}^{\infty} \lambda_k e_k(t) e_k(s). \quad (1.30)$$

From Eq. (1.30), it follows that

$$\begin{aligned} \int_0^1 R(t, s) e_n(s) ds &= \int_0^1 \sum_{k=1}^{\infty} \lambda_k e_k(t) e_k(s) e_n(s) ds \\ &= \sum_{k=1}^{\infty} \lambda_k e_k(t) \int_0^1 e_k(s) e_n(s) ds \\ &= \sum_{k=1}^{\infty} \lambda_k e_k(t) \delta_{kn} \\ &= \lambda_n e_n(t). \end{aligned}$$

Consequently, in order for the expansion (1.29) to be valid, the set $\{\lambda_n, e_n(t)\}_{n=1}^{\infty}$ has to be the set of eigenvalues and eigenfunctions of the integral operator whose kernel is the correlation function of X_t :

$$\int_0^1 R(t, s) e_n(s) ds = \lambda_n e_n(t). \quad (1.31)$$

To prove the expansion (1.29), we need to study the eigenvalue problem for the integral operator

$$\mathcal{R}f := \int_0^1 R(t, s) f(s) ds. \quad (1.32)$$

We consider \mathcal{R} as an operator from $L^2[0, 1]$ to $L^2[0, 1]$. We can show that this operator is self-adjoint and nonnegative in $L^2(0, 1)$:

$$\langle \mathcal{R}f, h \rangle = \langle f, \mathcal{R}h \rangle \quad \text{and} \quad \langle \mathcal{R}f, f \rangle \geq 0 \quad \forall f, h \in L^2(0, 1),$$

where $\langle \cdot, \cdot \rangle$ denotes the $L^2(0, 1)$ inner product. It follows that all its eigenvalues are real and nonnegative. Furthermore, it is a compact operator (if $\{\phi_n\}_{n=1}^{\infty}$ is a bounded sequence in $L^2(0, 1)$, then $\{\mathcal{R}\phi_n\}_{n=1}^{\infty}$ has a convergent subsequence). The spectral theorem for compact self-adjoint operators can be used to deduce that \mathcal{R} has a countable sequence of eigenvalues tending to 0. Furthermore, for every $f \in L^2(0, 1)$, we can write

$$f = f_0 + \sum_{n=1}^{\infty} f_n e_n(t),$$

where $\mathcal{R}f_0 = 0$ and $\{e_n(t)\}$ are the eigenfunctions of the operator \mathcal{R} corresponding to nonzero eigenvalues and where the convergence is in L^2 . Finally, Mercer's theorem states that for $R(t, s)$ continuous on $[0, 1] \times [0, 1]$, the expansion (1.30) is valid, where the series converges absolutely and uniformly.

Now we are ready to prove (1.29).

Theorem 1.4 (Karhunen–Loève). *Let $\{X_t, t \in [0, 1]\}$ be an L^2 process with zero mean and continuous correlation function $R(t, s)$. Let $\{\lambda_n, e_n(t)\}_{n=1}^\infty$ be the eigenvalues and eigenfunctions of the operator \mathcal{R} defined in (1.32). Then*

$$X_t = \sum_{n=1}^{\infty} \xi_n e_n(t), \quad t \in [0, 1], \quad (1.33)$$

where

$$\xi_n = \int_0^1 X_t e_n(t) dt, \quad \mathbb{E}\xi_n = 0, \quad \mathbb{E}(\xi_n \xi_m) = \lambda \delta_{nm}. \quad (1.34)$$

The series converges in L^2 to X_t , uniformly in t .

Proof. The fact that $\mathbb{E}\xi_n = 0$ follows from the fact that X_t is of mean zero. The orthogonality of the random variables $\{\xi_n\}_{n=1}^\infty$ follows from the orthogonality of the eigenfunctions of \mathcal{R} :

$$\begin{aligned} \mathbb{E}(\xi_n \xi_m) &= \mathbb{E} \int_0^1 \int_0^1 X_t X_s e_n(t) e_m(s) dt ds \\ &= \int_0^1 \int_0^1 R(t, s) e_n(t) e_m(s) ds dt \\ &= \lambda_n \int_0^1 e_n(s) e_m(s) ds = \lambda_n \delta_{nm}. \end{aligned}$$

Consider now the partial sum $S_N = \sum_{n=1}^N \xi_n e_n(t)$. We have

$$\begin{aligned} \mathbb{E}|X_t - S_N|^2 &= \mathbb{E}X_t^2 + \mathbb{E}S_N^2 - 2\mathbb{E}(X_t S_N) \\ &= R(t, t) + \mathbb{E} \sum_{k, \ell=1}^N \xi_k \xi_\ell e_k(t) e_\ell(t) - 2\mathbb{E} \left(X_t \sum_{n=1}^N \xi_n e_n(t) \right) \\ &= R(t, t) + \sum_{k=1}^N \lambda_k |e_k(t)|^2 - 2\mathbb{E} \sum_{k=1}^N \int_0^1 X_t X_s e_k(s) e_k(t) ds \\ &= R(t, t) - \sum_{k=1}^N \lambda_k |e_k(t)|^2 \rightarrow 0, \end{aligned}$$

by Mercer's theorem. □

The Karhunen–Loève expansion is straightforward to apply to Gaussian stochastic processes. Let X_t be a Gaussian second-order process with continuous covariance $R(t, s)$. Then the random variables $\{\xi_k\}_{k=1}^\infty$ are Gaussian, since they are defined through the time integral of a Gaussian process. Furthermore, since they are Gaussian and orthogonal, they are also independent. Hence, for Gaussian processes, the Karhunen–Loève expansion becomes

$$X_t = \sum_{k=1}^{+\infty} \sqrt{\lambda_k} \xi_k e_k(t), \quad (1.35)$$

where $\{\xi_k\}_{k=1}^{\infty}$ are independent $\mathcal{N}(0, 1)$ random variables.

Example 1.7 (The Karhunen–Loève expansion for Brownian motion). The correlation function of Brownian motion is $R(t, s) = \min(t, s)$. The eigenvalue problem $\mathcal{R}\psi_n = \lambda_n \psi_n$ becomes

$$\int_0^1 \min(t, s) \psi_n(s) ds = \lambda_n \psi_n(t).$$

Let us assume that $\lambda_n > 0$ (we can check that 0 is not an eigenvalue). On setting $t = 0$, we obtain $\psi_n(0) = 0$. The eigenvalue problem can be rewritten in the form

$$\int_0^t s \psi_n(s) ds + t \int_t^1 \psi_n(s) ds = \lambda_n \psi_n(t).$$

We differentiate this equation once:

$$\int_t^1 \psi_n(s) ds = \lambda_n \psi_n'(t).$$

We set $t = 1$ in this equation to obtain the second boundary condition $\psi_n'(1) = 0$. A second differentiation yields

$$-\psi_n(t) = \lambda_n \psi_n''(t),$$

where primes denote differentiation with respect to t . Thus, in order to calculate the eigenvalues and eigenfunctions of the integral operator whose kernel is the covariance function of Brownian motion, we need to solve the Sturm–Liouville problem

$$-\psi_n(t) = \lambda_n \psi_n''(t), \quad \psi_n(0) = \psi_n'(1) = 0.$$

We can calculate the eigenvalues and (normalized) eigenfunctions:

$$\psi_n(t) = \sqrt{2} \sin\left(\frac{1}{2}(2n-1)\pi t\right), \quad \lambda_n = \left(\frac{2}{(2n-1)\pi}\right)^2.$$

Thus, the Karhunen–Loève expansion of Brownian motion on $[0, 1]$ is

$$W_t = \sqrt{2} \sum_{n=1}^{\infty} \xi_n \frac{2}{(2n-1)\pi} \sin\left(\frac{1}{2}(2n-1)\pi t\right). \quad (1.36)$$

1.6 Discussion and Bibliography

The material presented in this chapter is very standard and can be found in any textbook on stochastic processes. Consult, for example, [75, 112, 113, 119]. The proof of Bochner's theorem, Theorem 1.1, can be found in [125], where additional material on stationary processes can be found. See also [112].

The Ornstein–Uhlenbeck process was introduced by Ornstein and Uhlenbeck in 1930 as a model for the velocity of a Brownian particle [234]. An early reference on the derivation of formulas of the form (1.15) is [231].

Gaussian processes are studied in [1]. Simulation algorithms for Gaussian processes are presented in [8]. Fractional Brownian motion was introduced in [152].

The spectral theorem for compact self-adjoint operators that we used in the proof of the Karhunen–Loëve expansion can be found in [204]. The Karhunen–Loëve expansion can be used to generate random fields, i.e., a collection of random variables that are parameterized by a spatial (rather than temporal) parameter x . See [69]. The Karhunen–Loëve expansion is useful in the development of numerical algorithms for partial differential equations with random coefficients. See [217].

We can use the Karhunen–Loëve expansion to study the L^2 -regularity of stochastic processes. First, let R be a compact, symmetric positive definite operator on $L^2(0, 1)$ with eigenvalues and normalized eigenfunctions $\{\lambda_k, e_k(x)\}_{k=1}^{+\infty}$, and consider a function $f \in L^2(0, 1)$ with $\int_0^1 f(s) ds = 0$. We can define the one-parameter family of Hilbert spaces H^α through the norm

$$\|f\|_\alpha^2 = \|R^{-\frac{\alpha}{2}} f\|_{L^2}^2 = \sum_k |f_k|^2 \lambda_k^{-\alpha}.$$

The inner product can be obtained through polarization. This norm enables us to measure the regularity of the function $f(t)$.⁴ Let X_t be a mean-zero second-order (i.e., with finite second moment) process with continuous autocorrelation function. Define the space $\mathcal{H}^\alpha := L^2((\Omega, P), H^\alpha(0, 1))$ with (semi)norm

$$\|X_t\|_\alpha^2 = \mathbb{E}\|X_t\|_{H^\alpha}^2 = \sum_k |\lambda_k|^{1-\alpha}. \quad (1.37)$$

Note that the regularity of the stochastic process X_t depends on the decay of the eigenvalues of the integral operator $\mathcal{R} := \int_0^1 R(t, s) \cdot ds$.

As an example, consider the L^2 -regularity of Brownian motion. From Example 1.7, we know that $\lambda_k \sim k^{-2}$. Consequently, from (1.37), we get that in order for W_t to be an element of the space \mathcal{H}^α , we need that

$$\sum_k |k|^{-2(1-\alpha)} < +\infty,$$

⁴ Think of R as being the inverse of the Laplacian with periodic boundary conditions. In this case, H^α coincides with the standard fractional Sobolev space.

from which we obtain that $\alpha < 1/2$. This is consistent with the Hölder continuity of Brownian motion from Theorem 1.3.⁵

1.7 Exercises

1. Let Y_0, Y_1, \dots be a sequence of independent identically distributed random variables and consider the stochastic process $X_n = Y_n$.

- (a) Show that X_n is a strictly stationary process.
- (b) Assume that $\mathbb{E}Y_0 = \mu < +\infty$ and $\mathbb{E}Y_0^2 = \sigma^2 < +\infty$. Show that

$$\lim_{N \rightarrow +\infty} \mathbb{E} \left| \frac{1}{N} \sum_{j=0}^{N-1} X_j - \mu \right| = 0.$$

- (c) Let f be such that $\mathbb{E}f^2(Y_0) < +\infty$. Show that

$$\lim_{N \rightarrow +\infty} \mathbb{E} \left| \frac{1}{N} \sum_{j=0}^{N-1} f(X_j) - f(Y_0) \right| = 0.$$

2. Let Z be a random variable and define the stochastic process $X_n = Z$, $n = 0, 1, 2, \dots$. Show that X_n is a strictly stationary process.

3. Let A_0, A_1, \dots, A_m and B_0, B_1, \dots, B_m be uncorrelated random variables with mean zero and variances $\mathbb{E}A_i^2 = \sigma_i^2$, $\mathbb{E}B_i^2 = \sigma_i^2$, $i = 1, \dots, m$. Let $\omega_0, \omega_1, \dots, \omega_m \in [0, \pi]$ be distinct frequencies and define, for $n = 0, \pm 1, \pm 2, \dots$, the stochastic process

$$X_n = \sum_{k=0}^m \left(A_k \cos(n\omega_k) + B_k \sin(n\omega_k) \right).$$

Calculate the mean and the covariance of X_n . Show that it is a weakly stationary process.

4. Let $\{\xi_n : n = 0, \pm 1, \pm 2, \dots\}$ be uncorrelated random variables with $\mathbb{E}\xi_n = \mu$, $\mathbb{E}(\xi_n - \mu)^2 = \sigma^2$, $n = 0, \pm 1, \pm 2, \dots$. Let a_1, a_2, \dots be arbitrary real numbers and consider the stochastic process

$$X_n = a_1 \xi_n + a_2 \xi_{n-1} + \dots + a_m \xi_{n-m+1}.$$

- (a) Calculate the mean, variance, and the covariance function of X_n . Show that it is a weakly stationary process.
- (b) Set $a_k = 1/\sqrt{m}$ for $k = 1, \dots, m$. Calculate the covariance function and study the cases $m = 1$ and $m \rightarrow +\infty$.

⁵ Observe, however, that Wiener's theorem refers to a.s. Hölder continuity, whereas the calculation presented in this section is about L^2 -continuity.

5. Let $W(t)$ be a standard one-dimensional Brownian motion. Calculate the following expectations:
- $\mathbb{E}e^{iW(t)}$.
 - $\mathbb{E}e^{i(W(t)+W(s))}$, $t, s \in (0, +\infty)$.
 - $\mathbb{E}(\sum_{i=1}^n c_i W(t_i))^2$, where $c_i \in \mathbb{R}$, $i = 1, \dots, n$ and $t_i \in (0, +\infty)$, $i = 1, \dots, n$.
 - $\mathbb{E}e^{i(\sum_{i=1}^n c_i W(t_i))}$, where $c_i \in \mathbb{R}$, $i = 1, \dots, n$ and $t_i \in (0, +\infty)$, $i = 1, \dots, n$.
6. Let W_t be a standard one-dimensional Brownian motion and define
- $$B_t = W_t - tW_1, \quad t \in [0, 1].$$
- (a) Show that B_t is a Gaussian process with
- $$\mathbb{E}B_t = 0, \quad \mathbb{E}(B_t B_s) = \min(t, s) - ts.$$
- (b) Show that for $t \in [0, 1]$, an equivalent definition of B_t is obtained through the formula
- $$B_t = (1-t)W\left(\frac{t}{1-t}\right).$$
- (c) Calculate the distribution function of B_t .
7. Let X_t be a mean-zero second-order stationary process with autocorrelation function
- $$R(t) = \sum_{j=1}^N \frac{\lambda_j^2}{\alpha_j} e^{-\alpha_j|t|},$$
- where $\{\alpha_j, \lambda_j\}_{j=1}^N$ are positive real numbers.
- (a) Calculate the spectral density and the correlation time of this process.
- (b) Show that the assumptions of Theorem 1.3 are satisfied and use the argument presented in Sect. 1.2 (i.e., the Green–Kubo formula) to calculate the diffusion coefficient of the process $Z_t = \int_0^t X_s ds$.
- (c) Under what assumptions on the coefficients $\{\alpha_j, \lambda_j\}_{j=1}^N$ can you study the above questions in the limit $N \rightarrow +\infty$?
8. Show that the position of a Brownian particle whose velocity is described by the stationary Ornstein–Uhlenbeck process Eq. (1.10) is a mean-zero Gaussian stochastic process and calculate the covariance function.
9. Let $W(t)$ be the standard one-dimensional Brownian motion and let $\sigma, s_1, s_2 > 0$. Calculate
- $\mathbb{E}e^{\sigma W(t)}$.
 - $\mathbb{E}(\sin(\sigma W(s_1)) \sin(\sigma W(s_2)))$.
10. Let W_t be a one-dimensional Brownian motion and let $\mu, \sigma > 0$ and define

$$S_t = e^{t\mu + \sigma W_t}.$$

- (a) Calculate the mean and the variance of S_t .
 (b) Calculate the probability density function of S_t .
11. Prove Proposition 1.5.
 12. Use Lemma 1.6 to calculate the distribution function of the stationary Ornstein–Uhlenbeck process.
 13. Calculate the mean and the correlation function of the integral of a standard Brownian motion

$$Y_t = \int_0^t W_s ds.$$

14. Show that the process

$$Y_t = \int_t^{t+1} (W_s - W_t) ds, \quad t \in \mathbb{R},$$

is second-order stationary.

15. Let $V_t = e^{-t} W(e^{2t})$ be the stationary Ornstein–Uhlenbeck process. Give the definition and study the main properties of the Ornstein–Uhlenbeck bridge.
 16. The autocorrelation function of the velocity $Y(t)$ of a Brownian particle moving in a harmonic potential $V(x) = \frac{1}{2} \omega_0^2 x^2$ is

$$R(t) = e^{-\gamma|t|} \left(\cos(\delta|t|) - \frac{1}{\delta} \sin(\delta|t|) \right),$$

where γ is the friction coefficient and $\delta = \sqrt{\omega_0^2 - \gamma^2}$.

- (a) Calculate the spectral density of $Y(t)$.
 (b) Calculate the mean square displacement $\mathbb{E}(X(t))^2$ of the position of the Brownian particle $X(t) = \int_0^t Y(s) ds$. Study the limit $t \rightarrow +\infty$.
 17. Prove the scaling property (1.27) of the fractional Brownian motion.
 18. The Poisson process with intensity λ , denoted by $N(t)$, is an integer-valued continuous-time stochastic process with independent increments satisfying

$$\mathbb{P}[(N(t) - N(s)) = k] = \frac{e^{-\lambda(t-s)} (\lambda(t-s))^k}{k!}, \quad t > s \geq 0, k \in \mathbb{N}.$$

Use Theorem 1.2 to show that there does not exist a continuous modification of this process.

19. Show that the correlation function of a process X_t satisfying (1.28) is continuous in both t and s .
 20. Let X_t be a stochastic process satisfying (1.28), and $R(t, s)$ its correlation function. Show that the integral operator $\mathcal{R} : L^2[0, 1] \mapsto L^2[0, 1]$ defined in (1.32),

$$\mathcal{R}f := \int_0^1 R(t, s) f(s) ds,$$

is self-adjoint and nonnegative. Show that all of its eigenvalues are real and nonnegative. Show that eigenfunctions corresponding to different eigenvalues are orthogonal.

21. Let H be a Hilbert space. An operator $\mathcal{R} : H \rightarrow H$ is said to be Hilbert–Schmidt if there exists a complete orthonormal sequence $\{\phi_n\}_{n=1}^{\infty}$ in H such that

$$\sum_{n=1}^{\infty} \|\mathcal{R}e_n\|^2 < \infty.$$

Let $\mathcal{R} : L^2[0, 1] \mapsto L^2[0, 1]$ be the operator defined in (1.32) with $R(t, s)$ being continuous both in t and s . Show that it is a Hilbert–Schmidt operator.

22. Let X_t a mean-zero second-order stationary process defined in the interval $[0, T]$ with continuous covariance $R(t)$, and let $\{\lambda_n\}_{n=1}^{+\infty}$ be the eigenvalues of the covariance operator. Show that

$$\sum_{n=1}^{\infty} \lambda_n = T R(0).$$

23. Calculate the Karhunen–Loève expansion for a second-order stochastic process with correlation function $R(t, s) = ts$.
 24. Calculate the Karhunen–Loève expansion of the Brownian bridge on $[0, 1]$.
 25. Let X_t , $t \in [0, T]$, be a second-order process with continuous covariance and Karhunen–Loève expansion

$$X_t = \sum_{k=1}^{\infty} \xi_k e_k(t).$$

Define the process

$$Y(t) = f(t)X_{\tau(t)}, \quad t \in [0, S],$$

where $f(t)$ is a continuous function and $\tau(t)$ a continuous nondecreasing function with $\tau(0) = 0$, $\tau(S) = T$. Find the Karhunen–Loève expansion of $Y(t)$, in an appropriate weighted L^2 space, in terms of the KL expansion of X_t . Use this to calculate the KL expansion of the Ornstein–Uhlenbeck process.

26. Calculate the Karhunen–Loève expansion of a centered Gaussian stochastic process with covariance function $R(s, t) = \cos(2\pi(t - s))$.
 27. Use the Karhunen–Loève expansion to generate paths of
 (a) the Brownian motion on $[0, 1]$;
 (b) the Brownian bridge on $[0, 1]$;
 (c) the Ornstein–Uhlenbeck process on $[0, 1]$.

Study computationally the convergence of the Karhunen–Loève expansion for these processes. How many terms do you need to keep in the expansion in order to calculate accurate statistics of these processes? How does the computational cost compare with that of the standard algorithm for simulating Gaussian stochastic processes?

28. (See [69].) Consider the mean zero Gaussian random field $X(x)$ in \mathbb{R} with covariance function

$$\gamma(x, y) = e^{-a|x-y|}, \quad (1.38)$$

where $a > 0$.

- (a) Simulate this field: generate samples and calculate the first four moments.
- (b) Consider $X(x)$ for $x \in [-L, L]$. Calculate analytically the eigenvalues and eigenfunctions of the integral operator \mathcal{K} with kernel $\gamma(x, y)$,

$$\mathcal{K}f(x) = \int_{-L}^L \gamma(x, y) f(y) dy.$$

Use this to obtain the Karhunen–Loëve expansion for X . Plot the first five eigenfunctions when $a = 1$, $L = 0.5$. Investigate (either analytically or by means of numerical experiments) the accuracy of the KL expansion as a function of the number of modes kept.

- (c) Develop a numerical method for calculating the first few eigenvalues and eigenfunctions of \mathcal{K} with $a = 1$, $L = 0.5$. Use the numerically calculated eigenvalues and eigenfunctions to simulate $X(x)$ using the KL expansion. Compare with the analytical results and comment on the accuracy of the calculation of the eigenvalues and eigenfunctions and on the computational cost.

Chapter 2

Diffusion Processes

In this chapter, we study some of the basic properties of Markov stochastic processes, and in particular, the properties of diffusion processes. In Sect. 2.1, we present various examples of Markov processes in discrete and continuous time. In Sect. 2.2, we give the precise definition of a Markov process and we derive the fundamental equation in the theory of Markov processes, the Chapman–Kolmogorov equation. In Sect. 2.3, we introduce the concept of the generator of a Markov process. In Sect. 2.4, we study ergodic Markov processes. In Sect. 2.5, we introduce diffusion processes, and we derive the forward and backward Kolmogorov equations. Discussion and bibliographical remarks are presented in Sect. 2.6, and exercises can be found in Sect. 2.7.

2.1 Examples of Markov Processes

Roughly speaking, a Markov process is a stochastic process that retains no memory of where it has been in the past: only the current state of a Markov process can influence where it will go next. A bit more precisely, a Markov process is a stochastic process whose past and future are statistically independent, conditioned on its present state.

Perhaps the simplest example of a Markov process is a random walk in one dimension. Let ξ_i , $i = 1, \dots$, be independent identically distributed mean-0 and variance-1 random variables. The one-dimensional random walk is defined as

$$X_N = \sum_{n=1}^N \xi_n, \quad X_0 = 0.$$

Let i_1, i_2, \dots be a sequence of nonnegative integers. Then for all integers n and m , we have

$$\mathbb{P}(X_{n+m} = i_{n+m} | X_1 = i_1, \dots, X_n = i_n) = \mathbb{P}(X_{n+m} = i_{n+m} | X_n = i_n). \quad (2.1)$$

(In fact, it is sufficient to take $m = 1$ in (2.1). See Exercise 1.) In words, the probability that the random walk will be at i_{n+m} at time $n+m$ depends only on its current value (at time n) and not on how it got there.

The random walk is an example of a *discrete-time Markov chain*. We will say that a stochastic process $\{S_n; n \in \mathbb{N}\}$ with state space $S = \mathbb{Z}$ is a discrete-time Markov chain if the Markov property (2.1) is satisfied.

Consider now a continuous-time stochastic process X_t with state space $S = \mathbb{Z}$ and denote by $\{X_s, s \leq t\}$ the collection of values of the stochastic process up to time t . We will say that X_t is a Markov processes if

$$\mathbb{P}(X_{t+h} = i_{t+h} | \{X_s, s \leq t\}) = \mathbb{P}(X_{t+h} = i_{t+h} | X_t), \quad (2.2)$$

for all $h \geq 0$. A continuous-time discrete-state-space Markov process is called a continuous-time Markov chain. A standard example of a continuous-time Markov chain is a Poisson process of rate λ with

$$\mathbb{P}(N_{t+h} = j | N_t = i) = \begin{cases} 0 & \text{if } j < i, \\ \frac{e^{-\lambda h} (\lambda h)^{j-i}}{(j-i)!}, & \text{if } j \geq i. \end{cases} \quad (2.3)$$

Similarly, we can define a continuous-time Markov process with state space \mathbb{R} as a stochastic process whose future depends on its present state and not on how it got there:

$$\mathbb{P}(X_{t+h} \in \Gamma | \{X_s, s \leq t\}) = \mathbb{P}(X_{t+h} \in \Gamma | X_t) \quad (2.4)$$

for all Borel sets Γ . In this book, we will consider continuous-time Markov processes for which a *conditional probability density* exists:

$$\mathbb{P}(X_{t+h} \in \Gamma | X_t = x) = \int_{\Gamma} p(y, t+h | x, t) dy. \quad (2.5)$$

Example 2.1. Brownian motion is a Markov process with conditional probability density given by the following formula:

$$\mathbb{P}(W_{t+h} \in \Gamma | W_t = x) = \int_{\Gamma} \frac{1}{\sqrt{2\pi h}} \exp\left(-\frac{|y-x|^2}{2h}\right) dy. \quad (2.6)$$

The Markov property of Brownian motion follows from the fact that it has independent increments.

Example 2.2. The stationary Ornstein–Uhlenbeck process $V_t = e^{-t}W(e^{2t})$ is a Markov process with conditional probability density

$$p(y, t | x, s) = \frac{1}{\sqrt{2\pi(1-e^{-2(t-s)})}} \exp\left(-\frac{|y-xe^{-(t-s)}|^2}{2(1-e^{-2(t-s)})}\right). \quad (2.7)$$

To prove (2.7), we use the formula for the distribution function of the Brownian motion to calculate, for $t > s$,

$$\begin{aligned}
\mathbb{P}(V_t \leq y | V_s = x) &= \mathbb{P}(e^{-t}W(e^{2t}) \leq y | e^{-s}W(e^{2s}) = x) \\
&= \mathbb{P}(W(e^{2t}) \leq e^t y | W(e^{2s}) = e^s x) \\
&= \int_{-\infty}^{e^t y} \frac{1}{\sqrt{2\pi(e^{2t} - e^{2s})}} e^{-\frac{|z - xe^s|^2}{2(e^{2t} - e^{2s})}} dz \\
&= \int_{-\infty}^y \frac{e^t}{\sqrt{2\pi(e^{2t} - e^{2s})}} e^{-\frac{|\rho e^t - xe^s|^2}{2(e^{2t} - e^{2s})}} d\rho \\
&= \int_{-\infty}^y \frac{1}{\sqrt{2\pi(1 - e^{-2(t-s)})}} e^{-\frac{|\rho - xe^{-(t-s)}|^2}{2(1 - e^{-2(t-s)})}} d\rho.
\end{aligned}$$

Consequently, the transition probability density for the Ornstein–Uhlenbeck process is given by the formula

$$\begin{aligned}
p(y, t | x, s) &= \frac{\partial}{\partial y} \mathbb{P}(V_t \leq y | V_s = x) \\
&= \frac{1}{\sqrt{2\pi(1 - e^{-2(t-s)})}} \exp\left(-\frac{|y - xe^{-(t-s)}|^2}{2(1 - e^{-2(t-s)})}\right).
\end{aligned}$$

The Markov property enables us to obtain an evolution equation for the *transition probability* for a discrete-time or continuous-time Markov chain

$$\mathbb{P}(X_{n+1} = i_{n+1} | X_n = i_n), \quad \mathbb{P}(X_{t+h} = i_{t+h} | X_t = i_t), \quad (2.8)$$

or for the transition probability density defined in (2.5). This equation is the *Chapman–Kolmogorov equation*. Using this equation, we can study the evolution of a Markov process.

We will be concerned mostly with time-homogeneous Markov processes, i.e., processes for which the conditional probabilities are invariant under time shifts. For time-homogeneous discrete-time Markov chains, we have

$$\mathbb{P}(X_{n+1} = j | X_n = i) = \mathbb{P}(X_1 = j | X_0 = i) =: p_{ij}.$$

We will refer to the matrix $P = \{p_{ij}\}$ as the transition matrix. The transition matrix is a stochastic matrix, i.e., it has nonnegative entries and $\sum_j p_{ij} = 1$. Similarly, we can define the n -step transition matrix $P_n = \{p_{ij}(n)\}$ as

$$p_{ij}(n) = \mathbb{P}(X_{m+n} = j | X_m = i).$$

We can study the evolution of a Markov chain through the Chapman–Kolmogorov equation:

$$p_{ij}(m+n) = \sum_k p_{ik}(m) p_{kj}(n). \quad (2.9)$$

Indeed, let $\mu_i^{(n)} := \mathbb{P}(X_n = i)$. The (possibly infinite-dimensional) vector $\mu^{(n)}$ determines the state of the Markov chain at time n . From the Chapman–Kolmogorov equation, we can obtain a formula for the evolution of the vector $\mu^{(n)}$:

$$\mu^{(n)} = \mu^{(0)} P^n, \quad (2.10)$$

where P^n denotes the n th power of the matrix P . Hence in order to calculate the state of the Markov chain at time n , what we need is the initial distribution μ^0 and the transition matrix P . Componentwise, the above equation can be written as

$$\mu_j^{(n)} = \sum_i \mu_i^{(0)} \pi_{ij}(n).$$

Consider now a continuous-time Markov chain with transition probability

$$p_{ij}(s, t) = \mathbb{P}(X_t = j | X_s = i), \quad s \leq t.$$

If the chain is homogeneous, then

$$p_{ij}(s, t) = p_{ij}(0, t - s) \quad \text{for all } i, j, s \leq t.$$

In particular,

$$p_{ij}(t) = \mathbb{P}(X_t = j | X_0 = i).$$

The Chapman–Kolmogorov equation for a continuous-time Markov chain is

$$\frac{dp_{ij}}{dt} = \sum_k p_{ik}(t) g_{kj}, \quad (2.11)$$

where the matrix G is called the *generator* of the Markov chain, defined as

$$G = \lim_{h \rightarrow 0} \frac{1}{h} (P_h - I),$$

with P_t denoting the matrix $\{p_{ij}(t)\}$. Equation (2.11) can also be written in matrix form:

$$\frac{dP}{dt} = P_t G.$$

Let now $\mu_t^i = \mathbb{P}(X_t = i)$. The vector μ_t is the distribution of the Markov chain at time t . We can study its evolution using the equation

$$\mu_t = \mu_0 P_t.$$

Thus, as in the case of discrete-time Markov chains, the evolution of a continuous-time Markov chain is completely determined by the initial distribution and transition matrix.

Consider now the case a continuous-time Markov process with continuous state space and with continuous paths. As we have seen in Example 2.1, Brownian motion

is such a process. The conditional probability density of the Brownian motion (2.6) is the fundamental solution (Green's function) of the diffusion equation:

$$\frac{\partial p}{\partial t} = \frac{1}{2} \frac{\partial^2 p}{\partial y^2}, \quad \lim_{t \rightarrow s} p(y, t | x, s) = \delta(y - x). \quad (2.12)$$

Similarly, the conditional distribution of the Ornstein–Uhlenbeck process satisfies the initial value problem

$$\frac{\partial p}{\partial t} = \frac{\partial(yp)}{\partial y} + \frac{1}{2} \frac{\partial^2 p}{\partial y^2}, \quad \lim_{t \rightarrow s} p(y, t | x, s) = \delta(y - x). \quad (2.13)$$

Brownian motion and the Ornstein–Uhlenbeck process are examples of a *diffusion process*: a continuous-time Markov process with continuous paths. A precise definition will be given in Sect. 2.5, where we will also derive evolution equations for the conditional probability density $p(y, t | x, s)$ of an arbitrary diffusion process, the forward Kolmogorov (Fokker–Planck) equation (2.54), and the backward Kolmogorov (2.47) equation.

2.2 Markov Processes and the Chapman–Kolmogorov Equation

In Sect. 2.1, we gave the definition of a Markov process whose time is either discrete or continuous and whose state space is countable. We also gave several examples of Markov chains as well as of processes with state space the real line. In this section, we give the precise definition of a Markov process with $t \in \mathbb{R}_+$ and with state space \mathbb{R}^d . We also introduce the Chapman–Kolmogorov equation.

In order to give the definition of a Markov process, we need to use the conditional expectation of the stochastic process conditioned on all past values. We can encode all past information about a stochastic process into an appropriate collection of σ -algebras. Let $(\Omega, \mathcal{F}, \mu)$ denote a probability space, and consider a stochastic process $X = X_t(\omega)$ with $t \in \mathbb{R}_+$ and state space $(\mathbb{R}^d, \mathcal{B})$, where \mathcal{B} denotes the Borel σ -algebra. We define the σ -algebra generated by $\{X_t, t \in \mathbb{R}_+\}$, denoted by $\sigma(X_t, t \in \mathbb{R}_+)$, to be the smallest σ -algebra such that the family of mappings $\{X_t, t \in \mathbb{R}_+\}$ is a stochastic process with sample space $(\Omega, \sigma(X_t, t \in \mathbb{R}_+))$ and state space $(\mathbb{R}^d, \mathcal{B})$.¹ In other words, the σ -algebra generated by X_t is the smallest σ -algebra such that X_t is a measurable function (random variable) with respect to it.

We define now a filtration on (Ω, \mathcal{F}) to be a nondecreasing family $\{\mathcal{F}_t, t \in \mathbb{R}_+\}$ of sub- σ -algebras of \mathcal{F} :

$$\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F} \quad \text{for } s \leq t.$$

¹ In later chapters, we will also consider Markov processes with state space a subset of \mathbb{R}^d , for example the unit torus.

We set $\mathcal{F}_\infty = \sigma(\cup_{t \in T} \mathcal{F}_t)$. The filtration generated by our stochastic process X_t , where X_t is:

$$\mathcal{F}_t^X := \sigma(X_s; s \leq t). \quad (2.14)$$

Now we are ready to give the definition of a Markov process.

Definition 2.1. Let X_t be a stochastic process defined on a probability space $(\Omega, \mathcal{F}, \mu)$ with values in \mathbb{R}^d , and let \mathcal{F}_t^X be the filtration generated by $\{X_t; t \in \mathbb{R}_+\}$. Then $\{X_t; t \in \mathbb{R}_+\}$ is a Markov process if

$$\mathbb{P}(X_t \in \Gamma | \mathcal{F}_s^X) = \mathbb{P}(X_t \in \Gamma | X_s) \quad (2.15)$$

for all $t, s \in T$ with $t \geq s$, and $\Gamma \in \mathcal{B}(\mathbb{R}^d)$.

We remark that the filtration \mathcal{F}_t^X is generated by events of the form $\{\omega | X_{t_1} \in \Gamma_1, X_{t_2} \in \Gamma_2, \dots, X_{t_n} \in \Gamma_n\}$ with $0 \leq t_1 < t_2 < \dots < t_n \leq t$ and $\Gamma_i \in \mathcal{B}(\mathbb{R}^d)$. The definition of a Markov process is thus equivalent to the hierarchy of equations

$$\mathbb{P}(X_t \in \Gamma | X_{t_1}, X_{t_2}, \dots, X_{t_n}) = \mathbb{P}(X_t \in \Gamma | X_{t_n}) \quad \text{a.s.}$$

for $n \geq 1$ and $0 \leq t_1 < t_2 < \dots < t_n \leq t$ with $\Gamma \in \mathcal{B}(E)$.

We also remark that it is sometimes possible to describe a non-Markovian process X_t in terms of a Markov process Y_t in a higher-dimensional state space. The additional variables that we introduce account for the memory in the X_t . This is possible when the non-Markovian process has finite memory that can be represented by a finite number of additional degrees of freedom. We will use this approach in Chap. 8 when we derive stochastic differential equations from deterministic dynamical systems with random initial conditions; see Definition 8.1.

As an example, consider a Brownian particle whose velocity is described by the stationary Ornstein–Uhlenbeck process $Y_t = e^{-t}W(e^{2t})$; see (1.10) and (1.11). The particle position is given by the integral of the Ornstein–Uhlenbeck process

$$X_t = X_0 + \int_0^t Y_s ds.$$

The particle position depends on the past of the Ornstein–Uhlenbeck process and consequently, is not a Markov process. However, the joint position–velocity process $\{X_t, Y_t\}$ is a Markov process. Its transition probability density $p(x, y, t | x_0, y_0)$ satisfies the forward Kolmogorov equation

$$\frac{\partial p}{\partial t} = -p \frac{\partial p}{\partial x} + \frac{\partial}{\partial y}(yp) + \frac{1}{2} \frac{\partial^2 p}{\partial y^2}.$$

The Chapman–Kolmogorov Equation

With every continuous-time Markov process X_t (we always take $t \in \mathbb{R}_+$) defined in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and state space $(\mathbb{R}^d, \mathcal{B})$, we can associate the *transition function*

$$P(\Gamma, t | X_s, s) := \mathbb{P}[X_t \in \Gamma | \mathcal{F}_s^X],$$

for all $t, s \in \mathbb{R}_+$ with $t \geq s$ and all $\Gamma \in \mathcal{B}(\mathbb{R}^d)$. It is a function of four arguments: the initial time s and position X_s , and the final time t and the set Γ . The transition function $P(\Gamma, t | x, s)$ is, for fixed t, x, s , a probability measure on \mathbb{R}^d with $P(\mathbb{R}^d, t | x, s) = 1$; it is $\mathcal{B}(\mathbb{R}^d)$ -measurable in x , for fixed t, s, Γ , and satisfies the Chapman–Kolmogorov equation

$$P(\Gamma, t | x, s) = \int_{\mathbb{R}^d} P(\Gamma, t | y, u) P(dy, u | x, s) \quad (2.16)$$

for all $x \in \mathbb{R}^d$, $\Gamma \in \mathcal{B}(\mathbb{R}^d)$, and $s, u, t \in \mathbb{R}_+$ with $s \leq u \leq t$. Assume that $X_s = x$. Since $\mathbb{P}[X_t \in \Gamma | \mathcal{F}_s^X] = \mathbb{P}[X_t \in \Gamma | X_s]$, we can write

$$P(\Gamma, t | x, s) = \mathbb{P}[X_t \in \Gamma | X_s = x].$$

The derivation of the Chapman–Kolmogorov equation is based on the Markovian assumption and on properties of conditional probability. We can formally derive the Chapman–Kolmogorov equation as follows: We use the Markov property, together with Eqs. (B.9) and (B.10) from Appendix B and the fact that $s < u \Rightarrow \mathcal{F}_s^X \subset \mathcal{F}_u^X$ to calculate

$$\begin{aligned} P(\Gamma, t | x, s) &= \mathbb{P}(X_t \in \Gamma | X_s = x) = \mathbb{P}(X_t \in \Gamma | \mathcal{F}_s^X) \\ &= \mathbb{E}(I_\Gamma(X_t) | \mathcal{F}_s^X) = \mathbb{E}(\mathbb{E}(I_\Gamma(X_t) | \mathcal{F}_s^X) | \mathcal{F}_u^X) \\ &= \mathbb{E}(\mathbb{E}(I_\Gamma(X_t) | \mathcal{F}_u^X) | \mathcal{F}_s^X) = \mathbb{E}(\mathbb{P}(X_t \in \Gamma | X_u) | \mathcal{F}_s^X) \\ &= \mathbb{E}(\mathbb{P}(X_t \in \Gamma | X_u = y) | X_s = x) \\ &= \int_{\mathbb{R}^d} P(\Gamma, t | X_u = y) P(dy, u | X_s = x) \\ &=: \int_{\mathbb{R}^d} P(\Gamma, t | y, u) P(dy, u | x, s), \end{aligned}$$

where $I_\Gamma(\cdot)$ denotes the indicator function of the set Γ . In words, the Chapman–Kolmogorov equation tells us that for a Markov process, the transition from x at time s to the set Γ at time t can be done in two steps: first, the system moves from x to y at some intermediate time u . Then it moves from y to Γ at time t . In order to calculate the probability for the transition from x at time s to Γ at time t , we need to sum (integrate) the transitions from all possible intermediate states y .

The transition function and the initial distribution of X_t are sufficient to determine a Markov process uniquely. In fact, a process X_t is a Markov process with respect to its filtration \mathcal{F}_t^X defined in (2.14) with transition function $P(\cdot, t | \cdot, s)$ and initial

distribution v ($X_0 \sim v$) if and only if for all $0 = t_0 < t_1 < \dots < t_n$ and bounded measurable functions f_j , $j = 0, \dots, N$, we have

$$\mathbb{E}_v \prod_{j=0}^n f_j(X_{t_j}) = \int_{\mathbb{R}^d} f_0(x_0) v(dx_0) \prod_{j=1}^n \int_{\mathbb{R}^d} f_j(x_j) P(dx_j, t_j | x_{j-1}, t_{j-1}), \quad (2.17)$$

where we have used the notation \mathbb{E}_v to emphasize the dependence of the expectation on the initial distribution v . The proof that a Markov process with transition function P satisfies (2.17) follows from the Chapman–Kolmogorov equation (2.16) and an induction argument. In other words, the finite-dimensional distributions of X_t are uniquely determined by the initial distribution and the transition function:

$$\mathbb{P}(X_0 \in dx_0, X_{t_1} \in dx_1, \dots, X_{t_n} \in dx_n) = v(dx_0) \prod_{j=1}^n P(dx_j, t_j | x_{j-1}, t_{j-1}). \quad (2.18)$$

In this book, we will consider Markov processes whose transition function has a density with respect to the Lebesgue measure:

$$P(\Gamma, t | x, s) = \int_{\Gamma} p(y, t | x, s) dy.$$

We will refer to $p(y, t | x, s)$ as the *transition probability density*. It is a function of four arguments: the initial position and time x, s , and the final position and time y, t . For $t = s$, we have $P(\Gamma, s | x, s) = I_{\Gamma}(x)$. The Chapman–Kolmogorov equation becomes

$$\int_{\Gamma} p(y, t | x, s) dy = \int_{\mathbb{R}^d} \int_{\Gamma} p(y, t | z, u) p(z, u | x, s) dz dy,$$

and since $\Gamma \in \mathcal{B}(\mathbb{R}^d)$ is arbitrary, we obtain the Chapman–Kolmogorov equation for the transition probability density:

$$p(y, t | x, s) = \int_{\mathbb{R}^d} p(y, t | z, u) p(z, u | x, s) dz. \quad (2.19)$$

When the transition probability density exists, and assuming that the initial distribution v has density ρ , we can write $\mathbb{P}(X_0 \in dx_0, X_{t_1} \in dx_1, \dots, X_{t_n} \in dx_n) = p(x_0, t_0, \dots, x_n, t_n) \prod_{j=0}^n dx_j$, and we have

$$p(x_0, t_0, \dots, x_n, t_n) = \rho(x_0) \prod_{j=1}^n p(x_j, t_j | x_{j-1}, t_{j-1}). \quad (2.20)$$

The above formulas simplify when the (random) law of evolution of the Markov process X_t does not change in time. In this case, the conditional probability in (2.15) depends on the initial and final times t and s only through their difference: we will say that a Markov process is *time-homogeneous* if the transition function $P(\cdot, t | \cdot, s)$ depends only on the difference $t - s$ between the initial and final times:

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

for all $\Gamma \in \mathcal{B}(\mathbb{R}^d)$ and $x \in \mathbb{R}^d$. For time-homogeneous Markov processes we can fix the initial time, $s = 0$. The Chapman–Kolmogorov equation for a time-homogeneous Markov process then becomes

$$P(t+s, x, \Gamma) = \int_{\mathbb{R}^d} P(s, x, dz) P(t, z, \Gamma). \quad (2.21)$$

Furthermore, formulas (2.17) and (2.18) become respectively

$$\mathbb{E}_v \prod_{j=0}^n f_j(X_{t_j}) = \int_{\mathbb{R}^d} f_0(y_0) \mu(dy_0) \prod_{j=1}^n \int_{\mathbb{R}^d} f_j(y_j) P(t_j - t_{j-1}, y_{j-1}, dy_j) \quad (2.22)$$

and

$$\mathbb{P}(X_0 \in dx_0, X_{t_1} \in dx_1, \dots, X_{t_n} \in dx_n) = v(dx_0) \prod_{j=1}^n P(t_j - t_{j-1}, y_{j-1}, dy_j). \quad (2.23)$$

Given the initial distribution v and the transition function $P(x, t, \Gamma)$ of a Markov process X_t , we can calculate the probability of finding X_t in a set Γ at time t :

$$\mathbb{P}(X_t \in \Gamma) = \int_{\mathbb{R}^d} P(x, t, \Gamma) v(dx).$$

Furthermore, for an observable f , we can calculate the expectation using the formula

$$\mathbb{E}_v f(X_t) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f(x) P(t, x_0, dx) v(dx_0). \quad (2.24)$$

The Poisson process defined in (2.3) is a homogeneous Markov process. Another example of a time-homogeneous Markov process is Brownian motion. The transition function is the Gaussian

$$P(t, x, dy) = \gamma_{t, x}(y) dy, \quad \gamma_{t, x}(y) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{|x-y|^2}{2t}\right). \quad (2.25)$$

Let now X_t be a time-homogeneous Markov process and assume that the transition probability density exists, $P(t, x, \Gamma) = \int_{\Gamma} p(t, x, y) dy$. The Chapman–Kolmogorov equation $p(t, x, y)$ reads

$$p(t+s, x, y) = \int_{\mathbb{R}^d} p(s, x, z) p(t, z, y) dz. \quad (2.26)$$

2.3 The Generator of a Markov Process

Let X_t denote a time-homogeneous Markov process. The Chapman–Kolmogorov equation (2.21) suggests that a time-homogeneous Markov process can be described through a semigroup of operators, i.e., a one-parameter family of linear operators with the properties

$$P_0 = I, \quad P_{t+s} = P_t \circ P_s \quad \text{for all } t, s \geq 0.$$

Indeed, let $P(t, \cdot, \cdot)$ be the transition function of a homogeneous Markov process and let $f \in C_b(\mathbb{R}^d)$, the space of continuous bounded functions on \mathbb{R}^d , and define the operator

$$(P_t f)(x) := \mathbb{E}(f(X_t) | X_0 = x) = \int_{\mathbb{R}^d} f(y) P(t, x, dy). \quad (2.27)$$

This is a linear operator with

$$(P_0 f)(x) = \mathbb{E}(f(X_0) | X_0 = x) = f(x),$$

which means that $P_0 = I$. Furthermore,

$$\begin{aligned} (P_{t+s} f)(x) &= \int_{\mathbb{R}^d} f(y) P(t+s, x, dy) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f(y) P(s, z, dy) P(t, x, dz) \\ &= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} f(y) P(s, z, dy) \right) P(t, x, dz) = \int_{\mathbb{R}^d} (P_s f)(z) P(t, x, dz) \\ &= (P_t \circ P_s f)(x). \end{aligned}$$

Consequently,

$$P_{t+s} = P_t \circ P_s.$$

The semigroup P_t defined in (2.27) is an example of a Markov semigroup. We can study properties of a time-homogeneous Markov process X_t by studying properties of the Markov semigroup P_t .

Let now X_t be a time-homogeneous Markov process in \mathbb{R}^d , and let P_t denote the corresponding semigroup defined in (2.27). We consider this semigroup acting on continuous bounded functions and assume that $P_t f$ is also a $C_b(\mathbb{R}^d)$ function. We define by $\mathcal{D}(\mathcal{L})$ the set of all $f \in C_b(E)$ such that the strong limit

$$\mathcal{L}f := \lim_{t \rightarrow 0} \frac{P_t f - f}{t} \quad (2.28)$$

exists. The operator $\mathcal{L} : \mathcal{D}(\mathcal{L}) \rightarrow C_b(\mathbb{R}^d)$ is called the (infinitesimal) generator of the operator semigroup P_t . We will also refer to \mathcal{L} as the generator of the Markov process X_t .

The semigroup property and the definition of the generator of a Markov semigroup (2.28) imply that formally, we can write

$$P_t = e^{t\mathcal{L}}.$$

Consider the function $u(x, t) := (P_t f)(x) = \mathbb{E}(f(X_t) | X_0 = x)$. We calculate its time derivative:

$$\begin{aligned} \frac{\partial u}{\partial t} &= \frac{d}{dt}(P_t f) = \frac{d}{dt} \left(e^{t\mathcal{L}} f \right) \\ &= \mathcal{L} \left(e^{t\mathcal{L}} f \right) = \mathcal{L} P_t f = \mathcal{L} u. \end{aligned}$$

Furthermore, $u(x, 0) = P_0 f(x) = f(x)$. Consequently, $u(x, t)$ satisfies the initial value problem

$$\frac{\partial u}{\partial t} = \mathcal{L} u, \quad (2.29a)$$

$$u(x, 0) = f(x). \quad (2.29b)$$

Equation (2.29) is the *backward Kolmogorov equation*. It governs the evolution of the expectation of an observable $f \in C_b(\mathbb{R}^d)$. At this level, this is formal, since we do not have a formula for the generator \mathcal{L} of the Markov semigroup. When the Markov process is the solution of a stochastic differential equation, then the generator is a second-order elliptic differential operator, and the backward Kolmogorov equation becomes an initial value problem for a parabolic partial differential equation. See Sect. 2.5 and Chap. 4.

As an example, consider Brownian motion in one dimension. The transition function is given by (2.25), the fundamental solution of the heat equation in one dimension. The corresponding Markov semigroup is the heat semigroup $P_t = \exp\left(\frac{t}{2} \frac{d^2}{dx^2}\right)$. The generator of the one-dimensional Brownian motion is the one-dimensional Laplacian $\frac{1}{2} \frac{d^2}{dx^2}$. The backward Kolmogorov equation is the heat equation

$$\frac{\partial u}{\partial t} = \frac{1}{2} \frac{\partial^2 u}{\partial x^2}.$$

The Adjoint Semigroup

The semigroup P_t acts on bounded continuous functions. We can also define the adjoint semigroup P_t^* , which acts on probability measures:

$$P_t^* \mu(\Gamma) = \int_{\mathbb{R}^d} \mathbb{P}(X_t \in \Gamma | X_0 = x) d\mu(x) = \int_{\mathbb{R}^d} P(t, x, \Gamma) d\mu(x).$$

The image of a probability measure μ under P_t^* is again a probability measure. The operators P_t and P_t^* are (formally) adjoint in the L^2 -sense:

$$\int_{\mathbb{R}} P_t f(x) d\mu(x) = \int_{\mathbb{R}} f(x) d(P_t^* \mu)(x). \quad (2.30)$$

We can write

$$P_t^* = e^{t\mathcal{L}^*}, \quad (2.31)$$

where \mathcal{L}^* is the L^2 -adjoint of the generator of the process:

$$\int \mathcal{L} f h dx = \int f \mathcal{L}^* h dx.$$

Let X_t be a Markov process with generator X_t with $X_0 \sim \mu$, and let P_t^* denote the adjoint semigroup defined in (2.31). We define

$$\mu_t := P_t^* \mu. \quad (2.32)$$

This is the *law* of the Markov process. An argument similar to that used in the derivation of the backward Kolmogorov equation (2.29) enables us to obtain an equation for the evolution of μ_t :

$$\frac{\partial \mu_t}{\partial t} = \mathcal{L}^* \mu_t, \quad \mu_0 = \mu.$$

Assuming that the initial distribution μ and the law of the process μ_t each have a density with respect to Lebesgue measure, denoted by $\rho_0(\cdot)$ and $\rho(\cdot, t)$ respectively, this equation becomes

$$\frac{\partial \rho}{\partial t} = \mathcal{L}^* \rho, \quad \rho(y, 0) = \rho_0(y). \quad (2.33)$$

This is the *forward Kolmogorov equation*. When the Markov process starts at $X_0 = x$, deterministic, the initial for the Fokker–Planck equation becomes $\rho_0 = \delta(x - y)$. As with the backward Kolmogorov equation (2.29), this equation is still formal, since we do not have a formula for the adjoint \mathcal{L}^* of the generator of the Markov process X_t . In Sect. 2.5, we will derive the forward and backward Kolmogorov equations and a formula for the generator \mathcal{L} for diffusion processes.

2.4 Ergodic Markov Processes

In Sect. 1.2, we studied stationary stochastic processes, and we showed that such processes satisfy a form of the law of large numbers, Theorem 1.3. In this section we introduce a class of Markov processes for which the phase-space average, with respect to an appropriate probability measure, the *invariant measure* equals the long

time average. Such Markov processes are called *ergodic*. Ergodic Markov processes are characterized by the fact that an invariant measure, see Eq. (2.36) below, exists and is unique.

For the precise definition of an ergodic Markov process we require that all shift invariant sets, i.e. all sets of the form $(X_{t_1} \in \Gamma_1, X_{t_2} \in \Gamma_2, \dots, X_{t_k} \in \Gamma_k)$ that are invariant under time shifts, are trivial, i.e. they have probability 0 or 1. For our purposes it is more convenient to describe ergodic Markov processes in terms of the properties of their generators and of the corresponding Markov semigroup.

We will consider a Markov process X_t in \mathbb{R}^d with generator \mathcal{L} and Markov semigroup P_t . We will say that X_t is ergodic provided that 0 is a simple eigenvalue of \mathcal{L} or, equivalently, provided that the equation

$$\mathcal{L}g = 0 \quad (2.34)$$

has only constant solutions. Consequently, we can study the ergodic properties of a Markov process X_t by studying the null space of its generator. From (2.34), and using the definition of the generator of a Markov process (2.28), we deduce that a Markov process is ergodic if the equation

$$P_t g = g, \quad (2.35)$$

has only constant solutions for all $t \geq 0$. Using the adjoint semigroup, we can define an *invariant measure* as a probability measure that is invariant under the time evolution of X_t , i.e., a fixed point of the semigroup P_t^* :

$$P_t^* \mu = \mu. \quad (2.36)$$

This equation is the L^2 -adjoint of the equation $P_t g = g$ in (2.35). If there is a unique probability measure satisfying (2.36), then the Markov process is ergodic (with respect to the measure μ). Using this, we can obtain an equation for the invariant measure in terms of the adjoint \mathcal{L}^* of the generator, which is the generator of the semigroup P_t^* . Assume, for simplicity, that the measure μ has a density ρ with respect to Lebesgue measure. We divide (2.36) by t and pass to the limit as $t \rightarrow 0$ to obtain

$$\mathcal{L}^* \rho = 0. \quad (2.37)$$

When X_t is a diffusion process, this equation is the *stationary Fokker–Planck equation*. Equation (2.37), which is the adjoint of (2.34), can be used to calculate the invariant distribution ρ , i.e., the density of the invariant measure μ .

The invariant measure (distribution) governs the long-time dynamics of the Markov process. In particular, when $X_0 \sim \mu_0$ initially, we have that

$$\lim_{t \rightarrow +\infty} P_t^* \mu_0 = \mu. \quad (2.38)$$

Furthermore, the long-time average of an observable f converges to the equilibrium expectation with respect to the invariant measure

$$\lim_{T \rightarrow +\infty} \frac{1}{T} \int_0^T f(X_s) ds = \int f(x) \mu(dx).$$

This is the definition of an ergodic process that is quite often used in physics: the long-time average equals the phase-space average.

If X_0 is distributed according to μ , then so is X_t for all $t > 0$. The resulting stochastic process, with X_0 distributed in this way, is stationary; see Sect. 1.2.

Example 2.3. Brownian motion in \mathbb{R}^d is not an ergodic Markov process. On the other hand, if we consider it in a bounded domain with appropriate boundary conditions, then it becomes an ergodic process. Consider a one-dimensional Brownian motion on $[0, 1]$, with periodic boundary conditions. The generator of this Markov process \mathcal{L} is the differential operator $\mathcal{L} = \frac{1}{2} \frac{d^2}{dx^2}$, equipped with periodic boundary conditions on $[0, 1]$. This operator is self-adjoint. The null spaces of both \mathcal{L} and \mathcal{L}^* comprise constant functions on $[0, 1]$. Both the backward Kolmogorov and the Fokker–Planck equation reduce to the heat equation

$$\frac{\partial \rho}{\partial t} = \frac{1}{2} \frac{\partial^2 \rho}{\partial x^2} \quad (2.39)$$

with periodic boundary conditions in $[0, 1]$. We can solve the heat equation (2.39) using Fourier analysis to deduce that the solution converges to a constant at an exponential rate.

Example 2.4. The one-dimensional Ornstein–Uhlenbeck process is a Markov process with generator

$$\mathcal{L} = -\alpha x \frac{d}{dx} + D \frac{d^2}{dx^2}.$$

The null space of \mathcal{L} comprises constants in x . Hence, it is an ergodic Markov process. In order to calculate the invariant measure, we need to solve the stationary Fokker–Planck equation:

$$\mathcal{L}^* \rho = 0, \quad \rho \geq 0, \quad \int \rho(x) dx = 1. \quad (2.40)$$

We calculate the L^2 -adjoint of \mathcal{L} . Assuming that f, h decay sufficiently fast at infinity, we have

$$\begin{aligned} \int_{\mathbb{R}} \mathcal{L} f h dx &= \int_{\mathbb{R}} \left[\left(-\alpha x \frac{df}{dx} \right) h + \left(D \frac{d^2 f}{dx^2} \right) h \right] dx \\ &= \int_{\mathbb{R}} [f \partial_x (\alpha x h) + f (D \partial_x^2 h)] dx =: \int_{\mathbb{R}} f \mathcal{L}^* h dx, \end{aligned}$$

where

$$\mathcal{L}^* h := \frac{d}{dx} (axh) + D \frac{d^2 h}{dx^2}.$$

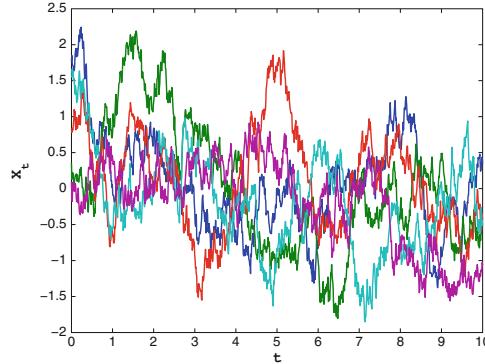


Fig. 2.1 Sample paths of the Ornstein–Uhlenbeck process

We can calculate the invariant distribution by solving Eq. (2.40). The invariant measure of this process is the Gaussian measure

$$\mu(dx) = \sqrt{\frac{\alpha}{2\pi D}} \exp\left(-\frac{\alpha}{2D}x^2\right) dx.$$

If the initial condition of the Ornstein–Uhlenbeck process is distributed according to the invariant measure, then the Ornstein–Uhlenbeck process is a stationary Gaussian process. Let X_t denote the one-dimensional Ornstein–Uhlenbeck process with $X_0 \sim \mathcal{N}(0, D/\alpha)$. Then X_t is a mean-zero Gaussian second-order stationary process on $[0, \infty)$ with correlation function

$$R(t) = \frac{D}{\alpha} e^{-\alpha|t|}$$

and spectral density

$$f(x) = \frac{D}{\pi} \frac{1}{x^2 + \alpha^2}.$$

The Ornstein–Uhlenbeck process is the only real-valued mean-zero Gaussian second-order stationary Markov process with continuous paths defined on \mathbb{R} . This is the content of Doob’s theorem. See Exercise 6. A few paths of the stationary Ornstein–Uhlenbeck process are presented in Fig. 2.1.

2.5 Diffusion Processes and the Forward and Backward Kolmogorov Equations

A Markov process consists of three parts: a drift, a random part, and a jump process. A diffusion process is a Markov process that has continuous sample paths (trajectories). Thus, it is a Markov process with no jumps. A diffusion process can be defined

by specifying its first two moments together with the requirement that there be no jumps. We begin with the definition of a diffusion process in one dimension.

Definition 2.2. A Markov process X_t in \mathbb{R} with transition function $P(\Gamma, t|x, s)$ is called a diffusion process if the following conditions are satisfied:

(i) (Continuity). For every x and every $\varepsilon > 0$,

$$\int_{|x-y|>\varepsilon} P(dy, t|x, s) = o(t-s), \quad (2.41)$$

uniformly over $s < t$.

(ii) (Definition of drift coefficient). There exists a function $b(x, s)$ such that for every x and every $\varepsilon > 0$,

$$\int_{|y-x|\leqslant\varepsilon} (y-x)P(dy, t|x, s) = b(x, s)(t-s) + o(t-s), \quad (2.42)$$

uniformly over $s < t$.

(iii) (Definition of diffusion coefficient). There exists a function $\Sigma(x, s)$ such that for every x and every $\varepsilon > 0$,

$$\int_{|y-x|\leqslant\varepsilon} (y-x)^2 P(dy, t|x, s) = \Sigma(x, s)(t-s) + o(t-s), \quad (2.43)$$

uniformly over $s < t$.

In Definition 2.2, we truncated the domain of integration, since do not know whether the first and second moments of X_t are finite. If we assume that there exists a $\delta > 0$ such that

$$\lim_{t \rightarrow s} \frac{1}{t-s} \int_{\mathbb{R}^d} |y-x|^{2+\delta} P(dy, t|x, s) = 0, \quad (2.44)$$

then we can extend the integration over all of \mathbb{R} and use expectations in the definition of the drift and the diffusion coefficient. Indeed, let $k = 0, 1, 2$ and notice that

$$\begin{aligned} \int_{|y-x|>\varepsilon} |y-x|^k P(dy, t|x, s) &= \int_{|y-x|>\varepsilon} |y-x|^{2+\delta} |y-x|^{k-(2+\delta)} P(dy, t|x, s) \\ &\leq \frac{1}{\varepsilon^{2+\delta-k}} \int_{|y-x|>\varepsilon} |y-x|^{2+\delta} P(dy, t|x, s) \\ &\leq \frac{1}{\varepsilon^{2+\delta-k}} \int_{\mathbb{R}^d} |y-x|^{2+\delta} P(dy, t|x, s). \end{aligned}$$

Using this estimate together with (2.44), we conclude that

$$\lim_{t \rightarrow s} \frac{1}{t-s} \int_{|y-x|>\varepsilon} |y-x|^k P(dy, t|x, s) = 0, \quad k = 0, 1, 2.$$

This implies that assumption (2.44) is sufficient for the sample paths to be continuous ($k = 0$) and for the replacement of the truncated integrals in (2.42) and (2.43) by integrals over \mathbb{R} ($k = 1$ and $k = 2$, respectively).

Assuming that the first two moments exist, we can write the formulas for the drift and diffusion coefficients in the following form:

$$\lim_{t \rightarrow s} \mathbb{E} \left(\frac{X_t - X_s}{t - s} \middle| X_s = x \right) = b(x, s) \quad (2.45)$$

and

$$\lim_{t \rightarrow s} \mathbb{E} \left(\frac{|X_t - X_s|^2}{t - s} \middle| X_s = x \right) = \Sigma(x, s). \quad (2.46)$$

The Backward Kolmogorov Equation

We can now use the definition of a diffusion process to obtain an explicit formula for the generator of such a process and to derive a partial differential equation for the conditional expectation $u(x, s) = \mathbb{E}(f(X_t) | X_s = x)$, as well as for the transition probability density $p(y, t | x, s)$. These are the backward and forward Kolmogorov equations. We will derive these equations for one-dimensional diffusion processes. The extension to multidimensional diffusion processes is discussed later in this section. In the following, we will assume that $u(x, s)$ is a smooth function of x and s .²

Theorem 2.1 (Kolmogorov). *Let $f(x) \in C_b(\mathbb{R})$, and let*

$$u(x, s) := \mathbb{E}(f(X_t) | X_s = x) = \int f(y) P(dy, t | x, s),$$

with t fixed. Assume, furthermore, that the functions $b(x, s)$, $\Sigma(x, s)$ are smooth in both x and s . Then $u(x, s)$ solves the final value problem

$$-\frac{\partial u}{\partial s} = b(x, s) \frac{\partial u}{\partial x} + \frac{1}{2} \Sigma(x, s) \frac{\partial^2 u}{\partial x^2}, \quad u(t, x) = f(x), \quad (2.47)$$

for $s \in [0, t]$.

Proof. First we observe that the continuity assumption (2.41), together with the fact that the function $f(x)$ is bounded, implies that

² In fact, all we need is that $u \in C^{2,1}(\mathbb{R} \times \mathbb{R}_+)$. This can be proved using our assumptions on the transition function, on f , and on the drift and diffusion coefficients.

$$\begin{aligned}
u(x, s) &= \int_{\mathbb{R}} f(y) P(dy, t|x, s) \\
&= \int_{|y-x| \leq \varepsilon} f(y) P(dy, t|x, s) + \int_{|y-x| > \varepsilon} f(y) P(dy, t|x, s) \\
&\leq \int_{|y-x| \leq \varepsilon} f(y) P(dy, t|x, s) + \|f\|_{L^\infty} \int_{|y-x| > \varepsilon} P(dy, t|x, s) \\
&= \int_{|y-x| \leq \varepsilon} f(y) P(dy, t|x, s) + o(t-s).
\end{aligned}$$

We add and subtract the final condition $f(x)$ and use the previous calculation to obtain

$$\begin{aligned}
u(x, s) &= \int_{\mathbb{R}} f(y) P(dy, t|x, s) = f(x) + \int_{\mathbb{R}} (f(y) - f(x)) P(dy, t|x, s) \\
&= f(x) + \int_{|y-x| \leq \varepsilon} (f(y) - f(x)) P(dy, t|x, s) + \int_{|y-x| > \varepsilon} (f(y) - f(x)) P(dy, t|x, s) \\
&\stackrel{(2.41)}{=} f(x) + \int_{|y-x| \leq \varepsilon} (f(y) - f(x)) P(dy, t|x, s) + o(t-s).
\end{aligned}$$

The final condition follows from the fact that $f(x) \in C_b(\mathbb{R})$ and the arbitrariness of ε .

Now we show that $u(s, x)$ solves the backward Kolmogorov equation (2.47). We use the Chapman–Kolmogorov equation (2.16) to obtain

$$\begin{aligned}
u(x, \sigma) &= \int_{\mathbb{R}} f(z) P(dz, t|x, \sigma) = \int_{\mathbb{R}} \int_{\mathbb{R}} f(z) P(dz, t|y, \rho) P(dy, \rho|x, \sigma) \\
&= \int_{\mathbb{R}} u(y, \rho) P(dy, \rho|x, \sigma).
\end{aligned} \tag{2.48}$$

We use Taylor's theorem to obtain

$$u(z, \rho) - u(x, \rho) = \frac{\partial u(x, \rho)}{\partial x} (z - x) + \frac{1}{2} \frac{\partial^2 u(x, \rho)}{\partial x^2} (z - x)^2 (1 + \alpha_\varepsilon), \quad |z - x| \leq \varepsilon, \tag{2.49}$$

where

$$\alpha_\varepsilon = \sup_{\rho, |z-x| \leq \varepsilon} \left| \frac{\partial^2 u(x, \rho)}{\partial x^2} - \frac{\partial^2 u(z, \rho)}{\partial x^2} \right|$$

and $\lim_{\varepsilon \rightarrow 0} \alpha_\varepsilon = 0$.

We combine now (2.48) with (2.49) to calculate

$$\begin{aligned}
\frac{u(x, s) - u(x, s+h)}{h} &= \frac{1}{h} \left(\int_{\mathbb{R}} P(dy, s+h|x, s) u(y, s+h) - u(x, s+h) \right) \\
&= \frac{1}{h} \int_{\mathbb{R}} P(dy, s+h|x, s) (u(y, s+h) - u(x, s+h)) \\
&= \frac{1}{h} \int_{|x-y| < \varepsilon} P(dy, s+h|x, s) (u(y, s+h) - u(x, s+h)) + o(1)
\end{aligned}$$

$$\begin{aligned}
&= \frac{\partial u}{\partial x}(x, s+h) \frac{1}{h} \int_{|x-y|<\varepsilon} (y-x) P(dy, s+h|x, s) \\
&\quad + \frac{1}{2} \frac{\partial^2 u}{\partial x^2}(x, s+h) \frac{1}{h} \int_{|x-y|<\varepsilon} (y-x)^2 P(dy, s+h|x, s) (1 + \alpha_\varepsilon) + o(1) \\
&= b(x, s) \frac{\partial u}{\partial x}(x, s+h) + \frac{1}{2} \Sigma(x, s) \frac{\partial^2 u}{\partial x^2}(x, s+h) (1 + \alpha_\varepsilon) + o(1).
\end{aligned}$$

Equation (2.47) follows by taking the limits $\varepsilon \rightarrow 0$, $h \rightarrow 0$.

Note that the backward Kolmogorov equation (2.47) is a final value problem for a partial differential equation of parabolic type. For time-homogeneous diffusion processes, where the drift and the diffusion coefficients are independent of time, $b = b(x)$ and $\Sigma = \Sigma(x)$, we can rewrite it as an initial value problem. Let $T = t - s$, and introduce the function $U(x, T) = u(x, t - s)$. The backward Kolmogorov equation now becomes

$$\frac{\partial U}{\partial T} = b(x) \frac{\partial U}{\partial x} + \frac{1}{2} \Sigma(x) \frac{\partial^2 U}{\partial x^2}, \quad U(x, 0) = f(x). \quad (2.50)$$

In the time-homogeneous case, we can set the initial time $s = 0$. We then have that the conditional expectation $u(x, t) = \mathbb{E}(f(X_t | X_0 = x))$ is the solution to the initial value problem

$$\frac{\partial u}{\partial t} = b(x) \frac{\partial u}{\partial x} + \frac{1}{2} \Sigma(x) \frac{\partial^2 u}{\partial x^2}, \quad u(x, 0) = f(x). \quad (2.51)$$

The differential operator that appears on the right-hand side of (2.51) is the generator of the diffusion process X_t . In this book, we will use the backward Kolmogorov equation in the form (2.51).³

Assume now that the transition function has a density $p(y, t|x, s)$. In this case, the formula for $u(x, s)$ becomes

$$u(x, s) = \int_{\mathbb{R}} f(y) p(y, t|x, s) dy.$$

Substituting this into the backward Kolmogorov equation (2.47), we obtain

$$\int_{\mathbb{R}} f(y) \left(\frac{\partial p(y, t|x, s)}{\partial s} + \mathcal{L}_{s,x} p(y, t|x, s) \right) = 0, \quad (2.52)$$

where

$$\mathcal{L}_{s,x} := b(x, s) \frac{\partial}{\partial x} + \frac{1}{2} \Sigma(x, s) \frac{\partial^2}{\partial x^2}.$$

³ The backward Kolmogorov equation can also be derived using Itô's formula. See Chap. 3.

Equation (2.52) is valid for arbitrary continuous bounded functions f . Consequently, from (2.52) we obtain a partial differential equation for the transition probability density:

$$-\frac{\partial p(y,t|x,s)}{\partial s} = b(x,s) \frac{\partial p(y,t|x,s)}{\partial x} + \frac{1}{2} \Sigma(x,s) \frac{\partial^2 p(y,t|x,s)}{\partial x^2}. \quad (2.53)$$

Notice that the variation is with respect to the “backward” variables x,s .

The Forward Kolmogorov Equation

Assume that the transition function has a density with respect to the Lebesgue measure that is a smooth function of its arguments:

$$P(dy,t|x,s) = p(y,t|x,s) dy.$$

We can obtain an equation with respect to the “forward” variables y,t , namely the forward Kolmogorov or Fokker–Planck equation.

Theorem 2.2 (Kolmogorov). *Assume that conditions (2.41)–(2.43) are satisfied and that $p(y,t|\cdot,\cdot)$, $b(y,t)$, $\Sigma(y,t)$ are smooth functions of y,t . Then the transition probability density is the solution to the initial value problem*

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial y} (b(t,y)p) + \frac{1}{2} \frac{\partial^2}{\partial y^2} (\Sigma(t,y)p), \quad p(s,y|x,s) = \delta(x-y). \quad (2.54)$$

Proof. The initial condition follows from the definition of the transition probability density $p(y,t|x,s)$. Fix now a function $f(y) \in C_0^2(\mathbb{R})$. An argument similar to that used in the proof of the backward Kolmogorov equation gives

$$\lim_{h \rightarrow 0} \frac{1}{h} \left(\int f(y) p(y, s+h|x, s) dy - f(x) \right) = b(x, s) \frac{df}{dx}(x) + \frac{1}{2} \Sigma(x, s) \frac{d^2 f}{dx^2}(x), \quad (2.55)$$

where subscripts denote differentiation with respect to x . On the other hand,

$$\begin{aligned} \int f(y) \frac{\partial}{\partial t} p(y, t|x, s) dy &= \frac{\partial}{\partial t} \int f(y) p(y, t|x, s) dy \\ &= \lim_{h \rightarrow 0} \frac{1}{h} \int (p(y, t+h|x, s) - p(y, t|x, s)) f(y) dy \\ &= \lim_{h \rightarrow 0} \frac{1}{h} \left(\int p(y, t+h|x, s) f(y) dy - \int p(z, t|s, x) f(z) dz \right) \\ &= \lim_{h \rightarrow 0} \frac{1}{h} \left(\int \int p(y, t+s|z, t) p(z, t|x, s) f(y) dy dz - \int p(z, t|s, x) f(z) dz \right) \\ &= \lim_{h \rightarrow 0} \frac{1}{h} \left(\int p(z, t|x, s) \left(\int p(y, t+h|z, t) f(y) dy - f(z) \right) dz \right) \end{aligned}$$

$$\begin{aligned}
&= \int p(z, t|x, s) \left(b(z, t) \frac{df}{dz}(z) + \frac{1}{2} \Sigma(z) \frac{d^2 f}{dz^2}(z) \right) dz \\
&= \int \left(-\frac{\partial}{\partial z} (b(z, t)p(z, t|x, s)) + \frac{1}{2} \frac{\partial^2}{\partial z^2} (\Sigma(z, t)p(z, t|x, s)) \right) f(z) dz.
\end{aligned}$$

In the above calculation, we used the Chapman–Kolmogorov equation. We have also performed two integrations by parts and used the fact that since the test function f has compact support, the boundary terms vanish. Since the above equation is valid for every test function f , the forward Kolmogorov equation follows.

Assume now that the initial distribution of X_t is $\rho_0(x)$, and set $s = 0$ (the initial time) in (2.54). Define

$$p(y, t) := \int p(y, t|x, 0) \rho_0(x) dx. \quad (2.56)$$

We multiply the forward Kolmogorov equation (2.54) by $\rho_0(x)$ and integrate with respect to x to obtain the equation

$$\frac{\partial p(y, t)}{\partial t} = -\frac{\partial}{\partial y} (a(y, t)p(y, t)) + \frac{1}{2} \frac{\partial^2}{\partial y^2} (b(y, t)p(y, t)), \quad (2.57)$$

together with the initial condition

$$p(y, 0) = \rho_0(y). \quad (2.58)$$

The solution of equation (2.57) provides us with the probability that the diffusion process X_t , which initially was distributed according to the probability density $\rho_0(x)$, is equal to y at time t . Alternatively, we can think of the solution to (2.54) as the Green's function for the partial differential equation (2.57). Using (2.57), we can calculate the expectation of an arbitrary function of the diffusion process X_t :

$$\begin{aligned}
\mathbb{E}(f(X_t)) &= \int \int f(y) p(y, t|x, 0) p(x, 0) dx dy \\
&= \int f(y) p(y, t) dy,
\end{aligned}$$

where $p(y, t)$ is the solution of (2.57).

The solution of the Fokker–Planck equation provides us with the transition probability density. The Markov property enables us to calculate joint probability densities using Eq. (2.20). For example, let X_t denote a diffusion process with $X_0 \sim \pi$, let $0 = t_0 < t_1 < \dots < t_n$, and let $f(x_0, \dots, x_n)$ be a measurable function. Denoting by \mathbb{E}_π the expectation with respect to π , we have

$$\mathbb{E}_\pi f(X_{t_0}, X_{t_1}, \dots, X_{t_n}) = \int \dots \int f(x_0, \dots, x_n) \pi(x_0) dx_0 \prod_{j=1}^n p(x_j, t_j | x_{j-1}, t_{j-1}) dx_j. \quad (2.59)$$

In particular, the autocorrelation function of X_t at times t and 0 is given by the formula

$$C(t) := \mathbb{E}_\pi(X_t X_0) = \int \int y x p(y, t | x, 0) \pi(x) dx dy. \quad (2.60)$$

Multidimensional Diffusion Processes

The backward and forward Kolmogorov equations can be derived for multidimensional diffusion processes using the same calculations and arguments that were used in the proofs of Theorems 2.1 and 2.2. Let X_t be a diffusion process in \mathbb{R}^d . The drift and diffusion coefficients of a diffusion process in \mathbb{R}^d are defined as follows, for all $\varepsilon > 0$:

$$\lim_{t \rightarrow s} \frac{1}{t-s} \int_{|y-x|<\varepsilon} (y-x) P(dy, t | x, s) = \mathbf{b}(x, s)$$

and

$$\lim_{t \rightarrow s} \frac{1}{t-s} \int_{|y-x|<\varepsilon} (y-x) \otimes (y-x) P(dy, t | x, s) = \Sigma(x, s).$$

The drift coefficient $\mathbf{b}(x, s)$ is a d -dimensional vector field, and the diffusion coefficient $\Sigma(x, s)$ is a $d \times d$ symmetric nonnegative matrix. The generator of a d -dimensional diffusion process is

$$\begin{aligned} \mathcal{L} &= \mathbf{b}(x, s) \cdot \nabla + \frac{1}{2} \Sigma(x, s) : \nabla \nabla \\ &= \sum_{j=1}^d b_j(x, s) \frac{\partial}{\partial x_j} + \frac{1}{2} \sum_{i,j=1}^d \Sigma_{ij}(x, s) \frac{\partial^2}{\partial x_i \partial x_j}. \end{aligned}$$

Assuming that the first and second moments of the multidimensional diffusion process exist, we can write the formulas for the drift vector and diffusion matrix as

$$\lim_{t \rightarrow s} \mathbb{E} \left(\frac{X_t - X_s}{t-s} \middle| X_s = x \right) = \mathbf{b}(x, s) \quad (2.61)$$

and

$$\lim_{t \rightarrow s} \mathbb{E} \left(\frac{(X_t - X_s) \otimes (X_t - X_s)}{t-s} \middle| X_s = x \right) = \Sigma(x, s). \quad (2.62)$$

The backward and forward Kolmogorov equations for multidimensional diffusion processes are

$$-\frac{\partial u}{\partial s} = \mathbf{b}(x, s) \cdot \nabla_x u + \frac{1}{2} \Sigma(x, s) : \nabla_x \nabla_x u, \quad u(t, x) = f(x), \quad (2.63)$$

and

$$\frac{\partial p}{\partial t} = \nabla_y \cdot \left(-\mathbf{b}(t, \mathbf{y})p + \frac{1}{2} \nabla_y \cdot (\Sigma(t, \mathbf{y})p) \right), \quad p(\mathbf{y}, s | \mathbf{x}, s) = \delta(x - y). \quad (2.64)$$

As for one-dimensional time-homogeneous diffusion processes, the backward Kolmogorov equation for a time-homogeneous multidimensional diffusion process can be written as an initial value problem:

$$\frac{\partial u}{\partial t} = \mathcal{L}u, \quad u(x, 0) = f(x), \quad (2.65)$$

for $u(x, t) = \mathbb{E}(f(X_t) | X_0 = x)$. For time-homogeneous processes, we can fix the initial time $s = 0$ in the forward Kolmogorov equation. Assuming furthermore that X_0 is a random variable with probability density $\rho_0(x)$, we see that the forward Kolmogorov equation becomes

$$\frac{\partial p}{\partial t} = \mathcal{L}^* p, \quad p(x, 0) = \rho_0(x), \quad (2.66)$$

for the transition probability density $p(y, t)$. In this book, we will use the forward and backward Kolmogorov equations in the forms (2.65) and (2.66).

2.6 Discussion and Bibliography

Markov chains in both discrete and continuous time are studied in [176, 228]. A standard reference on Markov processes is [53]. The proof that the transition function and the initial distribution of X_t are sufficient to determine uniquely a Markov process Eq. (2.17) can be found in [200, Proposition 1.4, Chap. III]. See also [53, Theorem 1.1, Chap. 4].

Operator semigroups are the main analytical tool for the study of diffusion processes; see, for example, [146]. Necessary and sufficient conditions for an operator \mathcal{L} to be the generator of a (contraction) semigroup are given by the Hille–Yosida theorem [55, Chap. 7].

The space $C_b(E)$ is natural in a probabilistic context for the study of Markov semigroups, but other function spaces often arise in applications; in particular, when there is a measure μ on E , the spaces $L^p(E; \mu)$ sometimes arise. We will quite often use the space $L^2(E; \mu)$, where μ is an invariant measure of the Markov process. Markov semigroups can be extended from the space of bounded continuous functions to the space $L^p(E; \mu)$ for every $p \geq 1$. The proof of this result, which follows from Jensen’s inequality and the Hahn–Banach theorem, can be found in [77, Proposition 1.14].

The generator is frequently taken as the starting point for the definition of a homogeneous Markov process. Conversely, let P_t be a *contraction semigroup* (let X be a Banach space and $T : X \rightarrow X$ a bounded operator; then T is a contraction if

$\|Tf\|_X \leq \|f\|_X \forall f \in X$), with $\mathcal{D}(P_t) \subset C_b(E)$, closed. Then under mild technical hypotheses, there is an E -valued homogeneous Markov process $X(t)$ associated with P_t defined through

$$\mathbb{E}[f(X(t)|\mathcal{F}_s^X)] = P_{t-s}f(X(s))$$

for all $t, s \in T$ with $t \geq s$ and $f \in \mathcal{D}(P_t)$.

The argument used in the derivation of the forward and backward Kolmogorov equations goes back to Kolmogorov's original work. See [70, 93]. A more modern approach to the derivation of the forward equation from the Chapman–Kolmogorov equation can be found in [226, Chap. 1]. The connection between Brownian motion and the corresponding Fokker–Planck equation, which is the heat equation, was made by Einstein [52]. Many of the early papers on the theory of stochastic processes have been reprinted in [239]. Early papers on Brownian motion, including the original papers by Fokker and by Planck, are available from <http://www.physik.uni-augsburg.de/theo1/hanggi/History/BM-History.html>. Very interesting historical comments can also be found in [161, 171].

We can also derive backward and forward Kolmogorov equations for continuous-time Markov processes with jumps. For such processes, an additional nonlocal-in-space term (an integral operator) appears in the Kolmogorov equations that accounts for the jumps.⁴ Details can be found in [68].

A diffusion process is characterized by the (almost sure) continuity of its paths and by specifying the first two moments. A natural question that arises is whether other types of stochastic processes can be defined by specifying a fixed number of moments greater than two. It turns out that this is not possible: we either need to retain two or all (i.e., infinitely many) moments. Specifying a finite number of moments greater than two leads to inconsistencies. This is the content of Pawula's theorem [186]. Example 2.3 can also be found in [185, Chap. 6].

The duality between the backward and forward Kolmogorov equations, the duality between studying the evolution of observables and states, is similar to the duality between the Heisenberg (evolution of observables) and Schrödinger (evolution of states) representations of quantum mechanics or the Koopman (evolution of observables) and Frobenius–Perron (evolution of states) operators in the theory of dynamical systems, i.e., the duality between the study of the evolution of observables and of states. See [134, 195, 232].

⁴ The generator of a Markov process with jumps is necessarily nonlocal: a local (differential) operator \mathcal{L} corresponds to a Markov process with continuous paths. See [226, Chap. 1].

2.7 Exercises

1. Let $\{X_n\}$ be a stochastic process with state space $S = \mathbb{Z}$. Show that it is a Markov process if and only if for all n ,

$$\mathbb{P}(X_{n+1} = i_{n+1} | X_1 = i_1, \dots, X_n = i_n) = \mathbb{P}(X_{n+1} = i_{n+1} | X_n = i_n).$$

2. Show that (2.6) is the solution of the initial value problem (2.12) as well as of the final value problem

$$-\frac{\partial p}{\partial s} = \frac{1}{2} \frac{\partial^2 p}{\partial x^2}, \quad \lim_{s \rightarrow t} p(y, t | x, s) = \delta(y - x).$$

3. Use (2.7) to show that the forward and backward Kolmogorov equations for the Ornstein–Uhlenbeck process are

$$\frac{\partial p}{\partial t} = \frac{\partial}{\partial y}(yp) + \frac{1}{2} \frac{\partial^2 p}{\partial y^2}$$

and

$$-\frac{\partial p}{\partial s} = -x \frac{\partial p}{\partial x} + \frac{1}{2} \frac{\partial^2 p}{\partial x^2}.$$

4. Let $W(t)$ be a standard one-dimensional Brownian motion, let $Y(t) = \sigma W(t)$ with $\sigma > 0$, and consider the process

$$X(t) = \int_0^t Y(s) ds.$$

Show that the joint process $\{X(t), Y(t)\}$ is Markovian and write down the generator of the process.

5. Let $Y(t) = e^{-t} W(e^{2t})$ be the stationary Ornstein–Uhlenbeck process and consider the process

$$X(t) = \int_0^t Y(s) ds.$$

Show that the joint process $\{X(t), Y(t)\}$ is Markovian and write down the generator of the process.

6. (a) Let X, Y be mean-zero Gaussian random variables with $\mathbb{E}X^2 = \sigma_X^2$, $\mathbb{E}Y^2 = \sigma_Y^2$ and correlation coefficient ρ (the correlation coefficient is $\rho = \frac{\mathbb{E}(XY)}{\sigma_X \sigma_Y}$). Show that

$$\mathbb{E}(X|Y) = \frac{\rho \sigma_X}{\sigma_Y} Y.$$

- (b) Let X_t be a mean-zero stationary Gaussian process with autocorrelation function $R(t)$. Use the previous result to show that

$$\mathbb{E}[X_{t+s}|X_s] = \frac{R(t)}{R(0)} X(s), \quad s, t \geq 0.$$

- (c) Use the previous result to show that the only stationary Gaussian Markov process with continuous autocorrelation function is the stationary Ornstein–Uhlenbeck process.
7. Show that a Gaussian process X_t is a Markov process if and only if
- $$\mathbb{E}(X_{t_n} | X_{t_1} = x_1, \dots, X_{t_{n-1}} = x_{n-1}) = \mathbb{E}(X_{t_n} | X_{t_{n-1}} = x_{n-1}).$$
8. Prove Eq. (2.55).
9. Derive the initial value problem (2.57), (2.58).
10. Prove Theorems 2.1 and 2.2 for multidimensional diffusion processes.

Chapter 3

Introduction to Stochastic Differential Equations

In this chapter, we study diffusion processes at the level of paths. In particular, we study stochastic differential equations (SDEs) driven by Gaussian white noise, defined formally as the derivative of Brownian motion. In Sect. 3.1, we introduce SDEs. In Sect. 3.2, we introduce the Itô and Stratonovich stochastic integrals. In Sect. 3.3, we present the concept of a solution to an SDE. The generator, Itô's formula, and the connection with the Fokker–Planck equation are covered in Sect. 3.4. Examples of SDEs are presented in Sect. 3.5. The Lamperti transformation and Girsanov's theorem are discussed briefly in Sect. 3.6. Linear SDEs are studied in Sect. 3.7. Bibliographical remarks and exercises can be found in Sects. 3.8 and 3.9, respectively.

3.1 Introduction

We consider SDEs of the form

$$\frac{dX(t)}{dt} = b(t, X(t)) + \sigma(t, X(t))\xi(t), \quad X(0) = x, \quad (3.1)$$

where $X(t) \in \mathbb{R}^d$, $b : [0, T] \times \mathbb{R}^d \mapsto \mathbb{R}^d$, and $\sigma : [0, T] \times \mathbb{R}^d \mapsto \mathbb{R}^{d \times m}$. We use the notation $\xi(t) = \frac{dW}{dt}$ to denote (formally) the derivative of Brownian motion in \mathbb{R}^m , i.e., the white noise process, which is a (generalized) mean-zero Gaussian vector-valued stochastic process with autocorrelation function

$$\mathbb{E}(\xi_i(t)\xi_j(s)) = \delta_{ij}\delta(t-s), \quad i, j = 1, \dots, m. \quad (3.2)$$

The initial condition x can be either deterministic or a random variable that is independent of the Brownian motion $W(t)$, in which case there are two different,

independent, sources of randomness in (3.1). We will use different notations for the solution of an SDE:

$$X(t), \quad X_t \quad \text{or} \quad X_t^x.$$

The latter notation will be used when we want to emphasize the dependence of the solution of the initial conditions.

We will consider mostly autonomous SDEs, i.e., equations whose coefficients do not depend explicitly on time. When we study stochastic resonance and Brownian motors in Sects. 7.4 and 7.5, respectively, it will be necessary to consider SDEs with time-dependent coefficients. It is also often useful to consider SDEs in bounded domains, for example in a box of size L , $[0, L]^d$, with periodic boundary conditions; see Sect. 6.6, on Brownian motion in periodic potentials.

The amplitude of the noise in (3.1) may be independent of the state of the system, $\sigma(x) \equiv \sigma$, a constant, in which case we will say that the noise in (3.1) is *additive*. When the amplitude of the noise depends on the state of the system, we will say that the noise in (3.1) is *multiplicative*. In the modeling of physical systems using SDEs, additive noise is usually due to thermal fluctuations, whereas multiplicative noise is due to noise in some control parameter.

Example 3.1. Consider the Landau–Stuart equation

$$\frac{dX}{dt} = X(\alpha - X^2),$$

where α is a parameter. Assume that this parameter fluctuates randomly in time or that we are uncertain about its actual value. Modeling this uncertainty as white noise, $\alpha \mapsto \alpha + \sigma \xi$, we obtain the stochastic Landau equation with multiplicative noise:

$$\frac{dX_t}{dt} = X_t(\alpha - X_t^2) + \sigma X_t \xi(t). \quad (3.3)$$

It is important to note that an equation of the form (3.3) is not sufficient for determining uniquely the stochastic process X_t ; we also need to determine how we choose to interpret the noise in the equation, e.g., whether the noise in (3.3) is Itô or Stratonovich. This is a separate modeling issue, which we will address in Sect. 5.1.

Since the white noise process $\xi(t)$ is defined only in a generalized sense, Eq. (3.1) is only formal. We will usually write it in the form

$$dX(t) = \mathbf{b}(t, X(t)) dt + \sigma(t, X(t)) dW(t), \quad (3.4)$$

together with the initial condition $X(0) = x$, or componentwise,

$$dX_i(t) = b_i(t, X(t)) dt + \sum_{j=1}^m \sigma_{ij}(t, X(t)) dW_j(t), \quad j = 1, \dots, d, \quad (3.5)$$

together with the initial conditions. In fact, the correct interpretation of the SDE (3.4) is as a stochastic integral equation

$$X(t) = x + \int_0^t \mathbf{b}(t, X(t)) dt + \int_0^t \sigma(t, X(t)) dW(t). \quad (3.6)$$

Even when writing the SDE as an integral equation, we are still facing several mathematical difficulties. First, we need to give an appropriate definition of the stochastic integral

$$I(t) := \int_0^t \sigma(t, X(t)) dW(t), \quad (3.7)$$

or more generally,

$$I(t) := \int_0^t h(t) dW(t), \quad (3.8)$$

for a sufficiently large class of functions. Since Brownian motion is not of bounded variation, the integral (3.7) cannot be defined as a Riemann–Stieltjes integral in a unique way. As we will see in the next section, different Riemann–Stieltjes approximations lead to different stochastic integrals, which, in turn, lead to SDEs with different properties.

After defining the stochastic integral in (3.6), we need to give a proper definition of a solution to an SDE. In particular, we need to give a definition that takes into account the randomness due to the Brownian motion and the initial conditions. Furthermore, we need to take into account the fact that since Brownian motion is not regular but only Hölder continuous with exponent $\alpha < 1/2$, solutions to an SDE of the form (3.1) cannot be very regular. As in the case of partial differential equations, there are different concepts of solution for an SDE of the form (3.1).

After having given an appropriate definition for the stochastic integral and developed an existence and uniqueness theory of solutions to SDEs, we would like to be able to calculate (the statistics of) functionals of the solution to the SDE. Let $X(t)$ be the solution of (3.4) and let $f(t, x)$ be a sufficiently regular function of t and x . We want to derive an equation for the function

$$z(t) = f(t, X(t)).$$

In the absence of noise, we can easily obtain an equation for $z(t)$ using the chain rule. However, the stochastic forcing in (3.1) and the lack of regularity of Brownian motion imply that the chain rule has to be modified appropriately. This is, roughly speaking, due to the fact that $\mathbb{E}(dW(t))^2 = dt$, and consequently, second-order differentials need to be kept in calculating the differential of $z(t)$. It turns out that whether a correction to the chain rule from standard calculus is needed depends on how we interpret the stochastic integral (3.7).

Furthermore, we would like to be able to calculate the statistics of solutions to SDEs. In Chap. 2, we saw that we can calculate the expectation value of an observable

$$u(x, t) = \mathbb{E}(f(X_t^x) | X_0^x = x) \quad (3.9)$$

by solving the backward Kolmogorov equation (2.47). In this chapter, we will see that the backward Kolmogorov equation is a consequence of *Itô's formula*, the chain rule of *Itô stochastic calculus*.

Quite often, it is important to be able to evaluate the statistics of solutions to SDEs at appropriate random times, the so-called *stopping times*. An example of a stopping time is the *first exit time* of the solution of an SDE of the form (3.1), which is defined as the first time the diffusion X_t^x exits an open domain $D \in \mathbb{R}^d$, with $x \in D$:

$$\tau_D = \inf_{t>0} \left\{ X_t \notin D \right\}. \quad (3.10)$$

The statistics of the first exit time will be needed in the calculation of the escape time of a diffusion process from a metastable state; see Chap. 7.

There is also an important modeling issue that we need to address: white noise is a stochastic process with zero correlation time. As such, it can be thought of only as an idealization of the noise that appears in physical, chemical, and biological systems. We are interested, therefore, in understanding the connection between an SDE driven by white noise and equations in which a more physically realistic noise is present. The question then is whether an SDE driven by white noise (3.1) can be obtained from an equation driven by noise with a nontrivial correlation structure through an appropriate limiting procedure. We will see in Sect. 5.1 that it is possible to describe more general classes of noise within the framework of diffusion processes by adding additional variables. Furthermore, we can obtain the SDE (3.1) in the limit of zero correlation time and for an appropriate definition of the stochastic integral.

3.2 The Itô and Stratonovich Stochastic Integrals

In this section, we define stochastic integrals of the form

$$I(t) = \int_0^t f(s) dW(s), \quad (3.11)$$

where $W(t)$ is a standard one-dimensional Brownian motion and $t \in [0, T]$. We are interested in the case that the integrand is a stochastic process whose randomness depends on the Brownian motion $W(t)$ —think of the stochastic integral in (3.6)—and in particular, that it is *adapted* to the *filtration* \mathcal{F}_t [see (2.14)] generated by the Brownian motion $W(t)$, i.e., that it is an \mathcal{F}_t -measurable function for all $t \in [0, T]$. Roughly speaking, this means that the integrand depends only the past history of the Brownian motion with respect to which we are integrating in (3.11). Furthermore, we will assume that the random process $f(\cdot)$ is square integrable:

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

Our goal is to define the stochastic integral $I(t)$ as the L^2 -limit of a Riemann sum approximation of (3.11). To this end, we introduce a partition of the interval $[0, T]$ by

setting $t_k = k\Delta t$, $k = 0, \dots, K-1$, and $K\Delta t = t$; we also define a parameter $\lambda \in [0, 1]$ and set

$$\tau_k = (1 - \lambda)t_k + \lambda t_{k+1}, \quad k = 0, \dots, K-1. \quad (3.12)$$

We define now the stochastic integral as the $L^2(\Omega)$ limit (Ω denoting the underlying probability space) of the Riemann sum approximation

$$I(t) := \lim_{K \rightarrow \infty} \sum_{k=0}^{K-1} f(\tau_k) (W(t_{k+1}) - W(t_k)). \quad (3.13)$$

In contrast to the case of the Riemann–Stieltjes integral, when we integrate against a smooth deterministic function, the result in (3.13) depends on the choice of $\lambda \in [0, 1]$ in (3.12). The two most common choices are $\lambda = 0$, in which case we obtain the *Itô stochastic integral*

$$I_I(t) := \lim_{K \rightarrow \infty} \sum_{k=0}^{K-1} f(t_k) (W(t_{k+1}) - W(t_k)), \quad (3.14)$$

and $\lambda = \frac{1}{2}$, which leads to the *Stratonovich stochastic integral*

$$I_S(t) := \lim_{K \rightarrow \infty} \sum_{k=0}^{K-1} f\left(\frac{1}{2}(t_k + t_{k+1})\right) (W(t_{k+1}) - W(t_k)). \quad (3.15)$$

We will use the notation

$$I_S(t) = \int_0^t f(s) \circ dW(s)$$

to denote the Stratonovich stochastic integral. In general, the Itô and Stratonovich stochastic integrals are different. When the integrand $f(t)$ depends on the Brownian motion $W(t)$ through $X(t)$, the solution of the SDE in (3.1), a formula exists for converting one stochastic integral into another.

When the integrand in (3.14) is a sufficiently smooth function, then the stochastic integral is independent of the parameter λ , and in particular, the Itô and Stratonovich stochastic integrals coincide.

Proposition 3.1. *Assume that there exist $C, \delta > 0$ such that*

$$\mathbb{E}(f(t) - f(s))^2 \leq C|t - s|^{1+\delta}, \quad 0 \leq s, t \leq T. \quad (3.16)$$

Then the Riemann sum approximation in (3.13) converges in $L^1(\Omega)$ to the same value for all $\lambda \in [0, 1]$.

The interested reader is invited to provide a proof of this proposition.

Example 3.2. Consider the Langevin equation (see Chap. 6) with a space-dependent friction coefficient:

$$m\ddot{q}_t = -\nabla V(q_t) - \gamma(q_t)\dot{q}_t + \sqrt{2\gamma(q_t)}\dot{W}_t,$$

where q denotes the position of a particle of mass m . Writing the Langevin equation as a system of first-order SDEs, we have

$$mdq_t = p_t dt, \quad (3.17a)$$

$$dp_t = -\nabla V(q_t) dt - \gamma(q_t) p_t dt + \sqrt{2\gamma(q_t)} dW_t. \quad (3.17b)$$

Assuming that the potential and the coefficients of friction are smooth functions, the particle's position q is a differentiable function of time.¹ Consequently, according to Proposition 3.1, the Itô and Stratonovich stochastic integrals in (3.17) coincide. This is not true in the limit of small mass, $m \rightarrow 0$. See Sect. 5.1 and the discussion in Sect. 3.8.

The Itô stochastic integral $I(t)$ is continuous almost surely in t . As expected, it satisfies the linearity property

$$\int_0^T (\alpha f(t) + \beta g(t)) dW(t) = \alpha \int_0^T f(t) dW(t) + \beta \int_0^T g(t) dW(t), \quad \alpha, \beta \in \mathbb{R},$$

for all square-integrable functions $f(t), g(t)$. Furthermore, the Itô stochastic integral satisfies the *Itô isometry*

$$\mathbb{E} \left(\int_0^T f(t) dW(t) \right)^2 = \int_0^T \mathbb{E} |f(t)|^2 dt, \quad (3.18)$$

from which it follows that for all square-integrable functions f, g (we use the Itô isometry with $f = h + g$),

$$\mathbb{E} \left(\int_0^T h(t) dW(t) \int_0^T g(s) dW(s) \right) = \mathbb{E} \int_0^T h(t) g(t) dt.$$

The Itô stochastic integral is a *martingale*:

Definition 3.1. Let $\{\mathcal{F}_t\}_{t \in [0, T]}$ be a filtration defined on the probability space $(\Omega, \mathcal{F}, \mu)$, and let $\{\mathcal{M}_t\}_{t \in [0, T]}$ be adapted to \mathcal{F}_t with $\mathcal{M}_t \in L^1(0, T)$. We say that \mathcal{M}_t is an \mathcal{F}_t martingale if

$$\mathbb{E}[\mathcal{M}_t | \mathcal{F}_s] = \mathcal{M}_s \quad \forall t \geq s.$$

For the Itô stochastic integral, we have

$$\mathbb{E} \int_0^t f(s) dW(s) = 0 \quad (3.19)$$

and

$$\mathbb{E} \left[\int_0^t f(\ell) dW(\ell) | \mathcal{F}_s \right] = \int_0^s f(\ell) dW(\ell) \quad \forall t \geq s, \quad (3.20)$$

¹ In fact, it is a $C^{1+\alpha}$ function of time, with $\alpha < 1/2$.

where \mathcal{F}_s denotes the filtration generated by $W(s)$. The quadratic variation of this martingale is

$$\langle I \rangle_t = \int_0^t (f(s))^2 ds.$$

The proofs of all these properties and the study of the Riemann sum (3.13) proceed as follows: first, these properties are proved for the simplest possible functions, namely step functions for which we can perform explicit calculations using the properties of Brownian increments. Then, an approximation step is used to show that square-integrable functions can be approximated by step functions. The details of these calculations can be found in the references listed in Sect. 3.8.

The above ideas are readily generalized to the case in which $W(t)$ is a standard d -dimensional Brownian motion and $f(t) \in \mathbb{R}^{m \times d}$ for each $t > 0$. In the multidimensional case, the Itô isometry takes the form

$$\mathbb{E}|I(t)|^2 = \int_0^t \mathbb{E}|f(s)|_F^2 ds, \quad (3.21)$$

where $|\cdot|_F$ denotes the Frobenius norm $|A|_F = \sqrt{\text{tr}(A^T A)}$.

Whether we choose the Itô or Stratonovich interpretation of the stochastic integral in (3.6) is a modeling issue that we will address later, in Sect. 5.1. Both interpretations have their advantages: The Itô stochastic integral is a martingale, and we can use the well-developed theory of martingales to study its properties; in particular, there are many inequalities and limit theorems for martingales that are very useful in the rigorous study of qualitative properties of solutions to SDEs. On the other hand, the Stratonovich stochastic integral leads to the standard Newton–Leibniz chain rule, in contrast to the Itô stochastic integral, which requires a correction to the Leibniz chain rule. Furthermore, as we will see in Sect. 5.1, SDEs driven by noise with nonzero correlation time converge, in the limit as the correlation time tends to 0, to the Stratonovich SDE. When the stochastic integral in (3.4) or (3.6) is an Itô integral, we will refer to the SDE as an *Itô SDE*, whereas when it is a Stratonovich integral, we will refer to the SDE as a *Stratonovich SDE*. There are other interpretations of the stochastic integral that arise in applications, e.g., the Klimontovich (kinetic) stochastic integral, which corresponds to the choice $\lambda = 1$ in (3.12).

An Itô SDE can be converted into a Stratonovich SDE and conversely. This transformation involves the addition (or subtraction) of a drift term. We calculate this correction to the drift in one dimension. Consider the Itô SDE

$$dX_t = b(X_t) dt + \sigma(X_t) dW_t. \quad (3.22)$$

We want to write it in Stratonovich form:

$$dX_t = \widehat{b}(X_t) dt + \widehat{\sigma}(X_t) \circ dW_t. \quad (3.23)$$

Let us calculate the Stratonovich stochastic integral in (3.23). We have, with $\alpha = \frac{1}{2}$ and using the notation $\Delta W_j = W(t_{j+1}) - W(t_j)$ and similarly for ΔX_j as well as a Taylor series expansion,

$$\begin{aligned}
\int_0^t \widehat{\sigma}(X_t) \circ dW_t &\approx \sum_j \widehat{\sigma}(X(j\Delta t + \alpha\Delta t)) \Delta W_j \\
&\approx \sum_j \widehat{\sigma}(X(j\Delta t)) \Delta W_j + \alpha \sum_j \frac{d\widehat{\sigma}}{dx}(X(j\Delta t)) \Delta X_j \Delta W_j \\
&\approx \int_0^t \widehat{\sigma}(X(t)) dW(t) + \alpha \sum_j \frac{d\widehat{\sigma}}{dx}(X(j\Delta t)) \left(b(X_j) \Delta t_j + \sigma(X_j) \Delta W_j \right) \Delta W_j \\
&\approx \int_0^t \widehat{\sigma}(X(t)) dW(t) + \alpha \int_0^t \frac{d\widehat{\sigma}}{dx}(X(t)) \sigma(X(t)) dt.
\end{aligned} \tag{3.24}$$

In the above calculation, we have used the formulas $\mathbb{E}(\Delta t \Delta W_j) = 0$ and $\mathbb{E}(\Delta W_j)^2 = \Delta t$; see Sect. 3.4. This calculation suggests that the Stratonovich stochastic integral, when evaluated at the solution of the Itô SDE (3.22), is equal to the Itô stochastic integral plus a drift correction. Note that the above calculation provides us with a correction for arbitrary choices of the parameter $\alpha \in [0, 1]$. The above heuristic argument can be made rigorous: we need to control the difference between $\int_0^t \widehat{\sigma}(X_t) \circ dW_t$ and the right-hand side of (3.24) in $L^2(\Omega)$; see Exercise 2.

Substituting (3.24) into the Stratonovich SDE (3.23), we obtain

$$dX_t = \left(\widehat{b}(X_t) + \frac{1}{2} \frac{d\widehat{\sigma}}{dx}(X_t) \sigma(X_t) \right) dt + \widehat{\sigma}(X_t) dW_t.$$

This is the Itô equation (3.22). Comparing the drift and the diffusion coefficients, we deduce that

$$\widehat{\sigma} = \sigma \quad \text{and} \quad \widehat{b} = b - \frac{1}{2} \sigma' \sigma. \tag{3.25}$$

Consequently, the Itô SDE

$$dX_t = b(X_t) dt + \sigma(X_t) dW_t$$

is equivalent to the Stratonovich SDE

$$dX_t = \left(b(X_t) - \frac{1}{2} \sigma'(X_t) \sigma(X_t) \right) dt + \sigma(X_t) \circ dW_t.$$

Conversely, the Stratonovich SDE

$$dX_t = b(X_t) dt + \sigma(X_t) \circ dW_t$$

is equivalent to the Itô SDE

$$dX_t = \left(b(X_t) + \frac{1}{2} \sigma'(X_t) \sigma(X_t) \right) dt + \sigma(X_t) dW_t.$$

The correction to the drift $\frac{1}{2} \sigma' \sigma$ is called the *Itô-to-Stratonovich correction*. Similar formulas can be obtained in arbitrary dimensions. The multidimensional Itô SDE

$$dX_t = \mathbf{b}(X_t) dt + \sigma(X_t) dW_t, \quad (3.26)$$

where $\mathbf{b} : \mathbb{R}^d \mapsto \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \mapsto \mathbb{R}^{d \times m}$, can be transformed into the Stratonovich SDE

$$dX_t = (\mathbf{b}(X_t) - \mathbf{h}(X_t)) dt + \sigma(X_t) \circ dW_t, \quad (3.27)$$

where the correction drift h is given by the formula

$$h_i(x) = \frac{1}{2} \sum_{j=1}^d \sum_{k=1}^m \sigma_{jk}(x) \frac{\partial \sigma_{ik}}{\partial x_j}(x), \quad i = 1, \dots, d. \quad (3.28)$$

Conversely, the multidimensional Stratonovich SDE

$$dX_t = \mathbf{b}(X_t) dt + \sigma(X_t) \circ dW_t \quad (3.29)$$

can be transformed into the Itô SDE

$$dX_t = (\mathbf{b}(X_t) + \mathbf{h}(X_t)) dt + \sigma(X_t) dW_t, \quad (3.30)$$

with h given by (3.28). The Itô-to-Stratonovich correction can be written in index-free notation:

$$\mathbf{h}(x) = \frac{1}{2} [\nabla \cdot \Sigma(x) - (\sigma \nabla \cdot \sigma^T)(x)], \quad \Sigma = \sigma \sigma^T. \quad (3.31)$$

To see this, we first note that

$$(\nabla \cdot \Sigma)_i = \sum_{k=1}^d \frac{\partial \Sigma_{ik}}{\partial x_k} = \sum_{k=1}^d \sum_{\ell=1}^m \left(\frac{\partial \sigma_{i\ell}}{\partial x_k} \sigma_{k\ell} + \sigma_{i\ell} \frac{\partial \sigma_{k\ell}}{\partial x_k} \right). \quad (3.32)$$

On the other hand,

$$(\sigma \nabla \cdot \sigma^T)_i = \sum_{k=1}^d \sum_{\ell=1}^m \sigma_{i\ell} \frac{\partial \sigma_{k\ell}}{\partial x_k}. \quad (3.33)$$

The equivalence between (3.31) and (3.28) follows on subtracting (3.33) from (3.32). Observe also that we can write

$$\mathbf{h} \cdot \ell = \frac{1}{2} \sigma^T : \nabla(\sigma^T \ell), \quad (3.34)$$

for all vectors $\ell \in \mathbb{R}^d$.

Note that in order to be able to transform a Stratonovich SDE into an Itô SDE, we need to assume differentiability of the matrix σ , an assumption that is not necessary for the existence and uniqueness of solutions to an SDE; see Sect. 3.3.

3.3 Solutions of Stochastic Differential Equations

In this section we present, without proof, a basic existence and uniqueness result for SDEs of the form

$$dX_t = \mathbf{b}(t, X_t) dt + \sigma(t, X_t) dW_t, \quad X(0) = x, \quad (3.35)$$

where $\mathbf{b}(\cdot, \cdot) : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ and $\sigma(\cdot, \cdot) : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times m}$ are measurable vector-valued and matrix-valued functions, respectively, and W_t denotes standard Brownian motion in \mathbb{R}^n . We assume that the initial condition is a random variable that is independent of the Brownian motion W_t . We will denote by \mathcal{F}_t the filtration generated by the Brownian motion W_t .

We will use the following concept of a solution to (3.35).

Definition 3.2. A process X_t with continuous paths defined on the probability space (Ω, \mathcal{F}, P) is called a *strong solution* to the SDE (3.35) if:

- (i) X_t is almost surely continuous and adapted to the filtration \mathcal{F}_t .
- (ii) $\mathbf{b}(\cdot, X_t) \in L^1((0, T); \mathbb{R}^d)$ and $\sigma(\cdot, X_t) \in L^2((0, T); \mathbb{R}^{d \times n})$ almost surely.
- (iii) For every $t \geq 0$, the stochastic integral equation

$$X_t = x + \int_0^t \mathbf{b}(s, X_s) ds + \int_0^t \sigma(s, X_s) dW_s, \quad X(0) = x, \quad (3.36)$$

holds almost surely.

The assumptions that we have to impose on the drift and diffusion coefficients in (3.35) so that a unique strong solution exists are similar to the Lipschitz continuity and linear growth assumptions that are familiar from the existence and uniqueness theory of (deterministic) ordinary differential equations. In particular, we make the following two assumptions on the coefficients: there exists a positive constant C such that for all $x \in \mathbb{R}^d$ and $t \in [0, T]$,

$$|\mathbf{b}(t, x)| + |\sigma(t, x)|_F \leq C(1 + |x|), \quad (3.37)$$

and for all $x, y \in \mathbb{R}^d$ and $t \in [0, T]$,

$$|\mathbf{b}(t, x) - \mathbf{b}(t, y)| + |\sigma(t, x) - \sigma(t, y)|_F \leq C|x - y|. \quad (3.38)$$

Note that for globally Lipschitz vector and matrix fields \mathbf{b} and σ (i.e., when (3.38) is satisfied), the linear growth condition (3.37) is equivalent to the requirement that $|\mathbf{b}(t, 0)|$ and $|\sigma(t, 0)|_F$ be bounded for all $t \geq 0$.

Under these assumptions, a *global, unique* solution exists for the SDE (3.35). By uniqueness of strong solutions we mean that if X_t and Y_t are strong solutions to (3.35), then

$$X_t = Y_t \quad \text{for all } t \text{ almost surely.}$$

Theorem 3.1. Let $\mathbf{b}(\cdot, \cdot)$ and $\sigma(\cdot, \cdot)$ satisfy assumptions (3.37) and (3.38). Assume, furthermore, that the initial condition x is a random variable independent of the Brownian motion W_t with

$$\mathbb{E}|x|^2 < \infty.$$

Then the SDE (3.35) has a unique strong solution X_t with

$$\mathbb{E} \left[\int_0^t |X_s|^2 ds \right] < \infty \quad (3.39)$$

for all $t > 0$.

Using Gronwall's inequality, we can also obtain a quantitative estimate in (3.39) as a function of the second moment of the solution that increases exponentially in time.

The solution of the SDE (3.35) satisfies the Markov property and has continuous paths: it is a diffusion process. In fact, solutions of SDEs are precisely the diffusion processes that we studied in Chap. 2.

The Stratonovich analogue of (3.35) is

$$dX_t = \mathbf{b}(t, X_t) dt + \sigma(t, X_t) \circ dW_t, \quad X(0) = x. \quad (3.40)$$

As with the Itô SDE, the correct interpretation of (3.35) is in the sense of the integral equation

$$X_t = x + \int_0^t \mathbf{b}(s, X_s) ds + \int_0^t \sigma(s, X_s) \circ dW_s, \quad X(0) = x. \quad (3.41)$$

Using the Itô-to-Stratonovich transformation (3.28), we can write (3.40) as an Itô SDE and then use Theorem 3.1 to prove the existence and uniqueness of strong solutions. Notice, however, that for this, we need to assume differentiability of the diffusion matrix σ and that we need to check that conditions (3.37) and (3.38) are satisfied for the modified drift in (3.30).

Just as with nonautonomous ordinary differential equations, it is possible to rewrite an SDE with time-dependent coefficients as a time-homogeneous equation by adding one additional variable. Consider, for simplicity, the one-dimensional equation

$$dX_t = b(t, X_t) dt + \sigma(t, X_t) dW_t. \quad (3.42)$$

We introduce the auxiliary variable τ_t to write (3.42) as the system of SDEs

$$dX_t = b(\tau_t, X_t) dt + \sigma(\tau_t, X_t) dW_t, \quad (3.43a)$$

$$d\tau_t = dt. \quad (3.43b)$$

Thus we obtain a system of two homogeneous SDEs for the variables (X_t, τ_t) . Note that noise acts only in the equation for X_t . The generator of the diffusion process (X_t, τ_t) (see Sect. 3.4) is

$$\mathcal{L} = \frac{\partial}{\partial \tau} + b(\tau, x) \frac{\partial}{\partial x} + \frac{1}{2} \sigma^2(\tau, x) \frac{\partial^2}{\partial x^2}. \quad (3.44)$$

Quite often, we are interested in studying the long-time properties of the solution of an SDE. For this, it is useful to rescale time so that we can focus on long time scales. Using the scaling property of Brownian motion, see Proposition 1.5,

$$W(ct) = \sqrt{c}W(t),$$

we have that if $s = ct$, then

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

the equivalence being, of course, in law. Hence, if we scale time to $s = ct$, the SDE

$$dX_t = \mathbf{b}(X_t) dt + \sigma(X_t) dW_t$$

becomes

$$dX_t = \frac{1}{c} \mathbf{b}(X_t) dt + \frac{1}{\sqrt{c}} \sigma(X_t) dW_t,$$

together with the initial condition $X_0 = x$. We will use such a change of time scale when we study Brownian motors in Sect. 7.5.

3.4 Itô's Formula

In Chap. 2, we showed that to a diffusion process X_t we can associate a second-order differential operator, the generator of the process. Consider the Itô SDE

$$dX_t = \mathbf{b}(X_t) dt + \sigma(X_t) dW_t, \quad (3.45)$$

where for simplicity, we have assumed that the coefficients are independent of time. Here X_t is a diffusion process with drift $\mathbf{b}(x)$ and diffusion matrix

$$\Sigma(x) = \sigma(x)\sigma(x)^T. \quad (3.46)$$

The *generator* \mathcal{L} is then defined as

$$\mathcal{L} = \mathbf{b}(x) \cdot \nabla + \frac{1}{2} \Sigma(x) : D^2, \quad (3.47)$$

where D^2 denotes the Hessian matrix. The generator can also be written as

$$\mathcal{L} = \sum_{j=1}^d b_j(x) \frac{\partial}{\partial x_j} + \frac{1}{2} \sum_{i,j=1}^d \Sigma_{ij} \frac{\partial^2}{\partial x_i \partial x_j}.$$

We can now write Itô's formula using the generator. This formula enables us to calculate the rate of change in time of functions $V : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ evaluated at the solution of an \mathbb{R}^d -valued SDE. First, recall that in the absence of noise, the rate of change of V can be written as

$$\frac{d}{dt}V(t, x(t)) = \frac{\partial V}{\partial t}(t, x(t)) + \mathcal{A}V(t, x(t)), \quad (3.48)$$

where $x(t)$ is the solution of the ODE $\dot{x} = b(x)$, and \mathcal{A} denotes the (backward) *Liouville operator*

$$\mathcal{A} = b(x) \cdot \nabla. \quad (3.49)$$

Let now X_t be the solution of (3.45) and assume that the white noise \dot{W} is replaced by a smooth function $\zeta(t)$. Then the rate of change of $V(t, X_t)$ is given by the formula

$$\frac{d}{dt}V(t, X_t) = \frac{\partial V}{\partial t}(t, X_t) + \mathcal{A}V(t, X_t) + \langle \nabla V(t, X_t), \sigma(X_t) \zeta(t) \rangle. \quad (3.50)$$

This formula is no longer valid when (3.45) is driven by white noise and not a smooth function. In particular, the Leibniz formula (3.50) has to be modified by the addition of a drift term that accounts for the lack of smoothness of the noise: the Liouville operator \mathcal{A} given by (3.49) in (3.50) has to be replaced by the generator \mathcal{L} given by (3.47). We have already encountered this additional term in the previous chapter, when we derived the backward Kolmogorov equation. Formally, we can write

$$\frac{d}{dt}V(t, X_t) = \frac{\partial V}{\partial t}(t, X_t) + \mathcal{L}V(t, X_t) + \left\langle \nabla V(t, X_t), \sigma(X_t) \frac{dW}{dt} \right\rangle.$$

The precise interpretation of the expression for the rate of change of V is in integrated form.

Lemma 3.2 (Itô's formula). *Assume that the conditions of Theorem 3.1 hold. Let X_t be the solution of (3.45), and let $V \in C^{1,2}([0, T] \times \mathbb{R}^d)$. Then the process $V(X_t)$ satisfies*

$$\begin{aligned} V(t, X_t) &= V(X_0) + \int_0^t \frac{\partial V}{\partial s}(s, X_s) ds + \int_0^t \mathcal{L}V(s, X_s) ds \\ &\quad + \int_0^t \langle \nabla V(s, X_s), \sigma(X_s) dW_s \rangle. \end{aligned} \quad (3.51)$$

The presence of the additional term in the drift is not very surprising, in view of the fact that the Brownian differential scales like the square root of the differential in time: $\mathbb{E}(dW_t)^2 = dt$. In fact, we can write Itô's formula in the form

$$\frac{d}{dt}V(t, X_t) = \frac{\partial V}{\partial t}(t, X_t) + \langle \nabla V(t, X_t), \dot{X}_t \rangle + \frac{1}{2} \langle \dot{X}_t, D^2V(t, X_t) \dot{X}_t \rangle, \quad (3.52)$$

or

$$dV(t, X_t) = \frac{\partial V}{\partial t} dt + \sum_{i=1}^d \frac{\partial V}{\partial x_i} dX_i + \frac{1}{2} \sum_{i,j=1}^d \frac{\partial^2 V}{\partial x_i \partial x_j} dX_i dX_j, \quad (3.53)$$

where we have suppressed the argument (t, X_t) from the right-hand side. In writing (3.53), we have used the convention $dW_i(t) dW_j(t) = \delta_{ij} dt$, $dW_i(t) dt = 0$, $i, j = 1, \dots, d$. Thus, we can think of (3.53) as a generalization of Leibniz's rule (3.50), where second-order differentials are kept. We can check that with the above convention, Itô's formula follows from (3.53).

The proof of Itô's formula is essentially the same as the proof of the validity of the backward Kolmogorov equation that we presented in Chap. 2, Theorem 2.1. Conversely, after Itô's formula has been proved, the backward Kolmogorov and Fokker-Planck (forward Kolmogorov) equations follow as corollaries. Let $\phi \in C^2(\mathbb{R}^d)$ and denote by X_t^x the solution of (3.45) with $X_0^x = x$. Consider the function

$$u(x, t) = \mathbb{E}\phi(X_t^x) := \mathbb{E}(\phi(X_t^x) | X_0^x = x), \quad (3.54)$$

where the expectation is with respect to all Brownian driving paths. We apply Itô's formula to $\phi(X_t^x)$, use the martingale property of the stochastic integral (3.20), and differentiate with respect to time to obtain the backward Kolmogorov equation

$$\frac{\partial u}{\partial t} = \mathcal{L}u, \quad (3.55)$$

together with the initial condition $u(x, 0) = \phi(x)$. In this formal derivation, we need to assume that the expectation \mathbb{E} and the generator \mathcal{L} commute; this can be justified, since as we have already seen, the expectation of a functional of the solution to the SDE is given by the semigroup generated by \mathcal{L} : $(\mathcal{P}_t f)(x) = (e^{t\mathcal{L}} f)(x) = \mathbb{E}f(X_t^x)$.

The Feynman–Kac Formula

Itô's formula can be used to obtain a probabilistic description of solutions to more general partial differential equations of parabolic type. Let X_t^x be a diffusion process with drift $b(\cdot)$, diffusion $\Sigma(\cdot) = \sigma\sigma^T(\cdot)$, and generator \mathcal{L} with $X_0^x = x$, and let $f \in C_0^2(\mathbb{R}^d)$ and $V \in C(\mathbb{R}^d)$, bounded from below. Then the function

$$u(x, t) = \mathbb{E} \left(e^{-\int_0^t V(X_s^x) ds} f(X_t^x) \right) \quad (3.56)$$

is the solution to the initial value problem

$$\frac{\partial u}{\partial t} = \mathcal{L}u - Vu, \quad (3.57a)$$

$$u(0, x) = f(x). \quad (3.57b)$$

To derive this result, we introduce the variable $Y_t = \exp(-\int_0^t V(X_s) ds)$. We rewrite the SDE for X_t as

$$dX_t^x = b(X_t^x) dt + \sigma(X_t^x) dW_t, \quad X_0^x = x, \quad (3.58a)$$

$$dY_t^x = -V(X_t^x) dt, \quad Y_0^x = 0. \quad (3.58b)$$

The process $\{X_t^x, Y_t^x\}$ is a diffusion process with generator

$$\mathcal{L}_{x,y} = \mathcal{L} - V(x) \frac{\partial}{\partial y}.$$

We can write

$$\mathbb{E} \left(e^{-\int_0^t V(X_s^x) ds} f(X_t^x) \right) = \mathbb{E}(\phi(X_t^x, Y_t^x)),$$

where $\phi(x, y) = f(x)e^y$. We now apply Itô's formula to this function [or equivalently, write the backward Kolmogorov equation for the function $u(x, y, t) = \mathbb{E}(\phi(X_t^x, Y_t^x))$] to obtain (3.57).

The representation formula (3.56) for the solution of the initial value problem (3.57) is called the *Feynman–Kac formula*. It is very useful for the theoretical analysis of initial value problems for parabolic PDEs of the form (3.57) as well as for their numerical solution using a Monte Carlo approach, based on solving (3.58) numerically. See Sect. 5.2.

Derivation of the Fokker–Planck Equation

Starting from Itô's formula, we can obtain the backward Kolmogorov equation (3.55). Using now the backward Kolmogorov equation and the fact that the Fokker–Planck operator is the L^2 -adjoint of the generator, we can obtain the Fokker–Planck (forward Kolmogorov) equation, which we will study in detail in Chap. 4. This line of argument provides us with an alternative, and perhaps simpler, derivation of the forward and backward Kolmogorov equations to that presented in Sect. 2.5.

The Fokker–Planck operator corresponding to the SDE (3.45) reads

$$\mathcal{L}^* \cdot = \nabla \cdot \left(-\mathbf{b}(x) \cdot + \frac{1}{2} \nabla \cdot (\Sigma \cdot) \right). \quad (3.59)$$

We can derive this formula from the formula for the generator \mathcal{L} of X_t and two integrations by parts:

$$\int_{\mathbb{R}^d} \mathcal{L} f h dx = \int_{\mathbb{R}^d} f \mathcal{L}^* h dx,$$

for all $f, h \in C_0^2(\mathbb{R}^d)$.

Proposition 3.3. *Let X_t denote the solution of the Itô SDE (3.45) and assume that the initial condition X_0 is a random variable, independent of the Brownian motion driving the SDE, with density $\rho_0(x)$. Assume that the law of the Markov process X_t has a density $\rho(x, t) \in C^{2,1}(\mathbb{R}^d \times (0, +\infty))$.² Then ρ is the solution of the initial value problem for the Fokker–Planck equation*

$$\frac{\partial \rho}{\partial t} = \mathcal{L}^* \rho \quad \text{for } (x, t) \in \mathbb{R}^d \times (0, \infty), \quad (3.60a)$$

$$\rho = \rho_0 \quad \text{for } x \in \mathbb{R}^d \times \{0\}. \quad (3.60b)$$

Proof. Let \mathbb{E}^μ denote the expectation with respect to the product measure induced by the measure μ with density ρ_0 on X_0 and the Wiener measure of the Brownian motion that is driving the SDE. Averaging over the random initial conditions, distributed with density $\rho_0(x)$, and denoting by $\phi(x, t)$ the solution of the backward Kolmogorov equation we obtain

$$\begin{aligned} \mathbb{E}^\mu \phi(X_t) &= \int_{\mathbb{R}^d} \phi(x, t) \rho_0(x) dx \\ &= \int_{\mathbb{R}^d} (e^{\mathcal{L}t} \phi)(x) \rho_0(x) dx \\ &= \int_{\mathbb{R}^d} (e^{\mathcal{L}^*t} \rho_0)(x) \phi(x) dx. \end{aligned}$$

But since $\rho(x, t)$ is the density of X_t , we also have

$$\mathbb{E}^\mu \phi(X_t) = \int_{\mathbb{R}^d} \rho(t) \phi(x) dx.$$

Equating these two expressions for the expectation at time t , we obtain

$$\int_{\mathbb{R}^d} (e^{\mathcal{L}^*t} \rho_0)(x) \phi(x) dx = \int_{\mathbb{R}^d} \rho(x, t) \phi(x) dx.$$

We use a density argument that allows the identity to be extended to all $\phi \in L^2(\mathbb{R}^d)$. Hence, from the above equation, we deduce that

$$\rho(x, t) = (e^{\mathcal{L}^*t} \rho_0)(x).$$

Differentiation of the above equation gives (3.60a). Setting $t = 0$ gives the initial condition (3.60b).

² This is the case, for example, when the SDE has smooth coefficients and the diffusion matrix $\Sigma = \sigma \sigma^T$ is strictly positive definite.

The Chain Rule for Stratonovich Equations

For a Stratonovich SDE, the rules of standard calculus apply:

Proposition 3.4. *Let $X_t : \mathbb{R}^+ \mapsto \mathbb{R}^d$ be the solution of the Stratonovich SDE*

$$dX_t = \mathbf{b}(X_t) dt + \sigma(X_t) \circ dW_t, \quad (3.61)$$

where $\mathbf{b} : \mathbb{R}^d \mapsto \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \mapsto \mathbb{R}^{d \times d}$. Then the generator and Fokker–Planck operator of X_t are, respectively,

$$\mathcal{L} \cdot = \mathbf{b} \cdot \nabla \cdot + \frac{1}{2} \sigma^T : \nabla(\sigma^T \nabla \cdot) \quad (3.62)$$

and

$$\mathcal{L}^* \cdot = \nabla \cdot (-\mathbf{b} \cdot + \frac{1}{2} \sigma \nabla \cdot (\sigma^T \cdot)). \quad (3.63)$$

Furthermore, the Newton–Leibniz chain rule applies.

Proof. We will use the summation convention and also use the notation ∂_j for the partial derivative with respect to x_j . We use the formula for the Itô-to-Stratonovich correction, Eq. (3.31), to write

$$\begin{aligned} \mathcal{L} &= b_j \partial_j + \frac{1}{2} \sigma_{jk} \partial_j \sigma_{ik} \partial_i f + \frac{1}{2} \sigma_{ik} \sigma_{jk} \partial_i \partial_j f \\ &= b_j \partial_j f_j + \frac{1}{2} \sigma_{jk} \partial_j (\sigma_{ik} \partial_i f) \\ &= b \cdot \nabla + \frac{1}{2} \sigma^T : \nabla(\sigma^T \nabla f). \end{aligned}$$

To obtain the formula for the Fokker–Planck operator, let f, h be two $C_0^2(\mathbb{R}^d)$ functions. We perform two integrations by parts to obtain

$$\begin{aligned} \int_{\mathbb{R}^d} \sigma_{jk} \partial_j (\sigma_{ik} \partial_i f) h dx &= \int_{\mathbb{R}^d} f \partial_i (\sigma_{ik} \partial_j (\sigma_{jk} h)) dx \\ &= \int_{\mathbb{R}^d} f \partial_i (\sigma_{ik} (\nabla \cdot (\sigma^T h))_k) dx \\ &= \int_{\mathbb{R}^d} f \partial_i (\sigma \cdot \nabla \cdot (\sigma^T h))_i dx \\ &= \int_{\mathbb{R}^d} f \nabla \cdot (\sigma \cdot \nabla \cdot (\sigma^T h)) dx. \end{aligned}$$

To show that the Stratonovich SDE satisfies the standard chain rule, let $\mathbf{h} : \mathbb{R}^d \mapsto \mathbb{R}^d$ be an invertible map, let $y = \mathbf{h}(x)$, and let $x = \mathbf{h}^{-1}(y) =: \mathbf{g}(y)$. Let $\mathbf{J} = \nabla_x \mathbf{h}$ denote the Jacobian of the transformation. We introduce the notation

$$\widehat{\mathbf{b}}(y) = \mathbf{b}(\mathbf{g}(y)), \quad \widehat{\sigma}(y) = \sigma(\mathbf{g}(y)), \quad \widehat{\mathbf{J}}(y) = \mathbf{J}(\mathbf{g}(y)). \quad (3.64)$$

We need to show that for $Y_t = \mathbf{h}(X_t)$,

$$dY_t = \widehat{\mathbf{J}}(Y_t) \left(\widehat{\mathbf{b}}(Y_t) dt + \widehat{\sigma}(Y_t) \circ dW_t \right). \quad (3.65)$$

It is important to note that in this equation, the (square root of the) diffusion matrix is $\widehat{\mathbf{J}}(y)\widehat{\sigma}(y)$. In particular, the Stratonovich correction in (3.65) is [see (3.28)]

$$b_\ell^S(y) = \frac{1}{2} (\widehat{J}_{j\rho} \widehat{\sigma}_{\rho k})(y) \frac{\partial (\widehat{J}_{\ell m} \widehat{\sigma}_{mk})}{\partial y_j}(y), \quad \ell = 1, \dots, d, \quad (3.66)$$

or, using (3.31),

$$\mathbf{b}^S(y) = \frac{1}{2} [\nabla \cdot (\widehat{\mathbf{J}} \widehat{\Sigma} \widehat{\mathbf{J}}^T) - \widehat{\mathbf{J}} \widehat{\sigma} \nabla \cdot (\widehat{\sigma}^T \widehat{\mathbf{J}}^T)](y).$$

To prove this, we first transform the Stratonovich SDE for X_t into an Itô SDE using (3.31). We then apply Itô's formula to calculate

$$\begin{aligned} dY_\ell &= (\mathcal{L}h_\ell)(X_t) dt + (\partial_i h_\ell \sigma_{ij})(X_t) dW_j \\ &= J_{\ell i}(X_t) (b_i(X_t) dt + \sigma_{ij}(X_t) dW_j) \\ &\quad + \frac{1}{2} \sigma_{jk}(X_t) \partial_j (\sigma_{ik}(X_t) J_{\ell i}(X_t)) dt, \end{aligned} \quad (3.67)$$

for $\ell = 1, \dots, d$. Equivalently, using (3.62), we have

$$\begin{aligned} dY_t &= \nabla_x \mathbf{h}(X_t) (\mathbf{b}(X_t) dt + \sigma(X_t) dW_t) \\ &\quad + \frac{1}{2} \sigma^T(X_t) : \nabla (\sigma^T(X_t) \nabla_x \mathbf{h}(X_t)) dt. \end{aligned} \quad (3.68)$$

Now we need to rewrite the right-hand side of this equation as a function of y . This follows essentially from the inverse function theorem, and in particular the fact that the Jacobian of the inverse transformation (from y to x) is given by the inverse of the Jacobian J of the transformation $y = \mathbf{h}(x)$:

$$\widehat{J}(y) = (\nabla_y \mathbf{g})^{-1}.$$

In particular,

$$\nabla_x f(x) = \widehat{J}(y) \nabla_y \widehat{f}(y). \quad (3.69)$$

For the first term, and using the notation (3.64), we have

$$\nabla_x \mathbf{h}(X_t) (\mathbf{b}(X_t) dt + \sigma(X_t) dW_t) = \widehat{\mathbf{J}}(Y_t) \left(\widehat{\mathbf{b}}(Y_t) dt + \widehat{\sigma}(Y_t) dW_t \right).$$

Now we need to show that the second term on the right-hand side of (3.67) is equal to the Stratonovich correction $\mathbf{b}^S(y)$ in (3.65), which is given by (3.66). This follows by applying the chain rule (3.69) to the right-hand side of (3.68) and using (3.34). Equivalently, using (3.67), we obtain

$$\begin{aligned}
dY_\ell &= \widehat{J}_{\ell i}(Y_t) \left(\widehat{b}_i(Y_t) dt + \widehat{\sigma}_{ij}(Y_t) dW_j \right) \\
&\quad + \frac{1}{2} \widehat{J}_{jp} \widehat{\sigma}_{pk}(Y_t) \partial_j \left(\widehat{\sigma}_{ik}(Y_t) \widehat{J}_{\ell i}(X_t) \right) dt \\
&= \widehat{J}_{\ell i}(Y_t) \left(\widehat{b}_i(Y_t) dt + \widehat{\sigma}_{ij}(Y_t) \circ dW_j \right).
\end{aligned}$$

3.5 Examples of SDEs

Brownian Motion

We consider an SDE with no drift and constant diffusion coefficient:

$$dX_t = \sqrt{2\sigma} dW_t, \quad X_0 = x, \quad (3.70)$$

where the initial condition can be either deterministic or random, independent of the Brownian motion W_t . The solution is

$$X_t = x + \sqrt{2\sigma} W_t.$$

This is just Brownian motion starting at x with diffusion coefficient σ .

Ornstein–Uhlenbeck Process

Adding a restoring force to (3.70), we obtain the SDE for the Ornstein–Uhlenbeck process that we have already encountered:

$$dX_t = -\alpha X_t dt + \sqrt{2\sigma} dW_t, \quad X(0) = x, \quad (3.71)$$

with $\alpha > 0$ and where, as in the previous example, the initial condition can be either deterministic or random, independent of the Brownian motion W_t . We solve this equation using the variation of constants formula:

$$X_t = e^{-\alpha t} x + \sqrt{2\sigma} \int_0^t e^{-\alpha(t-s)} dW_s. \quad (3.72)$$

We can use Itô's formula to obtain equations for the moments of the Ornstein–Uhlenbeck process.³ The generator is

$$\mathcal{L} = -\alpha x \frac{d}{dx} + \sigma \frac{d^2}{dx^2}.$$

³ In Sect. 4.2, we will redo this calculation using the Fokker–Planck equation.

We apply Itô's formula to the function $f(x) = x^n$ to obtain

$$\begin{aligned} dX_t^n &= \mathcal{L}X_t^n dt + \sqrt{2\sigma} \frac{d}{dx} X_t^n dW \\ &= -\alpha n X_t^n dt + \sigma n(n-1) X_t^{n-2} dt + n\sqrt{2\sigma} X_t^{n-1} dW. \end{aligned}$$

Consequently,

$$X_t^n = x^n + \int_0^t (-\alpha n X_s^n + \sigma n(n-1) X_s^{n-2}) ds + n\sqrt{2\sigma} \int_0^t X_s^{n-1} dW_s.$$

By taking the expectation in the above equation and using the fact that the stochastic integral is a martingale, in particular (3.19), we obtain an equation for the moments $M_n(t) = \mathbb{E}X_t^n$ of the Ornstein–Uhlenbeck process for $n \geq 2$:

$$M_n(t) = M_n(0) + \int_0^t (-\alpha n M_n(s) + \sigma n(n-1) M_{n-2}(s)) ds,$$

where we have considered random initial conditions distributed according to a distribution $\rho_0(x)$ with finite moments.

A variant of the SDE (3.71) is the *mean reverting Ornstein–Uhlenbeck process*

$$dX_t = (\mu - \alpha X_t) dt + \sqrt{2\sigma} dW_t, \quad X(0) = x. \quad (3.73)$$

We can use the variation of constants formula to solve this SDE; see Exercise 6.

Geometric Brownian Motion

(See also Sect. 4.2.) Consider the following scalar linear SDE with multiplicative noise:

$$dX_t = \mu X_t dt + \sigma X_t dW_t, \quad X_0 = x, \quad (3.74)$$

where we use the Itô interpretation of the stochastic differential. The solution to this equation is known as *geometric Brownian motion*. We can think of it as a very simple model of population dynamics in a fluctuating environment: we can obtain (3.74) from the exponential growth (or decay) model

$$\frac{dX_t}{dt} = \mu(t) X_t. \quad (3.75)$$

We assume that there are fluctuations (or uncertainty) in the growth rate $\mu(t)$. Modeling the uncertainty as Gaussian white noise, we can write

$$\mu(t) = \mu + \sigma \xi(t), \quad (3.76)$$

with $\xi(t)$ denoting the white noise process $\frac{dW_t}{dt}$. Substituting now (3.76) into (3.75), we obtain the SDE for geometric Brownian motion. The multiplicative noise in (3.74) is due to our lack of complete knowledge of the growth parameter $\mu(t)$. This is a general feature in the modeling of physical and biological systems using SDEs: multiplicative noise is quite often associated with fluctuations or uncertainty in the parameters of the model.

The generator of the geometric Brownian motion is

$$\mathcal{L} = \mu x \frac{d}{dx} + \frac{\sigma^2 x^2}{2} \frac{d^2}{dx^2}.$$

The solution of (3.74) is

$$X(t) = x \exp \left(\left(\mu - \frac{\sigma^2}{2} \right) t + \sigma W(t) \right). \quad (3.77)$$

To derive this formula, we apply Itô's formula to the function $f(x) = \log(x)$:

$$\begin{aligned} d \log(X_t) &= \mathcal{L} \left(\log(X_t) \right) dt + \sigma x \frac{d}{dx} \log(X_t) dW_t \\ &= \left(\mu X_t \frac{1}{X_t} + \frac{\sigma^2 X_t^2}{2} \left(-\frac{1}{X_t^2} \right) \right) dt + \sigma dW_t \\ &= \left(\mu - \frac{\sigma^2}{2} \right) dt + \sigma dW_t. \end{aligned}$$

Consequently

$$\log \left(\frac{X_t}{X_0} \right) = \left(\mu - \frac{\sigma^2}{2} \right) t + \sigma W(t),$$

from which (3.77) follows.

Observe that if we interpret the stochastic differential in (3.74) in the Stratonovich sense, then the solution is no longer given by (3.77). To see this, first note that from (3.28), it follows that the Stratonovich SDE

$$dX_t = \mu X_t dt + \sigma X_t \circ dW_t, \quad X_0 = x, \quad (3.78)$$

is equivalent to the Itô SDE

$$dX_t = \left(\mu + \frac{1}{2} \sigma^2 \right) X_t dt + \sigma X_t dW_t, \quad X_0 = x. \quad (3.79)$$

Consequently, from (3.77) and replacing μ with $\mu + \frac{1}{2} \sigma^2$, we conclude that the solution of (3.78) is

$$X(t) = x \exp(\mu t + \sigma W(t)). \quad (3.80)$$

Comparing (3.77) with (3.80), we immediately see that the Itô and Stratonovich interpretations of the stochastic integral lead to SDEs with different properties. For

example, from (3.77), we observe that the noise in (3.74) can change the qualitative behavior of the solution: for $\mu > 0$, the solution to the deterministic equation ($\sigma = 0$) increases exponentially as $t \rightarrow +\infty$, whereas for the solution of the stochastic equation, it can be shown that it converges to 0 with probability one, provided that $\mu - \frac{\sigma^2}{2} < 0$.

We now present a list of SDEs that appear in applications.

- The Cox–Ingersoll–Ross equation:

$$dX_t = \alpha(b - X_t) dt + \sigma \sqrt{X_t} dW_t, \quad (3.81)$$

where α, b, σ are positive constants.

- The stochastic Verhulst equation (population dynamics):

$$dX_t = (\lambda X_t - X_t^2) dt + \sigma X_t dW_t. \quad (3.82)$$

- Coupled Lotka–Volterra stochastic equations:

$$dX_i(t) = X_i(t) \left(a_i + \sum_{j=1}^d b_{ij} X_j(t) \right) dt + \sigma_i X_i(t) dW_i(t), \quad i = 1, \dots, d. \quad (3.83)$$

- Protein kinetics:

$$dX_t = (\alpha - X_t + \lambda X_t(1 - X_t)) dt + \sigma X_t(1 - X_t) \circ dW_t. \quad (3.84)$$

- Dynamics of a tracer particle (turbulent diffusion):

$$dX_t = u(X_t, t) dt + \sigma dW_t, \quad \nabla \cdot u(x, t) = 0. \quad (3.85)$$

- Josephson junction (pendulum with friction and noise):

$$\ddot{\phi}_t = -\sin(\phi_t) - \gamma \dot{\phi}_t + \sqrt{2\gamma\beta^{-1}} \dot{W}_t. \quad (3.86)$$

- Noisy Duffing oscillator (stochastic resonance)

$$\ddot{X}_t = -\beta X_t - \alpha X_t^3 - \gamma \dot{X}_t + A \cos(\omega t) + \sigma \dot{W}_t. \quad (3.87)$$

- The stochastic heat equation, which we can write formally as

$$\partial_t u = \partial_x^2 u + \partial_t W(x, t), \quad (3.88)$$

where $W(x, t)$ denotes an infinite-dimensional Brownian motion. We can consider (3.88) on $[0, 1]$ with Dirichlet boundary conditions. We can represent $W(x, t)$ as a Fourier series:

$$W(x, t) = \sum_{k=1}^{+\infty} e_k(x) W_k(t), \quad (3.89)$$

where $W_k(t)$, $k = 1, \dots, +\infty$ are one-dimensional independent Brownian motions and $\{e_k(x)\}_{k=1}^{+\infty}$ is the standard orthonormal basis in $L^2(0, 1)$ with Dirichlet boundary conditions, i.e., $e_k(x) = \frac{1}{\sqrt{2}} \sin(2\pi kx)$, $k = 1, \dots, +\infty$.

3.6 The Lamperti Transformation and Girsanov's Theorem

Stochastic differential equations with a nontrivial drift and multiplicative noise are hard to analyze, in particular in dimensions greater than one. In this section, we present two techniques that enable us to map equations with multiplicative noise to equations with additive noise, and to map equations with a nonconstant drift to equations with no drift and multiplicative noise. The first technique, the Lamperti transformation, works mostly in one dimension, whereas the second technique, Girsanov's theorem, is applicable in arbitrary, even infinite, dimensions.

The Lamperti Transformation

For SDEs in one dimension, it is possible to map multiplicative noise to additive noise through a generalization of the method that we used to obtain the solution of the equation for geometric Brownian motion. Consider a one-dimensional Itô SDE with multiplicative noise

$$dX_t = b(X_t) dt + \sigma(X_t) dW_t. \quad (3.90)$$

We ask whether there exists a transformation $z = h(x)$ that maps (3.90) into an SDE with additive noise. We apply Itô's formula to obtain

$$dZ_t = \mathcal{L}h(X_t) dt + h'(X_t) \sigma(X_t) dW_t,$$

where \mathcal{L} denotes the generator of X_t . In order to obtain an SDE with unit diffusion coefficient, we need to impose the condition

$$h'(x) \sigma(x) = 1,$$

from which we deduce that

$$h(x) = \int_{x_0}^x \frac{1}{\sigma(x)} dx, \quad (3.91)$$

where x_0 is arbitrary. We have

$$\mathcal{L}h(x) = \frac{b(x)}{\sigma(x)} - \frac{1}{2} \sigma'(x).$$

Consequently, the transformed SDE has the form

$$dY_t = b_Y(Y_t) dt + dW_t \quad (3.92)$$

with

$$b_Y(y) = \frac{b(h^{-1}(y))}{\sigma(h^{-1}(y))} - \frac{1}{2}\sigma'(h^{-1}(y)).$$

This is called the *Lamperti transformation*. As an application of this transformation, consider the Cox–Ingersoll–Ross equation (3.81)

$$dX_t = (\mu - \alpha X_t) dt + \sigma \sqrt{X_t} dW_t, \quad X_0 = x > 0.$$

From (3.91), we deduce that

$$h(x) = \frac{2}{\sigma} \sqrt{x}.$$

The generator of this process is

$$\mathcal{L} = (\mu - \alpha x) \frac{d}{dx} + \frac{\sigma^2}{2} x \frac{d^2}{dx^2}.$$

We have that

$$\mathcal{L}h(x) = \left(\frac{\mu}{\sigma} - \frac{\sigma}{4} \right) x^{-1/2} - \frac{\alpha}{\sigma} x^{1/2}.$$

The CIR equation becomes, for $Y_t = \frac{2}{\sigma} \sqrt{X_t}$,

$$\begin{aligned} dY_t &= \left(\frac{\mu}{\sigma} - \frac{\sigma}{4} \right) \frac{1}{\sqrt{X_t}} dt - \frac{\alpha}{\sigma} \sqrt{X_t} dt + dW_t \\ &= \left[\left(\frac{2\mu}{\sigma^2} - \frac{1}{2} \right) \frac{1}{Y_t} - \frac{\alpha}{2} Y_t \right] dt + dW_t. \end{aligned}$$

When $\mu = \frac{\sigma^2}{4}$, the equation above becomes the Ornstein–Uhlenbeck process for Y_t .

Apart from being a useful tool for obtaining solutions to one-dimensional SDEs with multiplicative noise, it is also often used in statistical inference for SDEs. See Sect. 5.3 and the discussion in Sect. 5.5. Such a transformation does not exist for arbitrary SDEs in higher dimensions. In particular, it is not possible, in general, to transform a multidimensional Itô SDE with multiplicative noise to an SDE with additive noise; see Exercise 7.

Girsanov's Theorem

Conversely, it is also sometimes possible to remove the drift term from an SDE and obtain an equation in which only (multiplicative) noise is present. Consider first the following one-dimensional SDE:

$$dX_t = b(X_t) dt + dW_t. \quad (3.93)$$

We introduce the following functional of X_t :

$$M_t = \exp \left(-\frac{1}{2} \int_0^t b^2(X_s) ds - \int_0^t b(X_s) dW_s \right). \quad (3.94)$$

We can write $M_t = e^{-Y_t}$, where Y_t is the solution to the SDE

$$M_t = e^{-Y_t}, \quad dY_t = \frac{1}{2} b^2(X_t) dt + b(X_t) dW_t, \quad Y_0 = 0. \quad (3.95)$$

We can now apply Itô's formula to obtain the equation

$$dM_t = -M_t b(X_t) dW_t. \quad (3.96)$$

Notice that this is an SDE without drift.

In fact, under appropriate conditions on the drift $b(\cdot)$, it is possible to show that the law of the process X_t , denoted by \mathbb{P} , which is a probability measure over the space of continuous functions, is absolutely continuous with respect to the Wiener measure \mathbb{P}_W , the law of the Brownian motion W_t . The Radon–Nikodym derivative between these two measures is the inverse of the stochastic process M_t given in (3.95):

$$\frac{d\mathbb{P}}{d\mathbb{P}_W}(X_t) = \exp \left(\frac{1}{2} \int_0^t b^2(X_s) ds + \int_0^t b(X_s) dW_s \right). \quad (3.97)$$

This is a form of *Girsanov's theorem*. Using now Eq. (3.93), we can rewrite (3.97) as

$$\frac{d\mathbb{P}}{d\mathbb{P}_W}(X_t) = \exp \left(\int_0^t b(X_s) dX_s - \frac{1}{2} \int_0^t |b(X_s)|^2 ds \right). \quad (3.98)$$

The Girsanov transformation (3.96) or (3.98) enables us to “compare” the process X_t with the Brownian motion W_t . This is a very useful result when the drift function $b(\cdot)$ in (3.93) is known up to parameters that we want to estimate from observations. In this context, the Radon–Nikodym derivative in (3.98) becomes the *likelihood function*. We will study the problem of maximum likelihood parameter estimation for SDEs in Sect. 5.3.

3.7 Linear Stochastic Differential Equations

In this section, we study linear SDEs in arbitrary finite dimensions. Let $A, \Sigma \in \mathbb{R}^{d \times d}$ be positive definite and positive semidefinite matrices, respectively, and let $W(t)$ be a standard Brownian motion in \mathbb{R}^d . We will consider the SDE⁴

⁴ We can also consider the case that $\sigma \in \mathbb{R}^{d \times m}$ with $n \neq m$, i.e., that the SDE is driven by an m -dimensional Brownian motion.

$$dX(t) = -AX(t)dt + \sigma dW(t), \quad (3.99)$$

or, componentwise,

$$dX_i(t) = -\sum_{j=1}^d A_{ij}X_j(t) + \sum_{j=1}^d \sigma_{ij}dW_j(t), \quad i = 1, \dots, d.$$

The initial conditions $X(0) = x$ can be taken to be either deterministic or random. The generator of the Markov process $X(t)$ is

$$\mathcal{L} = -Ax \cdot \nabla + \frac{1}{2}\Sigma : D^2,$$

where $\Sigma = \sigma\sigma^T$. The corresponding Fokker–Planck equation is

$$\frac{\partial p}{\partial t} = \nabla \cdot (Axp) + \frac{1}{2}\nabla \cdot (\Sigma \nabla p). \quad (3.100)$$

The solution of (3.99) is

$$X(t) = e^{-At}x + \int_0^t e^{-A(t-s)}\sigma dW(s). \quad (3.101)$$

We use the fact that the stochastic integral is a martingale to calculate the expectation of the process $X(t)$:

$$\mu(t) := \mathbb{E}X(t) = e^{-At}\mathbb{E}x.$$

To simplify the formulas below, we will set $\mu(t) = 0$. This is with no loss of generality, since we can define the new process $Y(t) = X(t) - \mu(t)$, which is of mean zero.

We think of $X(t) \in \mathbb{R}^{d \times 1}$ as a column vector. Consequently, $X^T(t) \in \mathbb{R}^{1 \times d}$ is a row vector. The autocorrelation matrix is

$$R(t, s) = \mathbb{E}(X(t)X^T(s)) = \mathbb{E}(X(t) \otimes X(s)).$$

Componentwise,

$$R_{ij}(t, s) = \mathbb{E}(X_i(t)X_j(s)).$$

We will denote the covariance matrix of the initial conditions x by R_0 , $\mathbb{E}xx^T =: R_0$.

Proposition 3.5. *The autocorrelation matrix of the process $X(t)$ is given by*

$$R(t, s) = e^{-At} \left(R_0 + \int_0^{\min(t, s)} e^{A\rho} \Sigma e^{A^T \rho} d\rho \right) e^{-A^T s}. \quad (3.102)$$

Furthermore, the variance at time t , $\Sigma(t) := R(t, t)$, satisfies the differential equation

$$\frac{d\Sigma(t)}{dt} = -A\Sigma(t) - \Sigma(t)A^T + \Sigma. \quad (3.103)$$

The steady-state variance Σ_∞ is the solution of the equation

$$A\Sigma_\infty + \Sigma_\infty A^T = \Sigma. \quad (3.104)$$

The solution to this equation is

$$\Sigma_\infty = \int_0^{+\infty} e^{A\rho} \Sigma e^{A^T \rho} d\rho. \quad (3.105)$$

The invariant distribution of (3.99) is Gaussian with mean 0 and variance Σ_∞ .

Proof. We will use the notation $S_t = e^{-At}$ and $B_t = \Sigma W(t)$. The solution (3.101) can be written in the form

$$X_t = S_t x + \int_0^t S_{t-s} dB_s.$$

Consequently, using the properties of the stochastic integral, we have

$$\begin{aligned} R(t, s) &= \mathbb{E}\left((S_t x)(S_s x)^T\right) + \mathbb{E}\int_0^t \int_0^s (S_{t-\ell} dB_\ell) (S_{s-\rho} dB_\rho)^T \\ &=: S_t \mathbb{E}(xx^T) S_s^T + \int_0^t \int_0^s S_{t-\ell} \Sigma \mathbb{E}(dW_\ell dW_\rho^T) \Sigma^T S_{s-\rho}^T \\ &= S_t R_0 S_s + \int_0^t \int_0^s S_{t-\ell} \Sigma \delta(\ell - \rho) \Sigma^T S_{s-\rho}^T d\ell d\rho \\ &= S_t R_0 S_s + \int_0^{\min(t, s)} S_{t-\rho} \Sigma S_{s-\rho}^T d\rho \\ &= e^{-At} \left(R_0 + \int_0^{\min(t, s)} e^{A\rho} \Sigma e^{A^T \rho} d\rho \right) e^{-A^T s}. \end{aligned}$$

From (3.102), it follows that

$$\Sigma(t) = e^{-At} \left(R(0) + \int_0^t e^{A\rho} \Sigma(t) e^{A^T \rho} d\rho \right) e^{-A^T t}. \quad (3.106)$$

On differentiating this equation, we obtain the equation for the variance:

$$\frac{d\Sigma(t)}{dt} = -A\Sigma(t) - \Sigma(t)A^T + \Sigma(t),$$

with $\Sigma(0) = R(0)$. We now set the left-hand side of the above equation to 0 to obtain the equation for the steady-state variance Σ_∞ :

$$A\Sigma_\infty + \Sigma_\infty A^T = \Sigma.$$

Equation (3.104) is an example of a *Lyapunov equation*. When Σ is strictly positive definite, it is possible to show that (3.105) is a well-defined unique solution of (3.104); see Exercise 3. The situation becomes more complicated when Σ is positive semidefinite. See the discussion in Sect. 3.8.

We can use Proposition 3.5 to solve the Fokker–Planck equation (3.100) with initial conditions

$$p(\mathbf{x}, t | \mathbf{x}_0, 0) = \delta(\mathbf{x} - \mathbf{x}_0). \quad (3.107)$$

The solution of (3.99) with $X(0) = x_0$ deterministic is a Gaussian process with mean $\mu(t) = e^{-At}$ and variance Σ_t given by (3.106). Consequently, the solution of the Fokker–Planck equation with initial conditions (3.107) is

$$p(x, t | x_0, 0) = \frac{1}{(2\pi)^{d/2} \sqrt{\det(\Sigma(t))}} \exp\left(-\frac{1}{2} (x - e^{-At} x_0)^T \Sigma^{-1}(t) (x - e^{-At} x_0)\right). \quad (3.108)$$

This result can also be obtained using the Fourier transform. See Exercise 5.

3.8 Discussion and Bibliography

Stochastic differential equations and stochastic calculus are treated in many textbooks. See, for example, [7, 54, 64, 65, 98, 154, 178] as well as [111, 200, 211, 212]. The reader is strongly encouraged to study carefully the proofs of the construction of the Itô integral, Itô’s formula, and the basic existence and uniqueness theorem for SDEs from the above references.

In this book, we consider stochastic equations in finite dimensions. There is also a very well developed theory of stochastic partial differential equations; see [190]. Proposition 3.1 is taken from [178, Exercise 3.10]. Theorem 3.1 is taken from [178], where a proof can be found; see also [119, Chap. 21] and [191, Theorem 3.1.1]. The assumption that the drift and diffusion coefficients are globally Lipschitz can be weakened when a priori bounds on the solution can be found. This is the case, for example, when a *Lyapunov function* can be constructed. See, e.g., [154]. Several examples of stochastic equations that can be solved analytically can be found in [67].

The derivation of Itô’s formula is similar to the derivation of the backward Kolmogorov equation that was presented in Chap. 2, in particular Theorem 2.1. The proof can be found in all of the books on SDEs mentioned above. The Feynman–Kac formula, which is a generalization of Itô’s formula, was first derived in the context of quantum mechanics, providing a path integral solution to the Schrödinger equation. Path integrals and functional integration are studied in detail in [72, 219].

The concept of the solution used in Theorem 3.1 is that of a *strong solution*. It is also possible to define weak solutions of SDEs, in which case the Brownian motion that is driving the SDE (3.35) is not specified a priori. Roughly speaking, the concept of a weak solution to a stochastic equation is related to the solution of the corresponding Fokker–Planck equation, i.e., the law of the process X_t . One could argue that in some applications, such as physics and chemistry, the concept of a weak solution is more natural than that of a strong solution, since in those applications, one is usually interested in probability distribution functions rather than the actual paths of a diffusion process.

It is not very difficult to construct examples of SDEs that have weak but not strong solutions. A standard example is that of the *Tanaka equation*

$$dX_t = \operatorname{sgn}(X_t) dW_t, \quad (3.109)$$

where sgn denotes the sign function. One can show that this equation has no strong solution (which is not a big surprise, since the assumptions of Theorem 3.1 are not satisfied), but that it does have a unique weak solution. See [178] for the details. The fact that (3.109) is solvable in the weak sense follows from the fact that the Fokker–Planck equation for (3.109) is simply the heat equation, i.e., the Fokker–Planck equation for Brownian motion. Hence, every Brownian motion is a weak solution of the Tanaka equation.

Itô's formula also holds for a particular class of random times, the Markov or *stopping times*. Roughly speaking, a stopping time is a random time τ for which we can decide whether it has already occurred based on the information that is available to us, i.e., the solution of our SDE up to the present time.⁵ Let X_t be a diffusion process with generator \mathcal{L} , starting at x . We will denote the expectation with respect to the law of this process by \mathbb{E}^x . Furthermore, let $f \in C_0^2(\mathbb{R}^d)$ and let τ be a stopping time with $\mathbb{E}^x \tau < +\infty$. *Dynkin's formula* reads

$$\mathbb{E}^x f(X_\tau) = f(x) + \mathbb{E}^x \left[\int_0^\tau \mathcal{L} f(X_s) ds \right]. \quad (3.110)$$

The derivation of this formula and generalizations can be found in [178, Chaps. 7, 8]. Dynkin's formula and the Feynman–Kac formula are very useful for deriving partial differential equations for interesting functionals of the diffusion process X_t . Details, in particular for diffusion processes in one dimension, can be found in [113].

It is not possible, in general, to transform an SDE with multiplicative noise to one with additive noise in dimensions greater than one. In other words, the Lamperti transformation exists, unless additional assumptions on the diffusion matrix are imposed, only in one dimension. Extensions of this transformation to higher dimensions are discussed in [2]. See also Exercise 7.

Girsanov's theorem is one of the fundamental results in stochastic analysis and is presented in all the standard textbooks, e.g., [100, 111, 200]. A very detailed discussion of Girsanov's theorem and its connection with the construction of the likelihood function for diffusion processes can be found in [142, Chap. 7]. A form of Girsanov's theorem that is very useful in statistical inference for diffusion processes is the following [133, Sect. 1.1.4], [97, Sect. 1.12]: consider the two equations

⁵ More precisely, let $\{\mathcal{F}_t\}$ be a filtration. A function $\tau : \Omega \mapsto [0, +\infty]$ is called a (strict) stopping time with respect to $\{\mathcal{F}_t\}$ if

$$\{\omega; \tau(\omega) \leq t\} \in \mathcal{F}_t \quad \text{for all } t \geq 0.$$

$$dX_t = b_1(X_t) dt + \sigma(X_t) dW_t, \quad X_0 = x^1, \quad t \in [0, T], \quad (3.111a)$$

$$dX_t = b_2(X_t) dt + \sigma(X_t) dW_t, \quad X_0 = x^2, \quad t \in [0, T], \quad (3.111b)$$

where $\sigma(x) > 0$. We assume that we have existence and uniqueness of strong solutions for both SDEs. Assume that x^1 and x^2 are random variables with densities $f_1(x)$ and $f_2(x)$ with respect to the Lebesgue measure that have the same support, or are nonrandom and equal to the same constant. Let P_1 and P_2 denote the laws of these two SDEs. Then these two measures are equivalent⁶ and their Radon–Nikodym derivative is

$$\frac{dP_2}{dP_1}(X) = \frac{f_2(X_0)}{f_1(X_0)} \exp \left(\int_0^T \frac{b_2(X_t) - b_1(X_t)}{\sigma^2(X_t)} dX_t - \frac{1}{2} \int_0^T \frac{b_2^2(X_t) - b_1^2(X_t)}{\sigma^2(X_t)} dt \right). \quad (3.112)$$

Linear SDEs are studied in most textbooks on SDEs and stochastic processes; see, e.g., [7, Chap. 8], [206, Chap. 6].

3.9 Exercises

1. Calculate all moments of the geometric Brownian motion (3.74) for the Itô and Stratonovich interpretations of the noise in the equation.
2. Prove rigorously (3.24) by keeping careful track of the error terms.
3. Consider Eq. (3.104) and assume that all the eigenvalues of the matrix A have positive real parts. Assume furthermore that Σ is symmetric and positive definite. Show that there exists a unique positive solution to the Lyapunov equation (3.104).
4. Obtain (3.103) using Itô's formula.
5. Solve the Fokker–Planck equation (3.100) with initial conditions (3.107) (*Hint:* Take the Fourier transform and use the fact that the Fourier transform of a Gaussian function is Gaussian.) Assume that the matrices A and Σ commute. Calculate the stationary autocorrelation matrix using the formula

$$\mathbb{E}(X_0^T X_t) = \int \int x_0^T x p(x, t | x_0, 0) p_s(x_0) dx dx_0$$

and Gaussian integration.

6. Consider the mean reverting Ornstein–Uhlenbeck process

$$dX_t = (\mu - \alpha X_t) dt + \sqrt{2\lambda} dW_t, \quad X(0) = x. \quad (3.113)$$

⁶ Two probability measures P_1, P_2 are equivalent on a σ -field \mathcal{G} if and only if $P_1(A) = 0 \iff P_2(A) = 0$ for all $A \in \mathcal{G}$. In this case $\frac{dP_2}{dP_1}$ and $\frac{dP_1}{dP_2}$ exist.

Obtain a solution to this equation. Write down the generator and the Fokker–Planck equation. Obtain the transition probability density, the stationary distribution, and formulas for the moments of this process.

7. Consider the two-dimensional Itô stochastic equation

$$dX_t = b(X_t) dt + \sigma(X_t) dW_t, \quad (3.114)$$

where W_t is a standard two-dimensional Brownian motion and $\sigma(x) \in \mathbb{R}^{2 \times 2}$ a uniformly positive definite matrix. Investigate whether a generalization of the Lamperti transformation (3.91) to two dimensions exists, i.e., whether there exists a transformation that maps (3.114) to an SDE with additive noise. In particular, find conditions on $\sigma(x)$ such that such a transformation exists. What is the analogue of this transformation at the level of the backward and forward Kolmogorov equations? Is such a transformation still possible when we consider a diffusion process in a bounded domain with reflecting, absorbing, or periodic boundary conditions?

8. (See [93, Chap. 6].)

- (a) Consider the Itô equation

$$dX_t = f(X_t) dt + \sigma g(X_t) dW_t. \quad (3.115)$$

Define

$$Z(x) = \frac{f(x)}{g(x)} - \frac{1}{2} \frac{dg}{dx}(x) \quad \text{and} \quad \theta(x) = -\frac{1}{\frac{dZ}{dx}(x)} \frac{d}{dx} \left(g(x) \frac{dZ}{dx}(x) \right).$$

Assume that

$$\theta(x) = \text{const} \equiv \theta. \quad (3.116)$$

Define the diffusion process

$$Y_t = \exp(\theta B(X_t)), \quad B(x) = \int^x \frac{1}{g(z)} dz.$$

Show that when (3.116) is satisfied, Y_t is the solution of the linear SDE

$$dY_t = (\alpha + \beta Y_t) dt + (\gamma + \sigma Y_t) dW_t. \quad (3.117)$$

- (b) Apply this transformation to obtain the solution and the transition probability density of the Stratonovich equation

$$dX_t = -\frac{1}{2\sqrt{2}} \tanh(2\sqrt{2}X_t) dt + \frac{\sigma}{4} \operatorname{sech}(2\sqrt{2}X_t) \circ dW_t. \quad (3.118)$$

- (c) Do the same for the Verhulst SDE

$$dX_t = (\lambda X_t - X_t^2) dt + \sigma X_t dW_t. \quad (3.119)$$

Chapter 4

The Fokker–Planck Equation

In Chap. 2, we derived the backward and forward (Fokker–Planck) Kolmogorov equations.¹ The Fokker–Planck equation enables us to calculate the transition probability density, which we can use to calculate the expectation value of observables of a diffusion process. In this chapter, we study various properties of this equation such as existence and uniqueness of solutions, long-time asymptotics, boundary conditions, and spectral properties of the Fokker–Planck operator. We also study in some detail various examples of diffusion processes and of the associated Fokker–Planck equation. We will restrict attention to time-homogeneous diffusion processes, for which the drift and diffusion coefficients do not depend on time.

In Sect. 4.1, we study various basic properties of the Fokker–Planck equation, including existence and uniqueness of solutions and boundary conditions. In Sect. 4.2, we present some examples of diffusion processes and use the corresponding Fokker–Planck equation to calculate statistical quantities such as moments. In Sect. 4.3, we study diffusion processes in one dimension. In Sect. 4.4, we study the Ornstein–Uhlenbeck process and the spectral properties of the corresponding Fokker–Planck operator. In Sect. 4.5, we study stochastic processes whose drift is given by the gradient of a scalar function, the so-called Smoluchowski equation. In Sect. 4.6, we study properties of the Fokker–Planck equation corresponding to reversible diffusions. In Sect. 4.7, we solve the Fokker–Planck equation for a reversible diffusion using eigenfunction expansions. In Sect. 4.8, we introduce very briefly Markov chain Monte Carlo techniques. In Sect. 4.9, we study the connection between the Fokker–Planck operator, the generator of a diffusion process, and the Schrödinger operator. Discussion and bibliographical remarks are included in Sect. 4.10. Exercises can be found in Sect. 4.11.

¹ In this chapter we will call this equation the Fokker–Planck equation, which is more customary in the physics literature, rather than the forward Kolmogorov equation, which is more customary in the mathematics literature.

4.1 Basic Properties of the Fokker–Planck Equation

We consider a time-homogeneous diffusion process X_t on \mathbb{R}^d with drift vector $\mathbf{b}(x)$ and diffusion matrix $\Sigma(x)$. We assume that the initial condition X_0 is a random variable with probability density function $\rho_0(x)$. The transition probability density $p(x, t)$, if it exists and is a $C^{2,1}(\mathbb{R}^d \times \mathbb{R}^+)$ function, is the solution of the initial value problem for the Fokker–Planck (backward Kolmogorov) equation that we derived in Chap. 2:

$$\frac{\partial p}{\partial t} = \nabla \cdot \left(-\mathbf{b}(x)p + \frac{1}{2}\nabla \cdot (\Sigma(x)p) \right) \quad (4.1a)$$

$$= -\sum_{j=1}^d \frac{\partial}{\partial x_j} (b_j(x)p) + \frac{1}{2} \sum_{i,j=1}^d \frac{\partial^2}{\partial x_i \partial x_j} (\Sigma_{ij}(x)p),$$

$$p(x, 0) = \rho_0(x). \quad (4.1b)$$

The Fokker–Planck equation (4.1a) can be written in equivalent forms that are often useful. First, we can rewrite it in the form

$$\frac{\partial p}{\partial t} = \nabla \cdot (\hat{\mathbf{b}}(x)p) + \frac{1}{2}\nabla \cdot (\Sigma(x)\nabla p), \quad (4.2)$$

with

$$\hat{\mathbf{b}}(x) = -\mathbf{b}(x) + \frac{1}{2}\nabla \cdot \Sigma(x). \quad (4.3)$$

We can also write the Fokker–Planck equation in nondivergence form:

$$\frac{\partial p}{\partial t} = \frac{1}{2}\Sigma(x) : D^2 p + \tilde{\mathbf{b}}(x) \cdot \nabla p + c(x)p, \quad (4.4)$$

where

$$\tilde{\mathbf{b}}(x) = -\mathbf{b}(x) + \nabla \cdot \Sigma(x), \quad c(x) = -\nabla \cdot \mathbf{b}(x) + \frac{1}{2}\nabla \cdot (\nabla \cdot \Sigma)(x). \quad (4.5)$$

By definition [see Eq. (2.62)], the diffusion matrix is always symmetric and nonnegative. We will assume that it is uniformly positive definite: there exists a constant $\alpha > 0$ such that

$$\langle \xi, \Sigma(x)\xi \rangle \geq \alpha \|\xi\|^2, \quad \forall \xi \in \mathbb{R}^d, \quad (4.6)$$

uniformly in $x \in \mathbb{R}^d$. We will refer to this as the uniform ellipticity assumption. This assumption is sufficient to guarantee the existence of the transition probability density; see the discussion in Sect. 6.2.

Furthermore, we will assume that the coefficients in (4.4) are smooth and that they satisfy the growth conditions

$$\|\Sigma(x)\| \leq M, \quad \|\tilde{\mathbf{b}}(x)\| \leq M(1 + \|x\|), \quad \|c(x)\| \leq M(1 + \|x\|^2). \quad (4.7)$$

Definition 4.1. We will call a solution to the initial value problem for the Fokker–Planck equation (4.1) a classical solution if:

- (i) $p(x, t) \in C^{2,1}(\mathbb{R}^d, \mathbb{R}^+)$.
- (ii) $\forall T > 0$, there exists a $c > 0$ such that

$$\|p(x, t)\|_{L^\infty(0, T)} \leq c e^{\alpha \|x\|^2}.$$

- (iii) $\lim_{t \rightarrow 0} p(x, t) = \rho_0(x)$.

We can prove that under the regularity and uniform ellipticity assumptions, the Fokker–Planck equation has a unique smooth solution. Furthermore, we can obtain pointwise bounds on the solution.

Theorem 4.1. Assume that conditions (4.6) and (4.7) are satisfied, and assume that $|\rho_0(x)| \leq c e^{\alpha \|x\|^2}$. Then there exists a unique classical solution to the Cauchy problem for the Fokker–Planck equation. Furthermore, there exist positive constants K, δ such that

$$|p|, |p_t|, \|\nabla p\|, \|D^2 p\| \leq K t^{(-d+2)/2} \exp\left(-\frac{1}{2t} \delta \|x\|^2\right). \quad (4.8)$$

From estimates (4.8), it follows that all moments of a diffusion process whose diffusion matrix satisfies the uniform ellipticity assumption (4.6) exist. In particular, we can multiply the Fokker–Planck equation by monomials x^n , integrate over \mathbb{R}^d , and then integrate by parts. It also follows from the maximum principle for parabolic PDEs that the solution of the Fokker–Planck equation is nonnegative for all times when the initial condition $\rho_0(x)$ is nonnegative. Since $\rho_0(x)$ is the probability density function of the random variable X_0 , it is nonnegative and normalized: $\|\rho_0\|_{L^1(\mathbb{R}^d)} = 1$. The solution to the Fokker–Planck equation preserves these properties, as we would expect, since it is the transition probability density.

The Fokker–Planck equation can be written as the familiar continuity equation from continuum mechanics. We define the *probability flux (current)* to be the vector

$$\mathbf{J} := \mathbf{b}(x)p - \frac{1}{2}\nabla \cdot (\Sigma(x)p). \quad (4.9)$$

The Fokker–Planck equation can be written in the form

$$\frac{\partial p}{\partial t} + \nabla \cdot \mathbf{J} = 0. \quad (4.10)$$

Integrating the Fokker–Planck equation over \mathbb{R}^d and using the divergence theorem on the right-hand side of the equation together with (4.8) yields

$$\frac{d}{dt} \int_{\mathbb{R}^d} p(x, t) dx = 0.$$

Consequently,

$$\int_{\mathbb{R}^d} p(x, t) dx = \int_{\mathbb{R}^d} \rho_0(x) dx = 1. \quad (4.11)$$

Hence, the total probability is conserved, as expected.

The stationary Fokker–Planck equation, whose solutions give us the invariant distributions of the diffusion process X_t , can be written in the form

$$\nabla \cdot \mathbf{J}(p_s) = 0. \quad (4.12)$$

Consequently, the equilibrium probability flux is a divergence-free vector field.

Boundary Conditions for the Fokker–Planck Equation

There are many applications in which it is necessary to study diffusion processes in bounded domains. In such cases, we need to specify the behavior of the diffusion process on the boundary of the domain. Equivalently, we need to specify the behavior of the transition probability density on the boundary. Diffusion processes in bounded domains lead to initial boundary value problems for the corresponding Fokker–Planck equation.

To understand the type of boundary conditions that we can impose on the Fokker–Planck equation, let us consider the example of a random walk on the domain $\{0, 1, \dots, N\}$.² When the random walker reaches either the left or the right boundary, we can consider the following cases:

- (i) $X_0 = 0$ or $X_N = 0$, which means that the particle gets absorbed at the boundary.
- (ii) $X_0 = X_1$ or $X_N = X_{N-1}$, which means that the particle is reflected at the boundary.
- (iii) $X_0 = X_N$, which means that the particle is moving on a circle (i.e., we identify the left and right boundaries).

These three different boundary behaviors for the random walk correspond to absorbing, reflecting, and periodic boundary conditions.

Consider the Fokker–Planck equation posed in $\Omega \subset \mathbb{R}^d$, where Ω is a bounded domain with smooth boundary. Let \mathbf{J} denote the probability current, and let \mathbf{n} be the unit outward-pointing normal vector to the surface. The above boundary conditions become the following:

- (i) The transition probability density vanishes on an absorbing boundary:

$$p(x, t) = 0, \quad \text{on } \partial\Omega.$$

² Of course, a random walk is not a diffusion process. However, as we have already seen, Brownian motion can be defined as the limit of an appropriately rescaled random walk. A similar construction exists for more general diffusion processes.

(ii) There is no net flow of probability on a reflecting boundary:

$$\mathbf{n} \cdot \mathbf{J}(x, t) = 0, \quad \text{on } \partial\Omega.$$

(iii) Consider the case $\Omega = [0, L]^d$ and assume that the transition probability function is periodic in all directions with period L . We can then consider the Fokker–Planck equation in $[0, L]^d$ with periodic boundary conditions.

In the terminology of the theory of partial differential equations, absorbing boundary conditions correspond to Dirichlet boundary conditions, and reflecting boundary conditions correspond to Robin boundary conditions. We can, of course, consider mixed boundary conditions whereby part of the domain is absorbing and part of it is reflecting.

Consider now a diffusion process in one dimension on the interval $[0, L]$. The boundary conditions are

$$p(0, t) = p(L, t) = 0 \quad \text{absorbing,}$$

$$J(0, t) = J(L, t) = 0 \quad \text{reflecting,}$$

$$p(0, t) = p(L, t) \quad \text{periodic,}$$

where the probability current is defined in (4.9). An example of mixed boundary conditions is absorbing boundary conditions at the left end and reflecting boundary conditions at the right end:

$$p(0, t) = 0, \quad J(L, t) = 0.$$

4.2 Examples of Diffusion Processes and Their Fokker–Planck Equation

There are several examples of diffusion processes in one dimension for which the corresponding Fokker–Planck equation can be solved analytically. We present some examples in this section. In the next section, we study diffusion processes in one dimension using eigenfunction expansions.

Brownian Motion

First, we consider Brownian motion in \mathbb{R} . We set $b(x, t) \equiv 0$, $\Sigma(x, t) \equiv 2D > 0$. The Fokker–Planck equation for Brownian motion is the heat equation. We calculate the transition probability density for a Brownian particle that is at x_0 at time s . The Fokker–Planck equation for the transition probability density $p(x, t|x_0, s)$ is

$$\frac{\partial p}{\partial t} = D \frac{\partial^2 p}{\partial x^2}, \quad p(x, s|x_0, s) = \delta(x - x_0). \quad (4.13)$$

The solution to this equation is the Green's function (fundamental solution) of the heat equation:

$$p(x, t|y, s) = \frac{1}{\sqrt{4\pi D(t-s)}} \exp\left(-\frac{(x-y)^2}{4D(t-s)}\right). \quad (4.14)$$

Quite often, we can obtain information on the properties of a diffusion process—for example, we can calculate moments—without having to solve the Fokker–Planck equation but using only the structure of the equation. For example, using the Fokker–Planck equation (4.13), we can show that the mean squared displacement of a Brownian motion grows linearly in time. Assume that a Brownian particle was at x_0 initially. We calculate, by performing two integrations by parts [which can be justified in view of (4.14)],

$$\begin{aligned} \frac{d}{dt} \mathbb{E}W_t^2 &= \frac{d}{dt} \int_{\mathbb{R}} x^2 p(x, t|x_0, 0) dx \\ &= D \int_{\mathbb{R}} x^2 \frac{\partial^2 p(x, t|x_0, 0)}{\partial x^2} dx \\ &= 2D \int_{\mathbb{R}} p(x, t|x_0, 0) dx = 2D. \end{aligned}$$

From this calculation, we conclude that the one-dimensional Brownian motion W_t with diffusion coefficient D satisfies

$$\mathbb{E}(W_t - x_0)^2 = 2Dt.$$

Assume now that the initial condition W_0 of the Brownian particle is a random variable with distribution $\rho_0(x)$. To calculate the transition probability density of the Brownian particle, we need to solve the Fokker–Planck equation with initial condition $\rho_0(x)$. In other words, we need to take the average of the probability density function

$$p(x, t|x_0) := p(x, t|x_0, 0)$$

over all initial realizations of the Brownian particle. The solution of the Fokker–Planck equation with $p(x, 0) = \rho_0(x)$ is

$$p(x, t) = \int p(x, t|x_0) \rho_0(x_0) dx_0. \quad (4.15)$$

4.2.1 Brownian Motion with Absorbing Boundary Conditions

We can also consider Brownian motion in a bounded domain, with either absorbing, reflecting, or periodic boundary conditions. Consider the Fokker–Planck equation for Brownian motion with diffusion coefficient D , (4.13), on $[0, 1]$ with absorbing boundary conditions:

$$\frac{\partial p}{\partial t} = D \frac{\partial^2 p}{\partial x^2}, \quad p(0, t|x_0) = p(1, t|x_0) = 0, \quad p(x, 0|x_0) = \delta(x - x_0). \quad (4.16)$$

In view of the Dirichlet boundary conditions, we look for a solution to this equation in a sine Fourier series:

$$p(x, t) = \sum_{k=1}^{\infty} p_n(t) \sin(n\pi x). \quad (4.17)$$

With our choice (4.17), the boundary conditions are automatically satisfied. The Fourier coefficients of the initial condition $\delta(x - x_0)$ are

$$p_n(0) = 2 \int_0^1 \delta(x - x_0) \sin(n\pi x) dx = 2 \sin(n\pi x_0).$$

We substitute the expansion (4.17) into (4.16) and use the orthogonality properties of the Fourier basis to obtain the equations

$$\dot{p}_n = -n^2 D \pi^2 p_n \quad n = 1, 2, \dots$$

The solution of this equation is

$$p_n(t) = p_n(0) e^{-n^2 \pi^2 D t}.$$

Consequently, the transition probability density for the Brownian motion on $[0, 1]$ with absorbing boundary conditions is

$$p(x, t|x_0, 0) = 2 \sum_{n=1}^{\infty} e^{-n^2 \pi^2 D t} \sin(n\pi x_0) \sin(n\pi x).$$

Note that

$$\lim_{t \rightarrow \infty} p(x, t|x_0) = 0.$$

This is not surprising, since all Brownian particles will eventually get absorbed at the boundary.

4.2.2 Brownian Motion with Reflecting Boundary Condition

Consider now a Brownian motion with diffusion coefficient D on the interval $[0, 1]$ with reflecting boundary conditions. To calculate the transition probability density, we solve the Fokker–Planck equation, which in this case is the heat equation on $[0, 1]$ with Neumann boundary conditions:

$$\frac{\partial p}{\partial t} = D \frac{\partial^2 p}{\partial x^2}, \quad \partial_x p(0, t|x_0) = \partial_x p(1, t|x_0) = 0, \quad p(x, 0) = \delta(x - x_0).$$

The boundary conditions are satisfied by functions of the form $\cos(n\pi x)$. We look for a solution in the form of a cosine Fourier series

$$p(x, t) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty} a_n(t) \cos(n\pi x).$$

From the initial conditions, we obtain

$$a_n(0) = 2 \int_0^1 \cos(n\pi x) \delta(x - x_0) dx = 2 \cos(n\pi x_0).$$

We substitute the expansion into the PDE and use the orthonormality of the Fourier basis to obtain the equations for the Fourier coefficients:

$$\dot{a}_n = -n^2\pi^2 D a_n,$$

from which we deduce that

$$a_n(t) = a_n(0)e^{-n^2\pi^2 D t}.$$

Consequently,

$$p(x, t|x_0) = 1 + 2 \sum_{n=1}^{\infty} \cos(n\pi x_0) \cos(n\pi x) e^{-n^2\pi^2 D t}.$$

Brownian motion on $[0, 1]$ with reflecting boundary conditions is an ergodic Markov process. To see this, let us consider the stationary Fokker–Planck equation

$$\frac{\partial^2 p_s}{\partial x^2} = 0, \quad \partial_x p_s(0) = \partial_x p_s(1) = 0.$$

The unique normalized solution to this boundary value problem is $p_s(x) = 1$. Indeed, we multiply the equation by p_s , integrate by parts, and use the boundary conditions to obtain

$$\int_0^1 \left| \frac{dp_s}{dx} \right|^2 dx = 0,$$

from which it follows that $p_s(x) = 1$. Alternatively, by taking the limit of $p(x, t|x_0)$ as $t \rightarrow \infty$, we obtain the invariant distribution:

$$\lim_{t \rightarrow \infty} p(x, t|x_0) = 1.$$

Now we can calculate the stationary autocorrelation function:

$$\begin{aligned}\mathbb{E}(W_t W_0) &= \int_0^1 \int_0^1 x x_0 p(x, t | x_0) p_s(x_0) dx dx_0 \\ &= \int_0^1 \int_0^1 x x_0 \left(1 + 2 \sum_{n=1}^{\infty} \cos(n\pi x_0) \cos(n\pi x) e^{-n^2 \pi^2 D t} \right) dx dx_0 \\ &= \frac{1}{4} + \frac{8}{\pi^4} \sum_{n=0}^{+\infty} \frac{1}{(2n+1)^4} e^{-(2n+1)^2 \pi^2 D t}.\end{aligned}$$

The Ornstein–Uhlenbeck Process

We now set $b(x, t) = -\alpha x$ with $\alpha > 0$ and $\Sigma(x, t) = 2D > 0$ for the drift and diffusion coefficients, respectively. The Fokker–Planck equation for the transition probability density $p(x, t | x_0)$ is

$$\frac{\partial p}{\partial t} = \alpha \frac{\partial (xp)}{\partial x} + D \frac{\partial^2 p}{\partial x^2}, \quad (4.18a)$$

$$p(x, 0 | x_0) = \delta(x - x_0). \quad (4.18b)$$

This equation is posed on the real line, and the boundary conditions are that $p(x, t | x_0)$ decays sufficiently fast at infinity; see Definition 4.1. The corresponding stochastic differential equation is

$$dX_t = -\alpha X_t dt + \sqrt{2D} dW_t, \quad X_0 = x_0. \quad (4.19)$$

In addition to Brownian motion, there is a linear force pulling the particle toward the origin. We know that Brownian motion is not a stationary process, since the variance grows linearly in time. By adding a linear damping term, it is reasonable to expect that the resulting process can become stationary. When (4.19) is used to model the velocity or position of a particle, the noisy term on the right-hand side of the equation is related to thermal fluctuations. The diffusion coefficient D measures the strength of thermal fluctuations and is associated with the temperature:

$$D = k_B T =: \beta^{-1}, \quad (4.20)$$

where T denotes the absolute temperature and k_B is Boltzmann's constant. We will quite often use the notation β^{-1} and refer to β as the inverse temperature. We will also refer to α in (4.19) as the friction coefficient.

The solution of the stochastic differential equation (4.19) is given by (3.72). It follows then that

$$X_t \sim \mathcal{N} \left(x_0 e^{-\alpha t}, \frac{D}{\alpha} (1 - e^{-2\alpha t}) \right); \quad (4.21)$$

see Eq. (3.72) and Sect. 3.7. From this, we can immediately obtain the transition probability density, the solution of the Fokker–Planck equation

$$p(x, t|x_0) = \sqrt{\frac{\alpha}{2\pi D(1 - e^{-2\alpha t})}} \exp\left(-\frac{\alpha(x - x_0 e^{-\alpha t})^2}{2D(1 - e^{-2\alpha t})}\right). \quad (4.22)$$

We also studied the stationary Ornstein–Uhlenbeck process in Example (2.2) (for $\alpha = D = 1$) using the fact that it can be defined as a time change of the Brownian motion. We can also derive it by solving the Fokker–Planck equation (4.18), by taking the Fourier transform of (4.18a), solving the resulting first-order PDE using the method of characteristics and then taking the inverse Fourier transform. See Exercise 5 in Chap. 3.

In the limit as the friction coefficient α goes to 0, the transition probability (4.22) converges to the transition probability of the Brownian motion. Furthermore, by taking the long-time limit in (4.22), we obtain

$$\lim_{t \rightarrow +\infty} p(x, t|x_0) = \sqrt{\frac{\alpha}{2\pi D}} \exp\left(-\frac{\alpha x^2}{2D}\right),$$

irrespective of the initial position x_0 . This is to be expected, since as we have already seen, the Ornstein–Uhlenbeck process is an ergodic Markov process with a Gaussian invariant distribution

$$p_s(x) = \sqrt{\frac{\alpha}{2\pi D}} \exp\left(-\frac{\alpha x^2}{2D}\right). \quad (4.23)$$

Using now (4.22) and (4.23), we obtain the stationary joint probability density

$$\begin{aligned} p_2(x, t|x_0) &= p(x, t|x_0)p_s(x_0) \\ &= \frac{\alpha}{2\pi D \sqrt{1 - e^{-2\alpha t}}} \exp\left(-\frac{\alpha(x^2 + x_0^2 - 2xx_0 e^{-\alpha t})}{2D(1 - e^{-2\alpha t})}\right), \end{aligned}$$

or, starting at an arbitrary initial time s ,

$$p_2(x, t|x_0, s) = \frac{\alpha}{2\pi D \sqrt{1 - e^{-2\alpha|t-s|}}} \exp\left(-\frac{\alpha(x^2 + x_0^2 - 2xx_0 e^{-\alpha|t-s|})}{2D(1 - e^{-2\alpha|t-s|})}\right). \quad (4.24)$$

Now we can calculate the stationary autocorrelation function of the Ornstein–Uhlenbeck process

$$\begin{aligned} \mathbb{E}(X_t X_s) &= \int \int x x_0 p_2(x, t|x_0, s) dx dx_0 \\ &= \frac{D}{\alpha} e^{-\alpha|t-s|}. \end{aligned}$$

The derivation of this formula requires the calculation of Gaussian integrals, similar to the calculations presented in Sect. B.5. See Exercise 3 and Sect. 3.7.

Assume now that the initial condition of the Ornstein–Uhlenbeck process X_t is a random variable distributed according to a distribution $\rho_0(x)$. As in the case of a Brownian particle, the probability density function is given by the convolution integral

$$p(x, t) = \int_{\mathbb{R}} p(x, t|x_0) \rho_0(x_0) dx_0. \quad (4.25)$$

When X_0 is distributed according to the invariant distribution $p_s(x)$, given by (4.23), the Ornstein–Uhlenbeck process becomes stationary. The solution to the Fokker–Planck equation is now $p_s(x)$ at all times, and the joint probability density is given by (4.24).

Knowledge of the transition probability density enables us to calculate all moments of the Ornstein–Uhlenbeck process:

$$M_n := \mathbb{E}(X_t)^n = \int_{\mathbb{R}} x^n p(x, t) dx, \quad n = 0, 1, 2, \dots$$

In fact, we can calculate the moments using the Fokker–Planck equation rather than the explicit formula for the transition probability density. We assume that all moments of the initial distribution exist. We start with $n = 0$. We integrate (4.18a) over \mathbb{R} to obtain

$$\int \frac{\partial p}{\partial t} dx = \alpha \int \frac{\partial (xp)}{\partial x} dx + D \int \frac{\partial^2 p}{\partial x^2} dx = 0,$$

after an integration by parts and using the fact that $p(x, t)$ decays rapidly at infinity. Consequently,

$$\frac{d}{dt} M_0 = 0 \quad \Rightarrow \quad M_0(t) = M_0(0) = 1,$$

which simply means that

$$\int_{\mathbb{R}} p(x, t) dx = \int_{\mathbb{R}} \rho_0(x) dx = 1.$$

Let now $n = 1$. We multiply (4.18a) by x , integrate over \mathbb{R} , and perform an integration by parts to obtain

$$\frac{d}{dt} M_1 = -\alpha M_1.$$

Consequently,

$$M_1(t) = e^{-\alpha t} M_1(0).$$

Now we consider the case $n \geq 2$. We multiply (4.18a) by x^n and integrate by parts, once on the first term and twice on the second on the right-hand side of the equation), to obtain

$$\frac{d}{dt} M_n = -\alpha n M_n + D n(n-1) M_{n-2}, \quad n \geq 2.$$

This is a first-order linear inhomogeneous differential equation. We can solve it using the variation of constants formula

$$M_n(t) = e^{-\alpha n t} M_n(0) + D n (n-1) \int_0^t e^{-\alpha n(t-s)} M_{n-2}(s) ds. \quad (4.26)$$

We can use this formula, together with the formulas for the first two moments, to calculate all higher-order moments in an iterative way. For example, for $n = 2$, we have

$$\begin{aligned} M_2(t) &= e^{-2\alpha t} M_2(0) + 2D \int_0^t e^{-2\alpha(t-s)} M_0(s) ds \\ &= e^{-2\alpha t} M_2(0) + \frac{D}{\alpha} e^{-2\alpha t} (e^{2\alpha t} - 1) \\ &= \frac{D}{\alpha} + e^{-2\alpha t} \left(M_2(0) - \frac{D}{\alpha} \right). \end{aligned}$$

As expected, all moments of the Ornstein–Uhlenbeck process converge to their stationary values:

$$\begin{aligned} M_n^\infty &:= \sqrt{\frac{\alpha}{2\pi D}} \int_{\mathbb{R}} x^n e^{-\frac{\alpha x^2}{2D}} dx \\ &= \begin{cases} 1.3 \dots (n-1) \left(\frac{D}{\alpha}\right)^{n/2}, & n \text{ even}, \\ 0, & n \text{ odd}. \end{cases} \end{aligned}$$

In fact, it is clear from (4.26) that the moments converge to their stationary values exponentially fast and that the convergence is faster, the closer we start from the stationary distribution; see the formula for $M_2(t)$. Later in this chapter, we will study the problem of convergence to equilibrium for more general classes of diffusion processes; see Sect. 4.5.

Geometric Brownian Motion

We set $b(x) = \mu x$, $\Sigma(x) = \frac{1}{2}\sigma^2 x^2$. This is the geometric Brownian motion that we encountered in Chaps. 1 and 3. This diffusion process appears in mathematical finance and in population dynamics. The generator of this process is

$$\mathcal{L} = \mu x \frac{\partial}{\partial x} + \frac{\sigma x^2}{2} \frac{\partial^2}{\partial x^2}. \quad (4.27)$$

Observe that this operator is not uniformly elliptic, since the diffusion coefficient vanishes at $x = 0$. The Fokker–Planck equation is

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}(\mu p) + \frac{\partial^2}{\partial x^2}\left(\frac{\sigma^2 x^2}{2}p\right). \quad (4.28a)$$

$$p(x, 0|x_0) = \delta(x - x_0). \quad (4.28b)$$

Since the diffusion coefficient is not uniformly elliptic, it is not covered by Theorem 4.1. The corresponding stochastic differential equation is given by Eq. (3.74). As for the Ornstein–Uhlenbeck process, we can use the Fokker–Planck equation to obtain equations for the moments of a geometric Brownian motion:

$$\frac{d}{dt}M_1 = \mu M_1, \quad \frac{d}{dt}M_n = \left(\mu n + \frac{\sigma^2}{2}n(n-1)\right)M_n, \quad n \geq 2.$$

We can solve these equations to obtain

$$M_1(t) = e^{\mu t}M_1(0)$$

and

$$M_n(t) = e^{(\mu + (n-1)\frac{\sigma^2}{2})nt}M_n(0), \quad n \geq 2.$$

We remark that the n th moment might diverge as $t \rightarrow \infty$, depending on the values of μ and σ . Consider, for example, the second moment. We have

$$M_2(t) = e^{(2\mu + \sigma^2)t}M_2(0), \quad (4.29)$$

which diverges when $\sigma^2 + 2\mu > 0$.

4.3 Diffusion Processes in One Dimension

In this section, we study the Fokker–Planck equation for diffusion processes in one dimension in a bounded interval and with reflecting boundary conditions. Let X_t denote a diffusion process in the interval $[\ell, r]$ with drift and diffusion coefficients $b(x)$ and $\sigma(x)$, respectively. We will assume that $\sigma(x)$ is positive in $[\ell, r]$. The transition probability density $p(x, t|x_0)$ is the solution of the following initial boundary value problem for the Fokker–Planck equation:

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}\left(b(x)p - \frac{1}{2}\frac{\partial}{\partial x}(\sigma(x)p)\right) =: -\frac{\partial J}{\partial x}, \quad x \in (\ell, r), \quad (4.30a)$$

$$p(x, 0|x_0) = \delta(x - x_0), \quad (4.30b)$$

$$J(\ell, t) = J(r, t) = 0. \quad (4.30c)$$

The reflecting boundary conditions and the assumption on the positivity of the diffusion coefficient ensure that X_t is ergodic. We can calculate the unique invariant probability distribution. The stationary Fokker–Planck equation reads

$$\frac{dJ_s}{dx} = 0, \quad x \in (\ell, r), \quad (4.31a)$$

$$J_s(\ell) = J_s(r) = 0, \quad (4.31b)$$

where

$$J_s(x) := J(p_s(x)) = b(x)p_s(x) - \frac{1}{2} \frac{d}{dx}(\sigma(x)p_s(x)) \quad (4.32)$$

denotes the stationary probability flux, $p_s(x)$ being the stationary probability distribution. We use the reflecting boundary conditions (4.31) to write the stationary Fokker–Planck equation in the form

$$J_s(x) = 0, \quad x \in (\ell, r). \quad (4.33)$$

Thus, the stationary probability flux vanishes. This is called the *detailed balance* condition, and it will be discussed in Sect. 4.6. Equation (4.33) gives

$$b(x)p_s(x) - \frac{1}{2} \frac{d}{dx}(\sigma(x)p_s(x)) = 0, \quad x \in (\ell, r), \quad (4.34)$$

together with the normalization condition

$$\int_{\ell}^r p_s(x) dx = 1.$$

The detailed balance condition (4.33) results in the stationary Fokker–Planck equation becoming a first-order differential equation. The solution to (4.34) can be obtained up to one constant, which is determined from the normalization condition. The solution is

$$p_s(x) = \frac{1}{Z} \frac{1}{\sigma(x)} \exp \left(2 \int_{\ell}^x \frac{b(y)}{\sigma(y)} dy \right), \quad Z = \int_{\ell}^r \left(\frac{1}{\sigma(x)} \exp \left(2 \int_{\ell}^x \frac{b(y)}{\sigma(y)} dy \right) \right) dx. \quad (4.35)$$

Now we solve the time-dependent Fokker–Planck equation (4.30). We first transform the Fokker–Planck (forward Kolmogorov) equation into the backward Kolmogorov equation, since the boundary conditions for the generator \mathcal{L} of the diffusion process X_t are simpler than those for the Fokker–Planck operator \mathcal{L}^* . Let $p \in D(\mathcal{L}^*) := \{p \in C^2(\ell, r); J(p(\ell)) = J(p(r)) = 0\}$, the domain of definition of the Fokker–Planck operator with reflecting boundary conditions. We write $p(x) = f(x)p_s(x)$ and use the stationary Fokker–Planck equation and the detailed balance condition (4.33) to calculate

$$\begin{aligned} \mathcal{L}^* p &= \frac{d}{dx} \left(-b(x)f(x)p_s(x) + \frac{1}{2} \frac{d}{dx}(\sigma(x)f(x)p_s(x)) \right) \\ &= p_s \mathcal{L} f. \end{aligned} \quad (4.36)$$

Furthermore,

$$J(p) = J(fp_s) = -\frac{1}{2}\sigma(x)p_s(x)\frac{df}{dx}(x).$$

In particular, in view of the reflecting boundary conditions and the fact that both the diffusion coefficient and the invariant distribution are positive, we have

$$\frac{df}{dx}(\ell) = \frac{df}{dx}(r) = 0. \quad (4.37)$$

Consequently, the generator \mathcal{L} of the diffusion process X_t is equipped with Neumann boundary conditions, $D(\mathcal{L}) = \{f \in C^2(\ell, r), f'(\ell) = f'(r) = 0\}$.

Setting now $p(x, t|x_0) = f(x, t|x_0)p_s(x)$, we obtain the following initial boundary value problem:

$$\frac{\partial f}{\partial t} = b(x)\frac{\partial f}{\partial x} + \frac{1}{2}\sigma(x)\frac{\partial^2 f}{\partial x^2} =: \mathcal{L}f, \quad x \in (\ell, r), \quad (4.38a)$$

$$f(x, 0|x_0) = p_s^{-1}(x)\delta(x - x_0), \quad (4.38b)$$

$$f'(\ell, t|x_0) = f'(r, t|x_0) = 0. \quad (4.38c)$$

We solve this equation using separation of variables (we suppress the dependence on the initial condition x_0):

$$f(x, t) = \psi(x)c(t).$$

Substituting this into (4.38a), we obtain

$$\frac{\dot{c}}{c} = \frac{\mathcal{L}\psi}{\psi} = -\lambda,$$

where λ is a constant and

$$c(t) = c(0)e^{-\lambda t}, \quad -\mathcal{L}\psi = \lambda\psi.$$

Using the superposition principle, we deduce that the solution to the backward Kolmogorov equation (4.38) is

$$f(x, t) = \sum_{n=0}^{+\infty} c_n \psi_n(x) e^{-\lambda_n t}, \quad (4.39)$$

where $\{\lambda_n, \psi_n\}_{n=0}^{+\infty}$ are the eigenvalues and eigenfunctions of the generator of X_t equipped with Neumann boundary conditions:

$$-\mathcal{L}\psi_n = \lambda_n \psi_n, \quad \psi_n'(\ell) = \psi_n'(r) = 0. \quad (4.40)$$

The generator \mathcal{L} (with Neumann boundary conditions) is a self-adjoint operator in the space $L^2((\ell, r); p_s(x))$, the space of square-integrable functions in the interval

(ℓ, r) , weighted by the stationary distribution of the process. This is a Hilbert space with inner product

$$\langle f, h \rangle = \int_{\ell}^r f(x)h(x)p_s(x)dx$$

and corresponding norm $\|f\| = \sqrt{\langle f, f \rangle}$. The self-adjointness of \mathcal{L} follows from (4.36).³

$$\int_{\ell}^r \mathcal{L}f h p_s dx = \int_{\ell}^r f \mathcal{L}^*(h p_s) dx = \int_{\ell}^r f \mathcal{L} h p_s dx$$

for all $f, h \in D(\mathcal{L})$. Furthermore, $-\mathcal{L}$ is a positive operator: performing an integration by parts and using the stationary Fokker–Planck equation, we obtain

$$\int_{\ell}^r (-\mathcal{L}f) f p_s dx = \frac{1}{2} \int_{\ell}^r |f'|^2 \sigma p_s dx.$$

The generator \mathcal{L} has a discrete spectrum in the Hilbert space $L^2((\ell, r); p_s(x))$. In addition, the eigenvalues of $-\mathcal{L}$ are real, nonnegative, with $\lambda_0 = 0$ corresponding to the invariant distribution, and can be ordered $0 = \lambda_0 < \lambda_1 < \lambda_2 < \dots$. The eigenfunctions of $-\mathcal{L}$ form an orthonormal basis on $L^2((\ell, r); p_s(x))$: a function $f \in L^2((\ell, r); p_s(x))$ can be expanded in a generalized Fourier series $f = \sum f_n \psi_n$ with $f_n = \langle f, \psi_n \rangle$.

The solution to (4.38) is given by (4.39)

$$f(x, t | x_0) = \sum_{n=0}^{+\infty} c_n e^{-\lambda_n t} \psi_n(x),$$

where the constants $\{c_n\}_{n=0}^{+\infty}$ are determined from the initial conditions:

$$\begin{aligned} c_n &= \int_{\ell}^r f(x, 0 | x_0) \psi_n(x) p_s(x) dx = \int_{\ell}^r \delta(x - x_0) \psi_n(x) p_s(x) dx \\ &= \psi_n(x_0). \end{aligned}$$

Putting everything together, we obtain a solution to the time-dependent Fokker–Planck equation (4.30):

$$p(x, t | x_0) = p_s(x) \sum_{n=0}^{+\infty} e^{-\lambda_n t} \psi_n(x) \psi_n(x_0). \quad (4.41)$$

The main challenge in this approach to solving the Fokker–Planck equation is the calculation of the eigenvalues and eigenfunctions of the generator of X_t in $L^2((\ell, r); p_s(x))$. This can be done either analytically or, in most cases, numerically.

³ In fact, we prove only that \mathcal{L} is symmetric. An additional argument is needed to prove that it is self-adjoint. See the comments in Sect. 4.10.

If the initial condition of the diffusion process X_0 is distributed according to a probability distribution $\rho_0(x)$, then the solution of the stationary Fokker–Planck equation is

$$p(x, t) = p_s(x) \sum_{n=0}^{+\infty} c_n e^{-\lambda_n t} \psi_n(x), \quad c_n = \int_{\ell}^r \psi_n(x) \rho_0(x) dx. \quad (4.42)$$

Notice that from the above formula and the fact that all eigenvalues apart from the first are positive, we conclude that X_t , starting from an arbitrary initial distribution, converges to its invariant distribution exponentially fast in $L^2((\ell, r); p_s(x))$. We will consider multidimensional diffusion processes for which a similar result can be obtained later in this chapter.

4.4 The Ornstein–Uhlenbeck Process and Hermite Polynomials

The Ornstein–Uhlenbeck process that we already encountered in Sect. 4.2 is one of the few stochastic processes for which we can calculate explicitly the solution of the corresponding stochastic differential equation, the solution of the Fokker–Planck equation, as well as the eigenvalues and eigenfunctions of the generator of the process. In this section, we show that the eigenfunctions of the Ornstein–Uhlenbeck process are the Hermite polynomials and study various properties of the generator of the Ornstein–Uhlenbeck process. We will see that it has many of the properties of the generator of a diffusion process in one dimension with reflective boundary conditions that we studied in the previous section. In the next section, we will show that many of the properties of the Ornstein–Uhlenbeck process (ergodicity, self-adjointness of the generator, exponentially fast convergence to equilibrium, real discrete spectrum) are shared by a large class of diffusion processes, namely reversible diffusions.

We consider a diffusion process in \mathbb{R}^d with drift $\mathbf{b}(p) = -\alpha p$, $\alpha > 0$, and $\Sigma(p) = \beta^{-1}I$, where I denotes the $d \times d$ identity matrix. The generator of the d -dimensional Ornstein–Uhlenbeck process is

$$\mathcal{L} = -\alpha p \cdot \nabla_p + \beta^{-1} \Delta_p, \quad (4.43)$$

where, as explained in Sect. 4.2, β denotes the inverse temperature and α denotes the friction coefficient.

We have already seen that the Ornstein–Uhlenbeck process is an ergodic Markov process whose unique invariant density is the Gaussian

$$\rho_{\beta}(p) = \frac{1}{(2\pi\alpha^{-1}\beta^{-1})^{d/2}} e^{-\beta \frac{\alpha|p|^2}{2}}.$$

We can perform the same transformation as in the previous section: we have that

$$\mathcal{L}^*(h\rho_\beta(p)) = \rho_\beta(p)\mathcal{L}h. \quad (4.44)$$

The initial value problem for the Fokker–Planck equation

$$\frac{\partial \rho}{\partial t} = \mathcal{L}^*\rho, \quad \rho(x, 0) = \rho_0(x)$$

becomes

$$\frac{\partial h}{\partial t} = \mathcal{L}h, \quad h(x, 0) = \rho_\beta^{-1}(x)\rho_0(x).$$

Therefore, in order to study the Fokker–Planck equation for the Ornstein–Uhlenbeck process, it is sufficient to study the properties of the generator \mathcal{L} . As in the previous section, the natural function space for studying the generator of the Ornstein–Uhlenbeck process is the L^2 -space weighted by the invariant measure of the process. This is a (separable) Hilbert space with norm

$$\|f\|_\rho^2 := \int_{\mathbb{R}^d} f^2 \rho_\beta dp$$

and corresponding inner product

$$\langle f, h \rangle_\rho = \int_{\mathbb{R}^d} fh \rho_\beta dp.$$

We can also define weighted L^2 -spaces involving derivatives, i.e., weighted Sobolev spaces. See Exercise 6.

The generator of the Ornstein–Uhlenbeck process becomes a self-adjoint operator in this space. In fact, \mathcal{L} defined in (4.43) has many nice properties that are summarized in the following proposition.

Proposition 4.1. *The operator \mathcal{L} has the following properties:*

(i) *For every $f, h \in C^2(\mathbb{R}^d) \cap L^2(\rho_\beta)$,*

$$\langle -\mathcal{L}f, h \rangle_\rho = \beta^{-1} \int_{\mathbb{R}^d} \nabla f \cdot \nabla h \rho_\beta dp. \quad (4.45)$$

(ii) *\mathcal{L} is a nonpositive operator on $L^2(\rho_\beta)$.*

(iii) *The null space of \mathcal{L} consists of constants.*

Proof.

(i) Equation (4.45) follows from an integration by parts:

$$\begin{aligned} \langle \mathcal{L}f, h \rangle_\rho &= \int_{\mathbb{R}^d} -p \cdot \nabla f h \rho_\beta dp + \beta^{-1} \int_{\mathbb{R}^d} \Delta f h \rho_\beta dp \\ &= \int_{\mathbb{R}^d} -p \cdot \nabla f h \rho_\beta dp - \beta^{-1} \int_{\mathbb{R}^d} \nabla f \cdot \nabla h \rho_\beta dp + \int_{\mathbb{R}^d} p \cdot \nabla f h \rho_\beta dp \\ &= -\beta^{-1} \langle \nabla f, \nabla h \rangle_\rho. \end{aligned}$$

(ii) The nonpositivity of \mathcal{L} follows from (4.45) on setting $h = f$:

$$\langle \mathcal{L}f, f \rangle_\rho = -\beta^{-1} \|\nabla f\|_\rho^2 \leq 0. \quad (4.46)$$

(iii) Let $f \in \mathcal{N}(\mathcal{L})$ and use (4.46) to deduce that

$$\int_{\mathbb{R}^d} |\nabla f|^2 \rho_\beta dp = 0,$$

from which we deduce that $f \equiv \text{const.}$

□

The generator of the Ornstein–Uhlenbeck process has a spectral gap: For every $f \in C^2(\mathbb{R}^d) \cap L^2(\rho_\beta)$, we have

$$\langle -\mathcal{L}f, f \rangle_\rho \geq \text{Var}(f), \quad (4.47)$$

where $\text{Var}(f) = \int_{\mathbb{R}^d} f^2 \rho_\beta - (\int_{\mathbb{R}^d} f \rho_\beta)^2$. This statement is equivalent to the statement that the Gaussian measure $\rho_\beta(x) dx$ satisfies the *Poincaré's inequality*:

$$\int_{\mathbb{R}^d} f^2 \rho_\beta dp \leq (\alpha\beta)^{-1} \int_{\mathbb{R}^d} |\nabla f|^2 \rho_\beta dp \quad (4.48)$$

for all smooth functions with $\int f \rho_\beta = 0$. The Poincaré inequality for Gaussian measures can be proved using the fact that (tensor products of) Hermite polynomials form an orthonormal basis in $L^2(\mathbb{R}^d; \rho_\beta)$. We can also use the fact that the generator of the Ornstein–Uhlenbeck process is unitarily equivalent to the Schrödinger operator for the quantum harmonic oscillator, whose eigenfunctions are the Hermite functions:⁴ consider the generator \mathcal{L} in one dimension and set, for simplicity, $\beta = 1$, $\alpha = 2$. Then

$$\rho_\beta^{1/2} (-\mathcal{L}(h\rho_\beta^{-1/2})) = -\frac{d^2 h}{dp^2} + p^2 h - h := \mathcal{H}h. \quad (4.49)$$

The Poincaré inequality (4.48) follows from the fact that the operator $\widehat{\mathcal{H}} = -\frac{d^2}{dp^2} + p^2$ has a spectral gap,

$$\int_{\mathbb{R}} \widehat{\mathcal{H}} h h dp \geq \int_{\mathbb{R}} |h|^2 dp, \quad (4.50)$$

which, in turn, follows from the estimate

$$\|h\|_{L^2}^2 \leq 2\|ph\|_{L^2} \left\| \frac{dh}{dp} \right\|_{L^2}, \quad (4.51)$$

⁴ The transformation of the generator of a diffusion process to a Schrödinger operator is discussed in detail in Sect. 4.9.

for all smooth functions h . From (4.49), it follows that (4.48) is equivalent to

$$\int_{\mathbb{R}} \mathcal{H} h h dp \geq \int_{\mathbb{R}} h^2 dp, \quad \int_{\mathbb{R}} h e^{-p^2/2} dp = 0.$$

We can check that $\sqrt{\rho} = e^{-p^2/2}$ is the first eigenfunction of \mathcal{H} , corresponding to the zero eigenvalue, $\lambda_0 = 0$. The centering condition for f is equivalent to the condition that h is orthogonal to the ground state (i.e., the first eigenfunction) of \mathcal{H} .

Since now $\widehat{\mathcal{H}} = -\frac{d^2}{dp^2} + p^2$ is a self-adjoint operator in $L^2(\mathbb{R})$ that satisfies the spectral gap estimate (4.50), it has a discrete spectrum, and its eigenfunctions form an orthonormal basis in $L^2(\mathbb{R})$.⁵ Furthermore, its eigenvalues are positive (from (4.50), it follows that it is a positive operator), and we can check that its first nonzero eigenvalue is $\lambda_1 = 2$. Let $\{\lambda_n, \phi_n\}$ denote the eigenvalues and eigenfunctions of \mathcal{H} , and let h be a smooth L^2 function that is orthogonal to the ground state. We have

$$\int_{\mathbb{R}} \mathcal{H} h h dp = \sum_{n=1}^{\infty} \lambda_n h_n^2 \geq 2 \sum_{n=1}^{\infty} h_n^2, \quad (4.52)$$

from which (4.47), or equivalently, (4.48), follows. From Proposition 4.1 and the spectral gap estimate (4.47), it follows that the generator of the Ornstein–Uhlenbeck process is a self-adjoint operator in $L^2(\rho_\beta)$ with discrete spectrum and nonnegative eigenvalues, and its eigenfunctions form an orthonormal basis in $L^2(\rho_\beta)$.

The connection between the generator of the Ornstein–Uhlenbeck process and the Schrödinger operator for the quantum harmonic oscillator can be used to calculate the eigenvalues and eigenfunctions of \mathcal{L} . We present the results in one dimension. The multidimensional problem can be treated similarly by taking tensor products of the eigenfunctions of the one-dimensional problem.

Theorem 4.2. *Consider the eigenvalue problem for the generator of the one-dimensional Ornstein–Uhlenbeck process*

$$\mathcal{L} = -\alpha p \frac{d}{dp} + \beta^{-1} \frac{d^2}{dp^2}, \quad (4.53)$$

in the space $L^2(\rho_\beta)$:

$$-\mathcal{L} f_n = \lambda_n f_n. \quad (4.54)$$

Then the eigenvalues of \mathcal{L} are the nonnegative integers multiplied by the friction coefficient:

$$\lambda_n = \alpha n, \quad n = 0, 1, 2, \dots \quad (4.55)$$

The corresponding eigenfunctions are the normalized Hermite polynomials:

$$f_n(p) = \frac{1}{\sqrt{n!}} H_n \left(\sqrt{\alpha \beta} p \right), \quad (4.56)$$

⁵ These eigenfunctions are the Hermite functions.

where

$$H_n(p) = (-1)^n e^{\frac{p^2}{2}} \frac{d^n}{dp^n} \left(e^{-\frac{p^2}{2}} \right). \quad (4.57)$$

Note that the eigenvalues of \mathcal{L} are independent of the strength of the noise β^{-1} .⁶ This is a general property of the spectrum of the generator of linear stochastic differential equations. See Sects. 3.7 and 6.3.

From (4.57), we can see that H_n is a polynomial of degree n . Furthermore, only odd (even) powers appear in $H_n(p)$ when n is odd (even). In addition, the coefficient multiplying p^n in $H_n(p)$ is always 1. The orthonormality of the modified Hermite polynomials $f_n(p)$ defined in (4.56) implies that

$$\int_{\mathbb{R}} f_n(p) f_m(p) \rho_{\beta}(p) dp = \delta_{nm}.$$

The first few Hermite polynomials and the corresponding rescaled/normalized eigenfunctions of the generator of the Ornstein–Uhlenbeck process are

$$\begin{aligned} H_0(p) &= 1, & f_0(p) &= 1, \\ H_1(p) &= p, & f_1(p) &= \sqrt{\beta} p, \\ H_2(p) &= p^2 - 1, & f_2(p) &= \frac{\alpha\beta}{\sqrt{2}} p^2 - \frac{1}{\sqrt{2}}, \\ H_3(p) &= p^3 - 3p, & f_3(p) &= \frac{\alpha\beta^{3/2}}{\sqrt{6}} p^3 - \frac{3\sqrt{\alpha\beta}}{\sqrt{6}} p \\ H_4(p) &= p^4 - 3p^2 + 3, & f_4(p) &= \frac{1}{\sqrt{24}} \left((\alpha\beta)^2 p^4 - 3\alpha\beta p^2 + 3 \right) \\ H_5(p) &= p^5 - 10p^3 + 15p, & f_5(p) &= \frac{1}{\sqrt{120}} \left((\alpha\beta)^{5/2} p^5 - 10(\alpha\beta)^{3/2} p^3 + 15(\alpha\beta)^{1/2} p \right). \end{aligned}$$

Proof of Theorem 4.2. We already know that \mathcal{L} has a discrete nonnegative spectrum and that its eigenfunctions span $L^2(\rho_{\beta})$. We can calculate the eigenvalues and eigenfunctions by introducing appropriate creation and annihilation operators. We define the annihilation operator

$$a^- = \frac{1}{\sqrt{\beta}} \frac{d}{dp} \quad (4.58)$$

and the creation operator

$$a^+ = \sqrt{\beta} \alpha p - \frac{1}{\sqrt{\beta}} \frac{d}{dp}. \quad (4.59)$$

⁶ Of course, the function space $L^2(\rho_{\beta})$ in which we study the eigenvalue problem for \mathcal{L} does depend on β through ρ_{β} .

These two operators are $L^2(\rho_\beta)$ -adjoint:

$$\langle a^- f, h \rangle_\rho = \langle f, a^+ h \rangle_\rho,$$

for all C^1 functions f, h in $L^2(\rho_\beta)$. Using these operators we can write the generator \mathcal{L} in the form

$$\mathcal{L} = -a^+ a^-.$$

The eigenvalue problem (4.54) becomes

$$a^+ a^- f_n = \lambda_n f_n.$$

Furthermore, we easily check that a^+ and a^- satisfy the following commutation relation:

$$[a^+, a^-] = a^+ a^- - a^- a^+ = -\alpha.$$

Now we calculate

$$\begin{aligned} [\mathcal{L}, a^+] &= \mathcal{L}^+ a^+ - a^+ \mathcal{L} = -a^+ a^- a^+ + a^+ a^+ a^- \\ &= -a^+ a^- a^+ + a^+ (a^- a^+ - \alpha) \\ &= -\alpha. \end{aligned}$$

We proceed by induction to show that

$$[\mathcal{L}, (a^+)^n] = -\alpha n (a^+)^n. \quad (4.60)$$

Define now

$$\phi_n = (a^+)^n \mathbf{1}, \quad (4.61)$$

where $\phi_0 = \mathbf{1}$ is the “ground state” corresponding to the eigenvalue $\lambda_0 = 0$, $\mathcal{L}_0 \phi_0 = 0$. We use (4.60) and the fact that $\mathcal{L} \mathbf{1} = 0$ to check that ϕ_n is the n th unnormalized eigenfunction of the generator \mathcal{L} :

$$\begin{aligned} -\mathcal{L} \phi_n &= -\mathcal{L} (a^+)^n \mathbf{1} = -(a^+)^n \mathcal{L} \mathbf{1} - [\mathcal{L}, (a^+)^n] \mathbf{1} \\ &= \alpha n (a^+)^n \mathbf{1} = \alpha n \phi_n. \end{aligned}$$

We can check by induction that the eigenfunctions defined in (4.61) are the unnormalized Hermite polynomials. We present the calculation of the first few eigenfunctions:

$$\phi_0 = \mathbf{1}, \quad \phi_1 = a^+ \phi_0 = \sqrt{\beta} \alpha p, \quad \phi_2 = a^+ \phi_1 = \beta \alpha^2 p^2 - \alpha.$$

Since a^+ and a^- are $L^2(\rho_\beta)$ -adjoint, we have that

$$\langle \phi_n, \phi_m \rangle_\rho = 0, \quad n \neq m.$$

On normalizing $\{\phi_n\}_{n=0}^{+\infty}$, we obtain (4.56). The normalization constant

$$\|\phi_n\|_\rho = \sqrt{\langle (a^+)^n \mathbf{1}, (a^+)^n \mathbf{1} \rangle_\rho}$$

can be calculated by induction. \square

From the eigenfunctions and eigenvalues of \mathcal{L} and using the transformation (4.44), we conclude that the Fokker–Planck operator of the Ornstein–Uhlenbeck process has the same eigenvalues as the generator \mathcal{L} , and the eigenfunctions are obtained by multiplying those of the generator by the invariant distribution:

$$-\mathcal{L}^*(\rho_\beta f_n) = \alpha n \rho_\beta f_n, \quad n = 0, 1, \dots \quad (4.62)$$

Using the eigenvalues and eigenfunctions of the Fokker–Planck operator (or equivalently, of the generator), we can solve the time-dependent problem and obtain a formula for the probability density function [compare with (4.41)]:

$$\rho(p, t) = \rho_\beta(p) \sum_{n=0}^{+\infty} c_n e^{-\lambda_n t} f_n(p), \quad c_n = \int_{\mathbb{R}} f_n(p) \rho_0(p) dx, \quad (4.63)$$

where $\{f_n\}_{n=0}^{+\infty}$ denotes the eigenvalues of the generator. From this formula, we deduce that beginning from an arbitrary initial distribution $\rho_0(p) \in L^2(\mathbb{R}; \rho_\beta^{-1})$, the law of the process converges exponentially fast to the invariant distribution.

4.5 The Smoluchowski Equation

The Ornstein–Uhlenbeck process is an example of an ordinary differential equation with a gradient structure that is perturbed by noise: letting $V(x) = \frac{1}{2}\alpha|x|^2$, we can write the SDE for the Ornstein–Uhlenbeck process in the form

$$dX_t = -\nabla V(X_t) dt + \sqrt{2\beta^{-1}} dW_t.$$

The generator can be written as

$$\mathcal{L} = -\nabla V(x) \cdot \nabla + \beta^{-1} \Delta. \quad (4.64)$$

The Gaussian invariant distribution of the Ornstein–Uhlenbeck process can be written in the form

$$\rho_\beta(x) = \frac{1}{Z} e^{-\beta V(x)}, \quad Z = \int_{\mathbb{R}^d} e^{-\beta V(x)} dx.$$

In the previous section, we were able to obtain detailed information on the spectrum of the generator of the Ornstein–Uhlenbeck process, which in turn enabled us to solve the time-dependent Fokker–Planck equation and to obtain (4.63), from which exponentially fast convergence to equilibrium follows. We can obtain similar results using the same approach for more general classes of diffusion processes whose generators are of the form (4.64), for a quite general class of scalar functions $V(x)$. We will refer to $V(x)$ as the potential.

In this section, we consider stochastic differential equations of the form

$$dX_t = -\nabla V(X_t) dt + \sqrt{2\beta^{-1}} dW_t, \quad X_0 = x, \quad (4.65)$$

for more general potentials $V(x)$, not necessarily quadratic. The generator of the diffusion process X_t is

$$\mathcal{L} = -\nabla V(x) \cdot \nabla + \beta^{-1} \Delta. \quad (4.66)$$

Assume that the initial condition for X_t is a random variable with probability density function $\rho_0(x)$. The probability density function $\rho(x, t)$ of X_t is the solution of the initial value problem for the corresponding Fokker–Planck equation:

$$\frac{\partial \rho}{\partial t} = \nabla \cdot (\nabla V \rho) + \beta^{-1} \Delta \rho, \quad (4.67a)$$

$$\rho(x, 0) = \rho_0(x). \quad (4.67b)$$

The Fokker–Planck equation (4.67a) is often called the *Smoluchowski equation*. In the sequel, we will refer to this equation as either the Smoluchowski or the Fokker–Planck equation.

It is not possible to calculate the time-dependent solution of the Smoluchowski equation for arbitrary potentials. We can, however, always calculate the stationary solution if it exists.

Definition 4.2. A potential V will be called *confining* if $\lim_{|x| \rightarrow +\infty} V(x) = +\infty$ and

$$e^{-\beta V(x)} \in L^1(\mathbb{R}^d) \quad (4.68)$$

for all $\beta \in \mathbb{R}^+$.

In other words, a potential $V(\cdot)$ is confining if it grows sufficiently fast at infinity that (4.68) holds. The simplest example of a confining potential is the quadratic potential. A particle moving in a confining potential according to the dynamics (4.65) cannot escape to infinity. It is constrained to move in a bounded region in \mathbb{R}^d . It is reasonable, then, to expect the dynamics (4.65) in a confining potential to have nice ergodic properties.

Proposition 4.2. *Let $V(x)$ be a smooth confining potential. Then the Markov process with generator (4.66) is ergodic. The unique invariant distribution is the Gibbs distribution*

$$\rho_\beta(x) = \frac{1}{Z} e^{-\beta V(x)}, \quad (4.69)$$

where the normalization factor Z is the partition function

$$Z = \int_{\mathbb{R}^d} e^{-\beta V(x)} dx. \quad (4.70)$$

The fact that the Gibbs distribution is an invariant distribution follows by direct substitution. In fact, the stationary probability flux vanishes [compare with (4.12)]:

$$J(\rho_\beta) = -\beta^{-1} \nabla \rho_\beta - \nabla V \rho_\beta = 0.$$

Uniqueness follows from the fact that the Fokker–Planck operator has a spectral gap; see the discussion later in the section.

Just as with one-dimensional diffusion processes with reflecting boundary conditions and the Ornstein–Uhlenbeck process, we can obtain the solution of the Smoluchowski equation (4.67) in terms of the solution of the backward Kolmogorov equation by a simple transformation. We define $h(x, t)$ through

$$\rho(x, t) = h(x, t)\rho_\beta(x).$$

Then we can check that the function h satisfies the backward Kolmogorov equation:

$$\frac{\partial h}{\partial t} = -\nabla V \cdot \nabla h + \beta^{-1} \Delta h, \quad h(x, 0) = \rho_0(x)\rho_\beta^{-1}(x). \quad (4.71)$$

To derive (4.71), we calculate the gradient and Laplacian of the solution to the Fokker–Planck equation:

$$\nabla \rho = \rho_\beta \nabla h - \rho_\beta h \beta \nabla V \text{ and } \Delta \rho = \rho_\beta \Delta h - 2\rho_\beta \beta \nabla V \cdot \nabla h - h \beta \Delta V \rho_\beta + h |\nabla V|^2 \beta^2 \rho_\beta.$$

We substitute these formulas into the Fokker–Planck equation to obtain (4.71).

Consequently, in order to study properties of solutions to the Fokker–Planck equation, it is sufficient to study the backward equation (4.71). The generator \mathcal{L} is self-adjoint in the right function space, which is the space of square-integrable functions weighted by the invariant density of the process X_t :

$$L^2(\rho_\beta) := \left\{ f \mid \int_{\mathbb{R}^d} |f|^2 \rho_\beta dx < \infty \right\}, \quad (4.72)$$

where ρ_β denotes the Gibbs distribution. This is a Hilbert space with inner product

$$\langle f, h \rangle_\rho := \int_{\mathbb{R}^d} f h \rho_\beta dx \quad (4.73)$$

and corresponding norm $\|f\|_\rho = \sqrt{\langle f, f \rangle_\rho}$.

The generator of the Smoluchowski dynamics (4.65) has the same properties as those of the generator of the Ornstein–Uhlenbeck process.

Proposition 4.3. *Assume that $V(x)$ is a smooth potential and assume that condition (4.68) holds. Then the operator*

$$\mathcal{L} = -\nabla V(x) \cdot \nabla + \beta^{-1} \Delta$$

is self-adjoint in $L^2(\rho_\beta)$. Furthermore, it is nonpositive, and its kernel consists of constants.

Proof. Let $f, h \in C_0^2(\mathbb{R}^d) \cap L^2(\rho_\beta)$. We calculate

$$\begin{aligned} \langle \mathcal{L}f, h \rangle_\rho &= \int_{\mathbb{R}^d} (-\nabla V \cdot \nabla + \beta^{-1} \Delta) f h \rho_\beta \, dx \\ &= \int_{\mathbb{R}^d} (\nabla V \cdot \nabla f) h \rho_\beta \, dx - \beta^{-1} \int_{\mathbb{R}^d} \nabla f \nabla h \rho_\beta \, dx - \beta^{-1} \int_{\mathbb{R}^d} \nabla f h \nabla \rho_\beta \, dx \\ &= -\beta^{-1} \int_{\mathbb{R}^d} \nabla f \cdot \nabla h \rho_\beta \, dx, \end{aligned} \quad (4.74)$$

from which self-adjointness follows.⁷

If we now set $f = h$ in the above equation, we get

$$\langle \mathcal{L}f, f \rangle_\rho = -\beta^{-1} \|\nabla f\|_{\rho_\beta}^2,$$

which shows that \mathcal{L} is nonpositive.

Clearly, constants are in the null space of \mathcal{L} . Assume that $f \in \mathcal{N}(\mathcal{L})$. Then from the above equation, we get

$$\int |\nabla f|^2 \rho_\beta \, dx = 0,$$

from which we deduce that f is constant. \square

The expression

$$D_{\mathcal{L}}(f) := \langle -\mathcal{L}f, f \rangle_\rho \quad (4.75)$$

is called the *Dirichlet form* of the generator \mathcal{L} . In the case of a gradient flow, it takes the form

$$D_{\mathcal{L}}(f) = \beta^{-1} \int_{\mathbb{R}^d} |\nabla f|^2 \rho_\beta(x) \, dx. \quad (4.76)$$

Several properties of the diffusion process X_t can be studied by looking at the corresponding Dirichlet form.

Using now Proposition 4.3, we can study the problem of convergence to equilibrium for X_t . In particular, we can show that the solution of the Fokker–Planck equation (4.67) for an arbitrary initial distribution $\rho_0(x)$ converges to the Gibbs distribution exponentially fast. To prove this, we need a functional inequality that is a property of the potential V . In particular, we need to use the fact that under appropriate assumptions on V , the Gibbs measure $\mu(dx) = Z^{-1} e^{-\beta V(x)} dx$ satisfies a *Poincaré inequality*:

Theorem 4.3. *Let $V \in C^2(\mathbb{R}^d)$ and define $\mu(dx) = \frac{1}{Z} e^{-V} dx$. If*

$$\lim_{|x| \rightarrow +\infty} \left(\frac{|\nabla V(x)|^2}{2} - \Delta V(x) \right) = +\infty, \quad (4.77)$$

⁷ A complete proof of this result would require a more careful study of the domain of definition of the generator.

then $\mu(dx)$ satisfies the Poincaré inequality with constant $\lambda > 0$: for every $f \in C^1(\mathbb{R}^d) \cap L^2(\mu)$ with $\int f \mu(dx) = 0$, there exists a constant $\lambda > 0$ such that

$$\lambda \|f\|_{L^2(\mu)}^2 \leq \|\nabla f\|_{L^2(\mu)}^2. \quad (4.78)$$

For simplicity, we will sometimes say that the potential V , rather than the corresponding Gibbs measure, satisfies the Poincaré inequality. Clearly, if $\mu(dx) = \frac{1}{Z}e^{-V}dx$ satisfies (4.78), so does $\mu_\beta(dx) = \frac{1}{Z}e^{-\beta V}dx$ for all positive β . Examples of potentials (Gibbs measures) that satisfy the Poincaré inequality are quadratic potentials in \mathbb{R}^d of the form $V(x) = \frac{1}{2}x^T D x$ with $D \in \mathbb{R}^d$ a strictly positive symmetric matrix and the bistable potential $V(x) = -\frac{x^2}{2} + \frac{x^4}{4}$ in \mathbb{R} . A condition that ensures that the probability measure $\mu(dx) = \frac{1}{Z}e^{-V}dx$ satisfies the Poincaré inequality with constant λ is the convexity condition

$$D^2V \geq \lambda I. \quad (4.79)$$

This is the Bakry–Emery criterion.

Notice also that using the definition of the Dirichlet form (4.76) associated with the generator \mathcal{L} , we can rewrite the Poincaré inequality in the form

$$\lambda \beta^{-1} \text{Var}(f) \leq D_{\mathcal{L}}(f), \quad (4.80)$$

for functions f in the domain of definition of the Dirichlet form. The assumption that the potential V satisfies the Poincaré inequality is equivalent to the assumption that the generator \mathcal{L} has a *spectral gap* in $L^2(\rho_\beta)$.

The proof of Theorem 4.3, or of the equivalent spectral gap estimate (4.80), is beyond the scope of this book. We remark that just as in the case of Gaussian measures, we can link (4.80) to the study of an appropriate Schrödinger operator. Indeed, we have (see Sect. 4.9)

$$-\rho_\beta^{-1/2} \mathcal{L} \rho_\beta^{1/2} = -\beta^{-1} \Delta + \left(\frac{\beta}{4} |\nabla V|^2 - \frac{1}{2} \Delta V \right) := -\beta^{-1} \Delta + W(x) =: \mathcal{H}. \quad (4.81)$$

If we can prove that the operator \mathcal{H} has a spectral gap in $L^2(\mathbb{R}^d)$, we can then use the expansion of $L^2(\mathbb{R}^d)$ -functions in eigenfunctions of \mathcal{H} to prove (4.80); see estimate (4.52) for the quadratic potential. This amounts to proving the estimate

$$\beta^{-1} \int_{\mathbb{R}^d} |\nabla h|^2 dx + \int_{\mathbb{R}^d} W(x) h^2 dx \geq \lambda \int_{\mathbb{R}^d} |h|^2 dx.$$

To prove this, it is sufficient to prove an inequality analogous to (4.51):

$$\lambda \|h\|_{L^2}^2 \leq 2\sqrt{\beta^{-1}} \|\sqrt{W}h\|_{L^2} \|\nabla h\|_{L^2} \quad (4.82)$$

for all C^1 -functions with compact support. Without loss of generality we have assumed that the potential W is nonnegative. It is clear that the behavior of $W(\cdot)$ at infinity, assumption (4.77), plays a crucial role in obtaining such an estimate.

The Poincaré inequality yields exponentially fast convergence to equilibrium in the right function space.

Theorem 4.4. *Let $\rho(x, t)$ denote the solution of the Fokker–Planck equation (4.67) with $\rho_0(x) \in L^2(\mathbb{R}^d; \rho_\beta^{-1})$, and assume that the potential V satisfies a Poincaré inequality with constant λ . Then $\rho(x, t)$ converges to the Gibbs distribution ρ_β defined in (4.69) exponentially fast:*

$$\|\rho(\cdot, t) - \rho_\beta\|_{L^2(\rho_\beta^{-1})} \leq e^{-\lambda\beta^{-1}t} \|\rho_0(\cdot) - \rho_\beta\|_{L^2(\rho_\beta^{-1})}. \quad (4.83)$$

Proof. We can rewrite (4.71) for the mean-zero function $h - 1$:

$$\frac{\partial(h - 1)}{\partial t} = \mathcal{L}(h - 1).$$

We multiply this equation by $(h - 1)\rho_\beta$, integrate, and use (4.76) to obtain

$$\frac{1}{2} \frac{d}{dt} \|h - 1\|_\rho^2 = -D_{\mathcal{L}}(h - 1, h - 1).$$

We now apply the Poincaré inequality in the form (4.80) to deduce

$$\begin{aligned} \frac{1}{2} \frac{d}{dt} \|h - 1\|_\rho^2 &= -D_{\mathcal{L}}(h - 1, h - 1) \\ &\leq -\beta^{-1}\lambda \|h - 1\|_\rho^2. \end{aligned}$$

Our assumption on ρ_0 implies that $h \in L^2(\rho_\beta)$. Consequently, the above calculation shows that

$$\|h(\cdot, t) - 1\|_\rho \leq e^{-\lambda\beta^{-1}t} \|h(\cdot, 0) - 1\|_\rho.$$

This estimate, together with the definition of h through $\rho = \rho_\beta h$, leads to (4.83). \square

The proof of the above theorem is based on the fact that the weighted L^2 -norm of $h - 1$ is a *Lyapunov function* for the backward Kolmogorov equation (4.71). In fact, we can construct a whole family of Lyapunov functions for the diffusion process X_t .

Proposition 4.4. *Let $\phi(\cdot) \in C^2(\mathbb{R}^d)$ be a convex function on \mathbb{R} and define*

$$H(h) = \int \phi(h) \rho_\beta dx \quad (4.84)$$

with $\rho_\beta = \frac{1}{Z} e^{-\beta V}$. Furthermore, assume that V is a confining potential and let $h(t, \cdot)$ be the solution of (4.71). Then

$$\frac{d}{dt}(H(h(t, \cdot))) \leq 0. \quad (4.85)$$

Proof. We use (4.74) to calculate

$$\begin{aligned}\frac{d}{dt}H(h(t,\cdot)) &= \frac{d}{dt}\int\phi(h)\rho_\beta dx = \int\phi'(h)\frac{\partial h}{\partial t}\rho_\beta dx \\ &= \int\phi'(h)\mathcal{L}h\rho_\beta dx = -\beta^{-1}\int\nabla\phi'(h)\nabla h\rho_\beta dx \\ &= -\beta^{-1}\int\phi''(h)|\nabla h|^2\rho_\beta dx \\ &\leq 0,\end{aligned}$$

since $\phi(\cdot)$ is convex. \square

In Theorem 4.4, we used $\phi(h) = (h - 1)^2$. In view of the previous proposition, another choice is

$$\phi(h) = h \ln h - h + 1. \quad (4.86)$$

From a thermodynamic perspective, this a more natural choice, since in this case, the Lyapunov functional (4.84) becomes (a multiple of) the *free energy functional*

$$F(\rho) = \int V\rho dx + \beta^{-1} \int \rho \ln \rho dx + \beta^{-1} \ln Z. \quad (4.87)$$

To show this we calculate, using the notation $H(\rho)$ instead of $H(h)$,

$$\begin{aligned}H(\rho) &= \int \phi(h)\rho_\beta dx = \int (h \ln h - h + 1)\rho_\beta dx \\ &= \int \rho \ln \left(\frac{\rho}{\rho_\beta} \right) dx \\ &= \int \rho \ln \rho dx - \int \rho \ln \left(Z^{-1} e^{-\beta V} \right) dx \\ &= \beta \int V\rho dx + \int \rho \ln \rho dx + \ln Z.\end{aligned}$$

We will also refer to the functional

$$H(\rho) = \int \rho \ln \left(\frac{\rho}{\rho_\beta} \right) dx$$

as the *relative entropy* between the probability densities ρ and ρ_β and we will denote it by $H(\rho|\rho_\beta)$. It is possible to prove exponentially fast convergence to equilibrium for the Smoluchowski equation under appropriate assumptions on the potential V . This result requires that the measure $\frac{1}{Z}e^{-V}$ satisfy a *logarithmic Sobolev inequality*. See the discussion in Sect. 4.10.

4.6 Reversible Diffusions

The stationary ($X_0 \sim \rho_\beta(x) dx$) diffusion process X_t with generator (4.66) that we studied in the previous section is an example of a (time-) reversible Markov process.

Definition 4.3. A stationary stochastic process X_t is time-reversible if its law is invariant under time reversal: for every $T \in (0, +\infty)$, X_t and the time-reversed process X_{T-t} have the same distribution.

This definition means that the processes X_t and X_{T-t} have the same finite-dimensional distributions. Equivalently, for each $N \in \mathbb{N}^+$, a collection of times $0 = t_0 < t_1 < \dots < t_N = T$, and bounded measurable functions with compact support f_j , $j = 0, \dots, N$, we have

$$\mathbb{E}_\mu \prod_{j=0}^N f_j(X_{t_j}) = \mathbb{E}_\mu \prod_{j=0}^N f_j(X_{T-t_j}), \quad (4.88)$$

where $\mu(dx)$ denotes the invariant measure of X_t , and \mathbb{E}_μ denotes expectation with respect to μ .

Reversible diffusion processes can be characterized in terms of the properties of their generator. Indeed, time-reversal, which is a symmetry property of the process X_t , is equivalent to the self-adjointness (which is also a symmetry property) of the generator in the Hilbert space $L^2(\mathbb{R}^d; \mu)$.

Theorem 4.5. *A stationary diffusion process X_t in \mathbb{R}^d with generator \mathcal{L} and invariant measure μ is reversible if and only if its generator is self-adjoint in $L^2(\mathbb{R}^d; \mu)$.*

The calculations presented in the following proof are rather formal. In particular, we do not distinguish between a symmetric and a self-adjoint operator. For a fully rigorous proof of this result, we need to be more careful with issues such as the domain of definition of the generator and its adjoint. See the discussion in Sect. 4.10.

Proof. It is sufficient to show that (4.88) holds if and only if the generator is self-adjoint in $L^2(\mathbb{R}^d; \mu)$. Assume first (4.88). We take $N = 1$ and $t_0 = 0, t_1 = T$ to deduce that

$$\mathbb{E}_\mu (f_0(X_0) f_1(X_T)) = \mathbb{E}_\mu (f_0(X_T) f_1(X_0)), \quad \forall f_0, f_1 \in L^2(\mathbb{R}^d; \mu).$$

This is equivalent to

$$\int (e^{\mathcal{L}t} f_0(x)) f_1(x) \mu(dx) = \int f_0(x) (e^{\mathcal{L}t} f_1(x)) \mu(dx),$$

i.e.,

$$\langle e^{\mathcal{L}t} f_1, f_2 \rangle_{L^2_\mu} = \langle f_1, e^{\mathcal{L}t} f_2 \rangle_{L^2_\mu}, \quad \forall f_1, f_2 \in L^2(\mathbb{R}^d; \mu). \quad (4.89)$$

Consequently, the semigroup $e^{\mathcal{L}t}$ generated by \mathcal{L} is self-adjoint. Differentiating (4.89) at $t = 0$ gives that \mathcal{L} is self-adjoint.

Conversely, assume that \mathcal{L} is self-adjoint in $L^2(\mathbb{R}^d; \mu)$. We will use an induction argument. Our assumption of self-adjointness implies that (4.88) is true for $N = 1$:

$$\mathbb{E}_\mu \prod_{j=0}^1 f_j(X_{t_j}) = \mathbb{E}_\mu \prod_{j=0}^1 f_j(X_{T-t_j}). \quad (4.90)$$

Assume that it is true for $N = k$. Using Eq. (2.22) and denoting by $p(t, x, \Gamma)$ the transition function of the Markov process X_t , we have

$$\begin{aligned} \mathbb{E}_\mu \prod_{j=0}^k f_j(X_{t_j}) &= \int \dots \int f_0(x_0) \mu(dx_0) \prod_{j=1}^k f_j(x_j) p(t_j - t_{j-1}, x_{j-1}, dx_j) \\ &= \mathbb{E}_\mu \prod_{n=0}^k f_j(X_{T-t_{j-1}}) \\ &= \int \dots \int f_k(x_k) \mu(dx_k) \prod_{j=1}^k f_{j-1}(x_{j-1}) p(t_j - t_{j-1}, x_j, dx_{j-1}). \end{aligned} \quad (4.91)$$

Now we show that (4.88) is true for $N = k + 1$. We calculate, using (4.90) and (4.91),

$$\begin{aligned} \mathbb{E}_\mu \prod_{j=1}^{k+1} f_j(X_{t_j}) &= \mathbb{E}_\mu \prod_{j=1}^k f_j(X_{t_j}) f_{k+1}(X_{t_{k+1}}) \\ &\stackrel{(2.22)}{=} \int \dots \int \mu(dx_0) f_0(x_0) \prod_{j=1}^k f_j(x_j) p(t_j - t_{j-1}, x_{j-1}, dx_j) \times \\ &\quad f_{k+1}(x_{k+1}) p(t_{k+1} - t_k, x_k, dx_{k+1}) \\ &\stackrel{(4.91)}{=} \int \dots \int \mu(dx_k) f_0(x_k) \prod_{j=1}^k f_{j-1}(x_{j-1}) p(t_j - t_{j-1}, x_j, dx_{j-1}) \times \\ &\quad f_{k+1}(x_{k+1}) p(t_{k+1} - t_k, x_k, dx_{k+1}) \\ &\stackrel{(4.91)}{=} \int \dots \int \mu(dx_k) f_0(x_k) f_{k+1}(x_{k+1}) p(t_{k+1} - t_k, x_k, dx_{k+1}) \times \\ &\quad \prod_{j=1}^k f_{j-1}(x_{j-1}) p(t_j - t_{j-1}, x_j, dx_{j-1}) \\ &\stackrel{(4.90)}{=} \int \dots \int \mu(dx_{k+1}) f_0(x_{k+1}) \prod_{j=1}^{k+1} f_{j-1}(x_{j-1}) p(t_j - t_{j-1}, x_j, dx_{j-1}) \times \\ &\quad f_{k+1}(x_{k+1}) p(t_{k+1} - t_k, x_k, dx_{k+1}) \\ &= \mathbb{E}_\mu \prod_{j=0}^{k+1} f_j(X_{T-t_j}). \end{aligned}$$

□

In the previous section, we showed that the generator of the Smoluchowski dynamics

$$\mathcal{L} = -\nabla V \cdot \nabla + \beta^{-1} \Delta$$

is a self-adjoint operator in $L^2(\rho_\beta)$, which implies that the stationary solution of (4.66) is a reversible diffusion process. More generally, consider the Itô stochastic differential equation in \mathbb{R}^d (see Sect. 3.2),

$$dX_t = b(X_t) dt + \sigma(X_t) dW_t. \quad (4.92)$$

The generator of this Markov process is

$$\mathcal{L} \cdot = b(x) \cdot \nabla + \frac{1}{2} \text{Tr}(\Sigma(x) D^2), \quad (4.93)$$

where $\Sigma(x) = \sigma(x) \sigma^T(x)$, which we assume to be strictly positive definite; see (4.6). Note that we can also write

$$\text{Tr}(\Sigma(x) D^2) = \Sigma(x) : \nabla \cdot \nabla.$$

The Fokker–Planck operator is

$$\mathcal{L}^* \cdot = \nabla \cdot \left(-b(x) \cdot + \frac{1}{2} \nabla \cdot (\Sigma(x) \cdot) \right). \quad (4.94)$$

We assume that the diffusion process has a unique invariant distribution that is the solution of the stationary Fokker–Planck equation

$$\mathcal{L}^* \rho_s = 0. \quad (4.95)$$

The stationary Fokker–Planck equation can be written as [see Eq. (4.12)]

$$\nabla \cdot J(\rho_s) = 0 \quad \text{with} \quad J(\rho_s) = b\rho_s - \frac{1}{2} \nabla \cdot (\Sigma \rho_s).$$

Note that we can write the invariant distribution ρ_s in the form

$$\rho_s = e^{-\Phi}, \quad (4.96)$$

where Φ can be thought of as a generalized potential.⁸

Let now X_t be an ergodic diffusion process with generator (4.93). We consider the stationary solution of (4.92), i.e., we set $X_0 \sim \rho_s$. Our goal is to find under

⁸ Note that we can incorporate the normalization constant in the definition of Φ . Alternatively, we can write $\rho = \frac{1}{Z} e^{-\Phi}$, $Z = \int e^{-\Phi} dx$. See, for example, the formula for the stationary distribution of a one-dimensional diffusion process with reflecting boundary conditions, Eq. (4.35). We can write it in the form

$$\rho_s(x) = \frac{1}{Z} e^{-\Phi} \quad \text{with} \quad \Phi = \log(\sigma(x)) - \left(2 \int_\ell^x \frac{b(y)}{\sigma(y)} dy \right).$$

what conditions on the drift and diffusion coefficients this process is reversible. According to Theorem 4.5, it is sufficient to check under what conditions on the drift and diffusion coefficients the generator \mathcal{L} is symmetric in $\mathcal{H} := L^2(\mathbb{R}^d; \rho_s(x) dx)$. Let $f, h \in \mathcal{H} \cap C^2(\mathbb{R}^d)$. We perform integrations by parts to calculate

$$\int b \cdot \nabla f h \rho_s dx = - \int f b \cdot \nabla h \rho_s dx - \int f h \nabla \cdot (b \rho_s) dx$$

and

$$\begin{aligned} \int (\Sigma \nabla \cdot \nabla f) h \rho_s dx &= - \int \Sigma \nabla f \cdot \nabla h \rho_s dx - \int (\nabla f h) \cdot \nabla \cdot (\Sigma \rho_s) dx \\ &= - \int \Sigma \nabla f \cdot \nabla h \rho_s dx + \int (f \nabla h) \cdot \nabla (\Sigma \rho_s) dx \\ &\quad + \int f h \nabla \cdot (\nabla \cdot (\Sigma \rho_s)) dx. \end{aligned}$$

We combine the above calculations and use the stationary Fokker–Planck equation (4.95) and the definition of the stationary probability flux $J_s := J(\rho_s)$ to deduce that

$$\begin{aligned} \langle -\mathcal{L}f, h \rangle_\rho &= \int (-\mathcal{L}f) h \rho_s dx \\ &= \frac{1}{2} \int \Sigma \nabla f \cdot \nabla h \rho_s dx + \int f \nabla h \cdot J_s dx \\ &= \frac{1}{2} \langle \Sigma \nabla f, \nabla h \rangle_\rho + \langle f, \rho_s^{-1} \nabla h \cdot J_s \rangle_\rho. \end{aligned}$$

The generator \mathcal{L} is symmetric if and only if the last term on the right-hand side of the above equation vanishes, i.e., if and only if the stationary probability flux vanishes:

$$J(\rho_s) = 0. \quad (4.97)$$

This is the *detailed balance* condition. From the detailed balance condition (4.97), we obtain a relation between the drift vector b , the diffusion matrix Σ , and the generalized potential Φ :

$$b = \frac{1}{2} \rho_s^{-1} \nabla \cdot (\Sigma \rho_s) = \frac{1}{2} \nabla \cdot \Sigma - \frac{1}{2} \Sigma \nabla \Phi. \quad (4.98)$$

We summarize the above calculations in the following proposition.

Proposition 4.5. *Let X_t denote the stationary process (4.92) with invariant distribution ρ_s , the solution of (4.95). Then X_t is reversible if and only if the detailed balance condition (4.97) holds, or equivalently, there exists a scalar function Φ such that (4.98) holds.*

Consider now an arbitrary ergodic diffusion process X_t , the solution of (4.92) with invariant distribution ρ_s . We can decompose this process into a reversible and an irreversible part in the sense that the generator can be decomposed into a

symmetric and antisymmetric part in the space $L^2(\mathbb{R}^d; \rho_s)$. To check this, we add and subtract the term $\rho_s^{-1} J_s \cdot \nabla$ from the generator \mathcal{L} and use the formula for the stationary probability flux:

$$\begin{aligned}\mathcal{L} &= \frac{1}{2} \Sigma \nabla \cdot \nabla + (b - \rho_s^{-1} J_s) \cdot \nabla + \rho_s^{-1} J_s \cdot \nabla \\ &= \frac{1}{2} \Sigma \nabla \cdot \nabla + \frac{1}{2} \rho_s^{-1} \nabla \cdot (\Sigma \rho_s) \cdot \nabla + \rho_s^{-1} J_s \cdot \nabla \\ &= \frac{1}{2} \rho_s^{-1} \nabla \cdot (\Sigma \rho_s \nabla) + \rho_s^{-1} J_s \cdot \nabla \\ &=: \mathcal{S} + \mathcal{A}.\end{aligned}$$

Clearly, the operator $\mathcal{S} = \frac{1}{2} \rho_s^{-1} \nabla \cdot (\Sigma \rho_s \nabla)$ is symmetric in $L^2(\mathbb{R}^d; \rho_s)$. To prove that \mathcal{A} is antisymmetric in this space, we use the stationary Fokker–Planck equation written in the form $\nabla \cdot J_s = 0$:

$$\begin{aligned}\langle \mathcal{A}f, h \rangle_\rho &= \int J_s \cdot \nabla f h dx \\ &= - \int \nabla f J_s \cdot \nabla h dx - \int f h \nabla \cdot J_s dx \\ &= - \langle f, \mathcal{A}h \rangle_\rho.\end{aligned}$$

The generator of an arbitrary diffusion process in \mathbb{R}^d can be written in the useful form

$$\mathcal{L} = \rho_s^{-1} J_s \cdot \nabla + \frac{1}{2} \rho_s^{-1} \nabla \cdot (\Sigma \rho_s \nabla), \quad (4.99)$$

where the drift (advection) term on the right-hand side is antisymmetric, whereas the second-order divergence-form part is symmetric in $L^2(\mathbb{R}^d; \rho_s)$.

4.7 Eigenfunction Expansions for Reversible Diffusions

Let X_t denote the generator of a reversible diffusion. We write the generator \mathcal{L} in the form

$$\mathcal{L} = \frac{1}{2} \rho_s^{-1} \nabla \cdot (\Sigma \rho_s \nabla); \quad (4.100)$$

see Eq. (4.99). The corresponding Dirichlet form is

$$D_{\mathcal{L}}(f) := \langle -\mathcal{L}f, f \rangle_\rho = \frac{1}{2} \langle \Sigma \nabla f, \nabla f \rangle_\rho. \quad (4.101)$$

We assume that the diffusion matrix Σ is uniformly positive definite with constant α , Eq. (4.6). This implies that

$$D_{\mathcal{L}}(f) \geq \frac{\alpha}{2} \int_{\mathbb{R}^d} |\nabla f|^2 e^{-\Phi} dx,$$

where we have introduced the generalized potential Φ , $\rho_s = e^{-\Phi}$. To prove that the generator (4.100) has a spectral gap, we need to show that the probability measure $\rho_s dx$, or equivalently, the potential Φ , satisfies a Poincaré inequality. For this, it is sufficient to show that the generalized potential satisfies assumption (4.77) in Theorem 4.3.

Assume now that the generator \mathcal{L} has a spectral gap:

$$\lambda \text{Var}(f) \leq D_{\mathcal{L}}(f). \quad (4.102)$$

Then $-\mathcal{L}$ is a nonnegative self-adjoint operator in $L^2(\mathbb{R}^d; \rho_s)$ with discrete spectrum. The eigenvalue problem for the generator is

$$-\mathcal{L}\phi_n = \lambda_n \phi_n, \quad n = 0, 1, \dots \quad (4.103)$$

Notice that $\phi_0 = 1$ and $\lambda_0 = 0$. The eigenvalues of the generator are real and non-negative:

$$0 = \lambda_0 < \lambda_1 < \lambda_2 < \dots$$

Furthermore, the eigenfunctions $\{\phi_j\}_{j=0}^{\infty}$ span $L^2(\mathbb{R}^d; \rho_s)$: we can express every element of $L^2(\mathbb{R}^d; \rho_s)$ in the form of a generalized Fourier series:

$$f = \sum_{n=0}^{\infty} \phi_n f_n, \quad f_n = \langle f, \phi_n \rangle_{\rho} \quad (4.104)$$

with $\langle \phi_n, \phi_m \rangle_{\rho} = \delta_{nm}$. This enables us to solve the time-dependent Fokker–Planck equation in terms of an eigenfunction expansion, exactly as in the case of the one-dimensional diffusion process with reflecting boundary conditions that we studied in Sect. 4.3. The calculation is exactly the same as for the one-dimensional problem: consider first the initial value problem for the transition probability density $p(x, t|x_0)$:

$$\frac{\partial p}{\partial t} = \mathcal{L}^* p, \quad (4.105a)$$

$$p(x, 0|x_0) = \delta(x - x_0). \quad (4.105b)$$

The function $h(x, t|x_0) = p(x, t|x_0)\rho_s^{-1}(x)$ is the solution of the initial value problem

$$\frac{\partial h}{\partial t} = \mathcal{L}h, \quad (4.106a)$$

$$h(x, 0|x_0) = \rho_s^{-1}(x)\delta(x - x_0). \quad (4.106b)$$

We solve this equation using separation of variables and the superposition principle. Transforming back, we finally obtain

$$p(x, t|x_0) = \rho_s(x) \left(1 + \sum_{\ell=1}^{\infty} e^{-\lambda_{\ell} t} \phi_{\ell}(x) \phi_{\ell}(x_0) \right). \quad (4.107)$$

When the initial condition is a random variable with probability density function $\rho_0(x)$, then the formula for the probability distribution function $p(x, t)$, the solution of the Fokker–Planck equation with $p(x, 0) = \rho_0(x)$, is

$$p(x, t) = \rho_s(x) \left(1 + \sum_{\ell=1}^{+\infty} e^{-\lambda_{\ell} t} \phi_{\ell}(x) \rho_{\ell} \right), \quad \rho_{\ell} = \int_{\mathbb{R}^d} \rho_0(x) \phi_{\ell}(x) dx. \quad (4.108)$$

We use now (2.60) to define the stationary autocorrelation matrix

$$C(t) := \mathbb{E}(X_t \otimes X_0) = \int \int x_0 \otimes x p(x, t | x_0) \rho_s(x_0) dx dx_0. \quad (4.109)$$

Substituting (4.107) into (4.109), we obtain

$$C(t) = \sum_{\ell=0}^{\infty} e^{-\lambda_{\ell} |t|} \alpha_{\ell} \otimes \alpha_{\ell}, \quad \alpha_{\ell} = \int_{\mathbb{R}^d} \mathbf{x} \phi_{\ell}(x) \rho_s(x) dx, \quad (4.110)$$

with $\lambda_0 = 0$, $\phi_0 = 1$. Using now (1.7), we can obtain a formula for the spectral density, which in the multidimensional case is a $d \times d$ matrix. We present here the formula in one dimension, for a mean zero process:

$$S(\omega) = \frac{1}{\pi} \sum_{k=1}^{\infty} \frac{\alpha_k^2 \lambda_k}{\lambda_k^2 + \omega^2}. \quad (4.111)$$

The interested reader is invited to supply the details of these calculations (see Exercise 11). It is important to note that for a reversible diffusion, the spectral density is given as a sum of Cauchy–Lorentz functions, which is the spectral density of the Ornstein–Uhlenbeck process.

4.8 Markov Chain Monte Carlo

Suppose that we are given a probability distribution $\pi(x)$ in \mathbb{R}^d that is known up to the normalization constant.⁹ Our goal is to sample from this distribution and to calculate expectation values of the form

$$\mathbb{E}_{\pi} f = \int_{\mathbb{R}^d} f(x) \pi(x) dx, \quad (4.112)$$

for particular choices of functions $f(x)$. A natural approach to solving this problem is to construct an ergodic diffusion process whose invariant distribution is $\pi(x)$. We then run the dynamics for a sufficiently long time, until it reaches the stationary

⁹ The calculation of the normalization constant requires the calculation of an integral (the partition function) in a high-dimensional space, which might be computationally very expensive.

regime. The equilibrium expectation (4.112) can be calculated by taking the long-time average and using the ergodic theorem:

$$\lim_{T \rightarrow +\infty} \frac{1}{T} \int_0^T f(X_s) ds = \mathbb{E}_\pi f. \quad (4.113)$$

This is an example of the *Markov chain Monte Carlo (MCMC)* methodology.

There are many different diffusion processes that we can use: we have to choose the drift and diffusion coefficients so that the stationary Fokker–Planck equation is satisfied:

$$\nabla \cdot \left(-b\pi + \frac{1}{2} \nabla \cdot (\Sigma\pi) \right) = 0. \quad (4.114)$$

We have to solve the “inverse problem” for this partial differential equation: given its solution $\pi(x)$, we want to find the coefficients $b(x)$ and $\Sigma(x)$ such that this equation is satisfied. Clearly, there are (infinitely) many solutions to this problem. We can restrict the class of drift and diffusion coefficients (and of the corresponding diffusion process) that we consider by imposing the detailed balance condition $J(\pi) = 0$. The stationary Fokker–Planck equation is

$$-b\pi + \frac{1}{2} \nabla \cdot (\Sigma\pi) = 0. \quad (4.115)$$

Thus, we consider reversible diffusion processes in order to sample from $\pi(x)$. Even when we impose the detailed balance condition, there is still considerable freedom in choosing the drift and diffusion coefficients. A natural choice is to consider a constant diffusion matrix, $\Sigma = 2I$. The drift is

$$b = \pi^{-1} \nabla \pi = \nabla \log \pi.$$

This leads to the Smoluchowski dynamics that we studied in Sect. 4.5:¹⁰

$$dX_t = \nabla \log \pi(X_t) dt + \sqrt{2} dW_t. \quad (4.116)$$

Notice that to construct this diffusion process, we do not need to know the normalization constant, since only the gradient of the logarithm of $\pi(x)$ appears in (4.116). Provided that the “potential” $V(x) = -\log \pi(x)$ satisfies the Poincaré inequality, Theorem 4.3, we have exponentially fast convergence to the target distribution in $L^2(\mathbb{R}^d; \pi)$, Theorem 4.4. The rate of convergence to the target distribution $\pi(x)$ depends only on (the tails of) the distribution itself, since the Poincaré constant depends only on the potential.

When the target distribution is multimodal (or equivalently, $V(x) = -\log \pi(x)$ has many local minima), convergence to equilibrium for the dynamics (4.116) might be slow (see Chap. 7). In such a case, it might be useful to modify the dynamics

¹⁰ In the statistics literature, this is usually called the Langevin dynamics. We will use this term for the second-order stochastic differential equation that is obtained after adding dissipation and noise to a Hamiltonian system; see Chap. 6. Using the terminology that we will introduce there, the dynamics (4.116) correspond to the overdamped Langevin dynamics.

through either the drift or the diffusion in order to facilitate the escape of the dynamics for the local minima of $V(x)$. Ideally, we would like to choose the drift and diffusion coefficients in such a way that the corresponding dynamics converges to the target distribution as quickly as possible.¹¹ For the reversible dynamics, for which the generator is a self-adjoint operator in $L^2(\mathbb{R}^d; \pi)$, the optimal choice of the diffusion process is the one that maximizes the first nonzero eigenvalue of the generator, since this determines the rate of convergence to equilibrium. The first nonzero eigenvalue can be expressed in terms of the Rayleigh quotient:

$$\lambda_1 = \min_{\phi \in D(\mathcal{L}) \setminus \{0\}} \frac{D_{\mathcal{L}}(\phi)}{\|\phi\|_{\rho}}, \quad (4.117)$$

where $D_{\mathcal{L}}(\phi) = \langle -\mathcal{L}\phi, \phi \rangle_{\rho}$. The optimal choice of the drift and diffusion coefficients is the one that maximizes λ_1 , subject to the detailed balance condition

$$\lambda_1 = \max_{b, \Sigma J(\rho_s)=0} \min_{\phi \in D(\mathcal{L}) \setminus \{0\}} \frac{D_{\mathcal{L}}(\phi)}{\|\phi\|_{\rho}}. \quad (4.118)$$

We can also introduce perturbations in the drift of the Smoluchowski dynamics that lead to irreversible diffusions. We consider the following dynamics:

$$dX_t^{\gamma} = (\nabla \log \pi(X_t^{\gamma}) + \gamma(X_t^{\gamma})) dt + \sqrt{2} dW_t, \quad (4.119)$$

where $\gamma(x)$ is a smooth vector field that has to be chosen so that the invariant distribution of (4.119) is still $\pi(x)$. The stationary Fokker–Planck equation becomes

$$\nabla \cdot (\gamma(x) \pi(x)) = 0. \quad (4.120)$$

Consequently, all divergence-free vector field perturbations with respect to the target distribution can be used to construct irreversible ergodic dynamics whose invariant distribution is $\pi(x)$. There exist many such vector fields, for example

$$\gamma(x) = J \nabla \log \pi(x), \quad J = -J^T.$$

We can then ask whether it is possible to accelerate convergence to the target distribution by choosing the irreversible perturbation, i.e., the matrix J , appropriately. It is reasonable to expect that an irreversible perturbation that facilitates the escape of the dynamics from the local minima of $V(x) = -\nabla \log \pi(x)$ would speed up convergence of the modified dynamics X_t^{γ} to equilibrium.

¹¹ When implementing the MCMC algorithm, we need to discretize the stochastic differential equation and take the discretization and Monte Carlo errors into account. See Sect. 5.2.

4.9 Reduction to a Schrödinger Operator

In Sects. 4.4 and 4.5, we used the fact that the generator of the Ornstein–Uhlenbeck process and the generator of the Smoluchowski dynamics can be transformed into a Schrödinger operator, formulas (4.49) and (4.81). We used this transformation to study Poincaré inequalities for the Gibbs measure $\mu(dx) = \frac{1}{Z}e^{-V(x)}dx$. In this section, we study in more detail the connection between the generator \mathcal{L} , the Fokker–Planck operator, and an appropriately defined Schrödinger-like operator.

We begin by writing the generator of an ergodic diffusion process in the form (4.99)

$$\mathcal{L} \cdot = \frac{1}{2}\rho_s^{-1}\nabla \cdot (\Sigma\rho_s\nabla \cdot) + \rho_s^{-1}J_s \cdot \nabla. \quad (4.121a)$$

$$=: \mathcal{S} + \mathcal{A}, \quad (4.121b)$$

where ρ_s denotes the invariant density. The operators \mathcal{S} and \mathcal{A} are symmetric and antisymmetric, respectively, in the function space $L^2(\mathbb{R}^d; \rho_s)$. Using the definition of the $L^2(\mathbb{R}^d)$ -adjoint, we can check that the Fokker–Planck operator can be written in the form

$$\mathcal{L}^* \cdot = \frac{1}{2}\nabla \cdot (\rho_s \Sigma \nabla (\rho_s^{-1} \cdot)) - J_s \cdot \nabla (\rho_s^{-1} \cdot), \quad (4.122a)$$

$$=: \mathcal{S}^* + \mathcal{A}^*. \quad (4.122b)$$

The operators \mathcal{S}^* and \mathcal{A}^* are symmetric and antisymmetric, respectively, in the function space $L^2(\mathbb{R}^d; \rho_s^{-1})$. We introduce the operator [compare with Eq. (4.81)]

$$\mathcal{H} \cdot := \rho_s^{1/2} \mathcal{L}(\rho_s^{-1/2} \cdot), \quad (4.123)$$

acting on twice-differentiable functions that belong in $L^2(\mathbb{R}^d)$.

Lemma 4.6. *The operator \mathcal{H} defined in (4.123) has the form*

$$\mathcal{H} \cdot = \frac{1}{2}\nabla \cdot (\Sigma \nabla \cdot) + W(x) \cdot + \mathcal{A} \cdot, \quad (4.124)$$

where \mathcal{A} denotes the antisymmetric part of \mathcal{L} , and the scalar function W is given by the formula

$$W(x) = \sqrt{\rho_s} \mathcal{L} \sqrt{\rho_s^{-1}}. \quad (4.125)$$

Proof. Let $f \in C^2(\mathbb{R}^d)$. We calculate

$$\begin{aligned} \mathcal{H}f &= \rho_s^{1/2} \mathcal{L}(\rho_s^{-1/2} f) = \rho_s^{1/2} \mathcal{S}(\rho_s^{-1/2} f) + \rho_s^{1/2} \mathcal{A}(\rho_s^{-1/2} f) \\ &= \rho_s^{1/2} \mathcal{S}(\rho_s^{-1/2} f) + \mathcal{A}f + f \sqrt{\rho_s} \mathcal{A} \sqrt{\rho_s^{-1}}, \end{aligned} \quad (4.126)$$

since \mathcal{A} is a first-order differential operator. Let now ψ be another C^2 function. We have the identity

$$\begin{aligned}\nabla \cdot (\Sigma \rho_s \nabla (f \psi)) &= \nabla \cdot (\Sigma \nabla f) \rho_s \psi + \nabla \cdot (\Sigma \rho_s \nabla \psi) f \\ &\quad + [\Sigma \psi \nabla \rho_s + 2 \Sigma \rho_s \nabla \psi] \cdot f.\end{aligned}$$

In particular, for $\psi = \sqrt{\rho_s^{-1}}$, the second term on the right-hand side of the above equation vanishes:

$$\nabla \cdot \left(\Sigma \rho_s \nabla \left(f \sqrt{\rho_s^{-1}} \right) \right) = \nabla \cdot (\Sigma \nabla f) \sqrt{\rho_s} + \nabla \cdot \left(\Sigma \rho_s \nabla \sqrt{\rho_s^{-1}} \right) f.$$

This equation implies that

$$\rho_s^{1/2} \mathcal{S}(\rho_s^{-1/2} f) = \frac{1}{2} \nabla \cdot (\Sigma \nabla f) + \rho_s^{1/2} \mathcal{S}(\rho_s^{-1/2}).$$

We combine this with (4.126) to obtain (4.124). \square

The operator \mathcal{H} given by (4.124) is reminiscent of a Schrödinger operator in a magnetic field. We can write the “effective potential” W in a more explicit form. We use the notation $\rho_s = e^{-\Phi}$; see Eq. (4.96). We have

$$\begin{aligned}\rho_s^{1/2} \mathcal{S} \rho_s^{-1/2} &= \frac{1}{2} e^{\Phi/2} \nabla \cdot \left(\Sigma e^{-\Phi} \nabla e^{\Phi/2} \right) \\ &= \frac{1}{4} e^{\Phi/2} \nabla \cdot \left(\Sigma e^{-\Phi/2} \nabla \Phi \right) \\ &= \frac{1}{4} \nabla \cdot (\Sigma \nabla \Phi) - \frac{1}{8} \Sigma \nabla \Phi \cdot \nabla \Phi.\end{aligned}$$

Furthermore,

$$\begin{aligned}\rho_s^{1/2} \mathcal{A} \rho_s^{-1/2} &= \rho_s^{-1/2} J_s \cdot \nabla \rho_s^{-1/2} \\ &= \frac{1}{2} \rho_s^{-1} J_s \cdot \nabla \Phi.\end{aligned}$$

We combine these calculations to obtain

$$W(x) = \frac{1}{4} \nabla \cdot (\Sigma \nabla \Phi) - \frac{1}{8} \Sigma \nabla \Phi \cdot \nabla \Phi + \frac{1}{2} \rho_s^{-1} J_s \cdot \nabla \Phi. \quad (4.127)$$

For reversible diffusions, the stationary probability current vanishes, and the operator \mathcal{H} becomes

$$\mathcal{H} = \frac{1}{2} \nabla \cdot (\Sigma \nabla \cdot) + \left(\frac{1}{4} \nabla \cdot (\Sigma \nabla \Phi) - \frac{1}{8} \Sigma \nabla \Phi \cdot \nabla \Phi \right). \quad (4.128)$$

On the other hand, for irreversible perturbations of the Smoluchowski (overdamped Langevin) dynamics, Eq. (4.119), whose generator is given by

$$\mathcal{L} = (-\nabla V + \gamma) \cdot \nabla + \Delta, \quad \nabla \cdot (\gamma e^{-V}) = 0, \quad (4.129)$$

the operator \mathcal{H} takes the form

$$\mathcal{H} = \Delta + \left(\frac{1}{2} \Delta V - \frac{1}{4} |\nabla V|^2 + \frac{1}{2} \gamma \cdot \nabla V \right) + \gamma \cdot \nabla. \quad (4.130)$$

It is important to keep in mind that the three operators $\mathcal{L}, \mathcal{L}^*, \mathcal{H}$ are defined in different function spaces, in (dense subsets of) $L^2(\mathbb{R}^d; \rho_s), L^2(\mathbb{R}^d; \rho_s^{-1}), L^2(\mathbb{R}^d)$, respectively. These operators are related through a (unitary) transformation.¹² We have already shown this for the map from \mathcal{L} to \mathcal{H} , Eq. (4.123). Define the multiplication operator

$$U_{\mathcal{L}, \mathcal{H}} = \sqrt{\rho_s} : L^2(\mathbb{R}^d; \rho_s) \mapsto L^2(\mathbb{R}^d).$$

This is a unitary operator from $L^2(\rho_\beta)$ to $L^2(\mathbb{R}^d)$:

$$\langle U_{\mathcal{L}, \mathcal{H}} f, U_{\mathcal{L}, \mathcal{H}} h \rangle_{L^2} = \langle f, h \rangle_\rho \quad \forall f, h \in L^2(\mathbb{R}^d; \rho_s).$$

We can then rewrite (4.123) in the form

$$\mathcal{H} = U_{\mathcal{L}, \mathcal{H}} \mathcal{L} U_{\mathcal{L}, \mathcal{H}}^{-1}.$$

The generator and Fokker–Planck operators are also unitarily equivalent, up to a sign change. We denote the generator defined in (4.121) by \mathcal{L}_{J_s} to emphasize the dependence on the stationary flux J_s . We use (4.121) and (4.122) to calculate, for every $f \in L^2(\mathbb{R}^d; \rho_s^{-1})$,

$$\begin{aligned} \rho_s \mathcal{L}_{J_s}(\rho_s^{-1} f) &= \rho_s(\mathcal{S} + \mathcal{A})(\rho_s^{-1} f) \\ &= (\mathcal{S}^* - \mathcal{A}^*) f =: \mathcal{L}_{-J_s}^* f. \end{aligned}$$

Introducing then the unitary multiplication operator

$$U_{\mathcal{L}, \mathcal{L}^*} = \rho_s : L^2(\mathbb{R}^d; \rho_s) \mapsto L^2(\mathbb{R}^d; \rho_s^{-1}),$$

¹² Two operators A_1, A_2 defined in two Hilbert spaces H_1, H_2 with inner products $\langle \cdot, \cdot \rangle_{H_1}, \langle \cdot, \cdot \rangle_{H_2}$, respectively, are called unitarily equivalent if there exists a unitary transformation $U : H_1 \mapsto H_2$ (i.e., $\langle Uf, Uh \rangle_{H_2} = \langle f, h \rangle_{H_1}, \forall f, h \in H_1$) such that

$$U A_1 U^{-1} = A_2.$$

When the operators A_1, A_2 are unbounded, we need to be more careful with their domain of definition.

we can write

$$U_{\mathcal{L}, \mathcal{L}^*} \mathcal{L}_{J_s} U_{\mathcal{L}, \mathcal{L}^*}^{-1} = \mathcal{L}_{-J_s}^*.$$

It is important to note that under this transformation, the symmetric part of the generator (corresponding to the reversible part of the dynamics) is mapped to the symmetric part of the Fokker–Planck operator, whereas the antisymmetric part (corresponding to the irreversible part of the dynamics) is mapped to minus the antisymmetric part of the Fokker–Planck operator. As an example, consider the generator of the Smoluchowski dynamics perturbed by the divergence-free vector field γ . We have [see Eq. (4.129)]

$$\mathcal{L}_{-\gamma} = (-\nabla V - \gamma) \cdot \nabla + \Delta, \quad \nabla \cdot (\gamma e^{-V}) = 0,$$

and similarly for the corresponding Fokker–Planck operator:

$$\mathcal{L}_\gamma^* = \nabla \cdot (\nabla V - \gamma + \nabla) \quad (4.131)$$

and

$$\mathcal{L}_{-\gamma}^* = \nabla \cdot (\nabla V + \gamma + \nabla)$$

We summarize these calculations in the following proposition.

Proposition 4.7. *The operators \mathcal{L}_{J_s} , $\mathcal{L}_{-J_s}^*$, and \mathcal{H} defined on $L^2(\mathbb{R}^d; \rho_s)$, $L^2(\mathbb{R}^d; \rho_s^{-1})$, and $L^2(\mathbb{R}^d)$, respectively, are unitarily equivalent:*

$$\rho_s \mathcal{L}_{J_s} \rho_s^{-1} = \mathcal{L}_{-J_s}^*, \quad (4.132a)$$

$$\sqrt{\rho_s} \mathcal{L}_{J_s} \sqrt{\rho_s^{-1}} = \mathcal{H}, \quad (4.132b)$$

$$\sqrt{\rho_s^{-1}} \mathcal{L}_{-J_s}^* \sqrt{\rho_s} = \mathcal{H}. \quad (4.132c)$$

This proposition is presented graphically in Fig. 4.1.

For reversible diffusions, i.e., when the stationary probability flux vanishes, no sign reversal is needed, and the generator, the Fokker–Planck operator, and the corresponding Schrödinger-like operator \mathcal{H} are unitarily equivalent. Consider now the case $J_s = 0$ and assume that the generalized potential Φ is such that the generator \mathcal{L} has a spectral gap, that is, Eq. (4.102) holds. Then the generator has a discrete nonnegative spectrum, and its eigenfunctions span $L^2(\mathbb{R}^d; \rho_s)$; see Sect. 4.7. Then the three operators \mathcal{L} , \mathcal{L}^* , and \mathcal{H} have the same eigenvalues, and their eigenfunctions are related through a simple transformation. Indeed, let (T_i, H_i) , $i = 1, 2$, be two unitarily equivalent self-adjoint operators with discrete spectrum in the Hilbert spaces H_1, H_2 . Consider the following eigenvalue problem for T_1 :

$$T_1 \psi_k^1 = \lambda_k^1 \psi_k^1, \quad k = 1, \dots.$$

$$\begin{array}{ccc}
 \mathcal{L}_{J_s}^* & \xrightarrow{\rho_s^{-1} \mathcal{L}_{-J_s}^* \rho_s} & \mathcal{L}_{-J_s} \\
 \uparrow & & \swarrow \\
 \rho_s^{1/2} \mathcal{H}_{-J_s} \rho_s^{-1/2} & & \rho_s^{1/2} \mathcal{L}_{-J_s} \rho_s^{-1/2} \\
 & \mathcal{H}_{-J_s} &
 \end{array}$$

Fig. 4.1 Transformation between the Fokker–Planck operator (4.122), the generator (4.121), and the Schrödinger operator (4.124)

Let $U^{-1}T_2U = T_1$. Substituting this formula into the above equation and multiplying the resulting equation by U , we deduce that

$$\psi_k^2 = U\psi_k^1 \quad \text{and} \quad \lambda_k^2 = \lambda_k^1.$$

The eigenfunctions of the generator, of the Fokker–Planck operator, and of the operator \mathcal{H} , denoted by $\psi_k^{\mathcal{L}}$, $\psi_k^{\mathcal{L}^*}$, and $\psi_k^{\mathcal{H}}$, respectively, are related through the formulas

$$\psi_k^{\mathcal{L}^*} = \rho_s^{-1} \psi_k^{\mathcal{L}} \quad \psi_k^{\mathcal{H}} = \sqrt{\rho_s^{-1}} \psi_k^{\mathcal{L}}. \quad (4.133)$$

Mapping the eigenvalue problem for the Fokker–Planck operator (or the generator) to the eigenvalue problem for a Schrödinger operator is very useful, since the spectral problem for such operators is very well studied. Similarly, we can map the Fokker–Planck equation to a Schrödinger equation in imaginary time. Let us consider the Smoluchowski Fokker–Planck equation

$$\frac{\partial p}{\partial t} = \beta^{-1} \nabla \cdot (e^{-\beta V} \nabla (e^{\beta V} p)). \quad (4.134)$$

Define $\psi(x, t) = e^{\beta V/2} p(x, t)$. Then ψ solves the PDE

$$\frac{\partial \psi}{\partial t} = \beta^{-1} \Delta \psi - U(x) \psi, \quad U(x) := \frac{\beta |\nabla V|^2}{4} - \frac{\Delta V}{2}. \quad (4.135)$$

The operator \mathcal{H} can be written as the product of two first-order operators:

$$\mathcal{H} = -\beta^{-1} \mathcal{A}^* \mathcal{A}, \quad \mathcal{A} = \nabla + \frac{\beta \nabla V}{2}, \quad \mathcal{A}^* = -\nabla + \frac{\beta \nabla V}{2},$$

or

$$\mathcal{A} \cdot = e^{-\beta V/2} \nabla \left(e^{\beta V/2} \cdot \right), \quad \mathcal{A}^* \cdot = e^{-\beta V/2} \nabla \left(e^{-\beta V/2} \cdot \right).$$

4.10 Discussion and Bibliography

The proof of existence and uniqueness of classical solutions for the Fokker–Planck equation of a uniformly elliptic diffusion process with smooth drift and diffusion coefficients, Theorem 4.1, can be found in [63]. See also [226], in particular Theorem 1.1.9, for rigorous results on the backward and forward Kolmogorov equations for diffusion processes. Parabolic PDEs (in particular in bounded domains) are studied in detail in [55].

The condition that solutions to the Fokker–Planck equation not grow too fast, see Definition 4.1, is necessary to ensure uniqueness. In fact, there are infinitely many solutions of

$$\begin{aligned} \frac{\partial p}{\partial t} &= \Delta p \quad \text{in } \mathbb{R}^d \times (0, T) \\ p(x, 0) &= 0. \end{aligned}$$

Each of these solutions besides the trivial solution $p = 0$ grows very rapidly as $x \rightarrow +\infty$. More details can be found in [109, Chap. 7].

The Fokker–Planck equation is studied extensively in Risken’s monograph [206]. See also [68, 93, 220, 236]. In these references, several examples of diffusion processes whose Fokker–Planck equation can be solved analytically can be found. The connection between the Fokker–Planck equation and stochastic differential equations is presented in Chap. 5. See also [7, 64, 65].

Diffusion processes in one dimension are studied in [153]. There is a complete classification of boundaries and boundary conditions in one dimension, the *Feller classification*: the boundaries can be *regular*, *exit*, *entrance*, and *natural*. The Feller classification for one-dimensional diffusion processes can be found in [57, 113]. We particularly recommend [113, Chap. 15] for a very detailed presentation of diffusion processes in one dimension. Several examples of Fokker–Planck operators in one dimension whose spectrum can be calculated analytically and whose eigenfunctions can be expressed in terms of orthogonal polynomials are presented in [40]. The study of the Fokker–Planck operator in one dimension is closely related to the study of Sturm–Liouville problems. More information on the Sturm–Liouville problem and one-dimensional Schrödinger operators (which we obtain after the unitary transformation described in Sect. 4.9) can be found in [232, Chap. 9].

Hermite polynomials appear very frequently in applications. We can prove that the Hermite polynomials form an orthonormal basis for $L^2(\mathbb{R}^d, \rho_\beta)$ without using

the fact that they are the eigenfunctions of a symmetric operator with compact resolvent.¹³ The proof of Proposition 4.1 can be found in [227, Lemma 2.3.4].

In Sect. 4.6, we studied convergence to equilibrium for reversible diffusions using a functional-analytic approach and, in particular, the Poincaré inequality for the probability measure $Z^{-1}e^{-V}dx$. An alternative approach is the use of a *Lyapunov function* [134, 149, 150]: we will say that the function $U \in C^2(\mathbb{R}^d)$ is a Lyapunov function if

- (i) $U(x) \geq 0$ for all $x \in \mathbb{R}^d$;
- (ii) $\lim_{|x| \rightarrow +\infty} U(x) = +\infty$;
- (iii) there exist positive constants ρ and δ such that $U(x) \leq \rho e^{\delta|x|}$ and $|\nabla U(x)| \leq \rho e^{\delta|x|}$.

It is possible to show that the existence of a Lyapunov function satisfying

$$\mathcal{L}U(x) \leq -\alpha U(x) + \beta, \quad (4.136)$$

where α, β are positive constants, ensures convergence of the solution to the Fokker–Planck equation $p(x, t)$ to the unique steady state $p_s(x)$ (i.e., the solution of the stationary Fokker–Planck equation) for all initial conditions $p(x, 0)$:

$$\lim_{t \rightarrow +\infty} p(t, x) = p_s(x), \quad (4.137)$$

the convergence being in $L^1(\mathbb{R}^d)$. The Fokker–Planck equation and the corresponding probability density function are called globally asymptotically stable [150]. Lyapunov function techniques for stochastic differential equations are studied in detail in [83]. A comparison between functional-inequalities-based and Lyapunov-function-based techniques for studying convergence to equilibrium for diffusion processes is presented in [10]. A systematic use of Lyapunov functions in the study of the ergodic properties of Markov chains is presented in [167].

Dirichlet forms play an important role in the study of diffusion processes, both in finite and in infinite dimensions. Consider an ergodic diffusion process X_t with invariant measure $\mu(dx)$ and generator

$$\mathcal{L} = b(x) \cdot \nabla + \frac{1}{2} \Sigma(x) : D^2.$$

The *opérateur carré du champ*, defined, for example, on $C^2(\mathbb{R}^d) \times C^2(\mathbb{R}^d)$, is

$$\Gamma(f, g) = \mathcal{L}(fg) - f\mathcal{L}g - g\mathcal{L}f. \quad (4.138)$$

In particular,

$$\Gamma(f, f) = \mathcal{L}f^2 - 2f\mathcal{L}f = \langle \Sigma(x)\nabla f, \nabla f \rangle.$$

¹³ In fact, the Poincaré inequality for Gaussian measures can be proved using the fact that the Hermite polynomials form an orthonormal basis for $L^2(\mathbb{R}^d, \rho_\beta)$.

The Dirichlet form of the diffusion process X_t is then defined as

$$D_{\mathcal{L}}(f) = \int_{\mathbb{R}^d} \Gamma(f, f) \mu(dx). \quad (4.139)$$

Further information on Dirichlet forms and the study of diffusion processes can be found at [148].

Poincaré inequalities for probability measures is a vast subject with deep connections with the theory of Schrödinger operators and spectral theory. See [11] and the references therein. The proof of the Poincaré inequality under assumption (4.77), Theorem 4.3, can be found in [237, Theorem A.19]. As we saw in Sect. 4.5, the Poincaré inequality for the measure $\rho_\beta dx = \frac{1}{Z} e^{-\beta V} dx$ immediately implies exponentially fast convergence to equilibrium for the corresponding reversible diffusion process in the space $L^2(\mathbb{R}^d; \rho_\beta^{-1})$. However, Theorem 4.4 is not very satisfactory, since we are assuming that we are already close to equilibrium. Indeed, the assumption on the initial condition

$$\int_{\mathbb{R}^d} |\rho_0(x)|^2 \rho_\beta^{-1} < \infty$$

is very restrictive (think of the case $V = \frac{1}{2}x^2$). The function space $L^2(\mathbb{R}^d; \rho_\beta^{-1})$ in which we prove convergence is not the right space to use. Since $\rho(\cdot, t) \in L^1$, ideally we would like to prove exponentially fast convergence in $L^1(\mathbb{R}^d)$. We can prove such a result assuming that the Gibbs density ρ_β satisfies a *logarithmic Sobolev inequality* (LSI) [76]. We recall that the relative entropy between two probability densities ρ_1 and ρ_2 (more precisely, two probability measures $\rho_1(x) dx$ and $\rho_2(x) dx$) is defined as $H(\rho_1 | \rho_2) = \int \rho_1 \ln \left(\frac{\rho_1}{\rho_2} \right) dx$. We will say that the probability measure $e^{-V} dx$ satisfies a logarithmic Sobolev inequality with constant $\lambda > 0$ if for all probability measures $\rho(x) dx$,

$$H(\rho | e^{-V}) \leq \frac{1}{2\lambda} I(\rho | e^{-V}),$$

where

$$I(\rho | e^{-V}) = \int \rho \left| \nabla \left(\ln \frac{\rho}{e^{-V}} \right) \right|^2 dx$$

denotes the (relative) Fisher information. The logarithmic Sobolev inequality immediately implies exponentially fast convergence to equilibrium in relative entropy for the corresponding reversible diffusion. We have the following result (we use the notation $\rho_\beta = e^{-V}$).

Theorem 4.6. *Let ρ denote the solution of the Fokker–Planck equation (4.67) with $\beta = 1$, where the potential is smooth and satisfies the Bakry–Emery criterion (4.79). Assume that the initial conditions satisfy*

$$H(\rho_0 | \rho_\beta) < \infty.$$

Then ρ converges to ρ_β exponentially fast in relative entropy:

$$H(\rho(\cdot, t) | \rho_\beta) \leq e^{-2\lambda t} H(\rho_0 | \rho_\beta).$$

The relative entropy controls the L^1 norm:

$$\|\rho_1 - \rho_2\|_{L^1}^2 \leq 2H(\rho_1 | \rho_2).$$

This is the *Csiszar–Kullback* or *Pinsker* inequality. From this inequality and the above theorem it follows that for potentials V such that the probability measure $e^{-V} dx$ satisfies an LSI we have exponentially fast convergence in L^1 . Logarithmic Sobolev inequalities are studied in detail in [11]. The approach using relative entropy and logarithmic Sobolev inequalities to study convergence to equilibrium for the Fokker–Planck equation is presented in [6, 155]. Similar ideas can be used for studying convergence to equilibrium for other types of parabolic PDEs and kinetic equations, both linear and nonlinear. Further information can be found in [5, 43, 44]. The Bakry–Emery criterion (4.79) guarantees that the measure $e^{-V} dx$ satisfies a logarithmic Sobolev inequality with constant λ . This, in turn, implies that potentials of the form $V + v_0$, where $v_0 \in L^\infty(\mathbb{R}^d)$, also satisfy a logarithmic Sobolev inequality. This is the content of the Holley–Stroock perturbation lemma. See [155] and the references therein.

The connection between self-adjointness of the generator of an ergodic diffusion process in $L^2(\mu)$ and the gradient structure (existence of a potential function) of the drift is established in [170]. The equivalence between self-adjointness, the existence of a potential function, time-reversibility, and zero entropy production for a diffusion process is studied in detail in [106, 192]. Conditions on the drift and diffusion coefficients that ensure detailed balance are studied in [205]. Time-reversal for diffusion processes with time-dependent coefficients is studied in [84]. Consider the stochastic equation

$$dX_t = b(X_t, t) dt + \sigma(X_t, t) dW_t, \quad (4.140)$$

in \mathbb{R}^d and with $t \in (0, 1)$, where the drift and diffusion coefficients satisfy the assumptions of Theorem 3.1 (i.e., a unique strong solution exists), and assume, furthermore, that the probability density $p(t, x)$, the solution of the Fokker–Planck equation corresponding to (4.140), satisfies

$$\int_0^1 \int_{\mathcal{O}} \left[|p(x, t)|^2 + |\sigma(x, t) \cdot \nabla p(x, t)|^2 \right] dx dt < \infty, \quad (4.141)$$

for every open bounded set \mathcal{O} . Then the reversed process $\bar{X}_t = X_{1-t}$, $t \in [0, 1]$ is a Markov diffusion process satisfying the SDE

$$d\bar{X}_t = \bar{b}(\bar{X}_t, t) dt + \bar{\sigma}(\bar{X}_t, t) dW_t, \quad (4.142)$$

with

$$\bar{b}(x, t) = -b(x, 1-t) + p(x, 1-t)^{-1} \nabla \cdot (\Sigma(x, 1-t) p(1-t, x)), \quad (4.143)$$

where $\Sigma = \sigma \sigma^T$ and

$$\bar{\sigma}(x, t) = \sigma(x, 1 - t). \quad (4.144)$$

When the drift and diffusion coefficients in (4.140) are time-independent and X_t is stationary with stationary distribution $p_s(x)$, then the formulas for the drift and the diffusion coefficients become

$$\bar{b}(x) = -b(x) + p_s(x)^{-1} \nabla \cdot (\Sigma(x) p_s(x)) \quad (4.145)$$

and

$$\bar{\sigma}(x) = \sigma(x). \quad (4.146)$$

Markov chain Monte Carlo is the standard methodology for sampling from probability distributions in high-dimensional spaces [45, 143]. Usually, the stochastic dynamics are combined with an accept–reject (Metropolis–Hastings) step. When the Smoluchowski (overdamped Langevin) dynamics are combined with the Metropolis–Hastings step, the resulting algorithm is called the Metropolis adjusted Langevin algorithm (MALA) [207–209].

In Sect. 4.8, we saw that there are (infinitely many) different diffusion processes that can be used to sample from a given probability distribution $\pi(x)$. Choosing the diffusion process that converges the fastest to equilibrium leads to a computationally efficient algorithm. The Smoluchowski dynamics is not the optimal choice, since it can lead to a slow convergence to the target distribution. The drift vector and/or the diffusion matrix have to be modified in order to accelerate convergence. It turns out that the addition of an irreversible perturbation to the dynamics will in general speed up convergence to equilibrium; see [95, 96]. The optimal irreversible perturbation can be calculated for diffusions with linear drift, which can be used to sample from Gaussian distributions. See [137]. Introducing a time-dependent temperature can also accelerate convergence to the target distribution. This is related to the *simulated annealing* algorithm [90]. Another quantity of interest is the asymptotic variance σ_f^2 for an observable f , that we can take to be mean zero. We can show that

$$\sigma_f^2 := \text{Var}(f) = \langle (-\mathcal{L})^{-1} f, f \rangle_\pi, \quad (4.147)$$

where \mathcal{L} denotes the generator of the dynamics, π the distribution from which we want to sample, and $\langle \cdot, \cdot \rangle_\pi$ the inner product in $L^2(\mathbb{R}^d; \pi)$. It is possible to use techniques from the spectral theory of operators to study σ_f^2 . See [168] and the references therein. A detailed analysis of algorithms for sampling from the Gibbs distribution $\frac{1}{Z} e^{-\beta V}$ can be found in [138].

Mapping a Fokker–Planck operator to a Schrödinger operator is very useful, since Schrödinger operators are one of the most studied topics in mathematical physics; see, e.g., [194]. For example, the algebraic study of the spectrum of the generator of the Ornstein–Uhlenbeck process using creation and annihilation operators is a standard tool in quantum mechanics. See [232, Chap. 8]. In addition, semigroups generated by Schrödinger operators can be used to study properties of the corresponding Markov semigroup; see [218].

Conversely, it is possible to express the solution of the time-dependent Schrödinger equation in terms of the solution to an appropriate Fokker–Planck equation. This is the basis for Nelson’s *stochastic mechanics*. See [32, 172].

4.11 Exercises

1. Solve Eq. (4.18) by taking the Fourier transform, using the method of characteristics for first-order PDEs and taking the inverse Fourier transform.
2. Use (4.26) to obtain formulas for the moments of the Ornstein–Uhlenbeck process. Prove, using these formulas, that the moments of the Ornstein–Uhlenbeck process converge to their equilibrium values exponentially fast.
3. Show that the autocorrelation function of the stationary Ornstein–Uhlenbeck process is

$$\begin{aligned}\mathbb{E}(X_t X_0) &= \int_{\mathbb{R}} \int_{\mathbb{R}} x x_0 p_{OU}(x, t | x_0, 0) p_s(x_0) dx dx_0 \\ &= \frac{D}{2\alpha} e^{-\alpha|t|},\end{aligned}$$

where $p_{OU}(x, t | x_0, 0)$ denotes the transition probability function and $p_s(x)$ the invariant Gaussian distribution.

4. Let X_t be a one-dimensional diffusion process with drift and diffusion coefficients $a(y, t) = -a_0 - a_1 y$ and $b(y, t) = b_0 + b_1 y + b_2 y^2$, where $a_i, b_i \geq 0$, $i = 0, 1, 2$.
 - (a) Write down the generator and the forward and backward Kolmogorov equations for X_t .
 - (b) Assume that X_0 is a random variable with probability density $\rho_0(x)$ that has finite moments. Use the forward Kolmogorov equation to derive a system of differential equations for the moments of X_t .
 - (c) Find the first three moments M_0, M_1, M_2 in terms of the moments of the initial distribution $\rho_0(x)$.
 - (d) Under what conditions on the coefficients $a_i, b_i \geq 0$, $i = 0, 1, 2$, is M_2 finite for all times?
5. Consider a uniformly elliptic diffusion process in $\Omega \subset \mathbb{R}^d$ with reflecting boundary conditions and generator

$$\mathcal{L} = b(x) \cdot \nabla + \frac{1}{2} \Sigma : D^2. \quad (4.148)$$

Let $p(x, t)$ denote the probability density function, i.e., the solution of the Fokker–Planck equation, and $p_s(x)$ the stationary distribution. Show that the relative entropy

$$H(t) = \int_{\Omega} p(x, t) \ln \left(\frac{p(x, t)}{p_s(x)} \right) dx$$

is nonincreasing:

$$\frac{dH}{dt} \leq 0.$$

6. Let V be a confining potential in \mathbb{R}^d , $\beta > 0$, and let $\rho_\beta(x) = Z^{-1} e^{-\beta V(x)}$. Give the definition of the Sobolev space $H^k(\mathbb{R}^d; \rho_\beta)$ for k a positive integer and study some of its basic properties.
7. Let X_t be a multidimensional diffusion process on $[0, 1]^d$ with periodic boundary conditions. The drift vector is a periodic function $a(x)$, and the diffusion matrix is $2DI$, where $D > 0$ and I is the identity matrix.
 - (a) Write down the generator and the forward and backward Kolmogorov equations for X_t .
 - (b) Assume that $a(x)$ is divergence-free ($\nabla \cdot a(x) = 0$). Show that X_t is ergodic and find the invariant distribution.
 - (c) Show that the probability density $p(x, t)$ (the solution of the forward Kolmogorov equation) converges to the invariant distribution exponentially fast in $L^2([0, 1]^d)$.
8. The Rayleigh process X_t is a diffusion process that takes values on $(0, +\infty)$ with drift and diffusion coefficients $a(x) = -ax + \frac{D}{x}$ and $b(x) = 2D$, respectively, where $a, D > 0$.
 - (a) Write down the generator and the forward and backward Kolmogorov equations for X_t .
 - (b) Show that this process is ergodic and find its invariant distribution.
 - (c) Solve the forward Kolmogorov (Fokker–Planck) equation using separation of variables. (*Hint:* Use Laguerre polynomials).
9. Let $\mathbf{x}(t) = \{x(t), y(t)\}$ be the two-dimensional diffusion process on $[0, 2\pi]^2$ with periodic boundary conditions with drift vector $\mathbf{a}(x, y) = (\sin(y), \sin(x))$ and diffusion matrix $\mathbf{b}(x, y)$ with $b_{11} = b_{22} = 1$, $b_{12} = b_{21} = 0$.
 - (a) Write down the generator of the process $\{x(t), y(t)\}$ and the forward and backward Kolmogorov equations.
 - (b) Show that the constant function

$$\rho_s(x, y) = C$$

is the unique stationary distribution of the process $\{x(t), y(t)\}$ and calculate the normalization constant.

- (c) Let \mathbb{E} denote the expectation with respect to the invariant distribution $\rho_s(x, y)$. Calculate

$$\mathbb{E}(\cos(x) + \cos(y)) \quad \text{and} \quad \mathbb{E}(\sin(x) \sin(y)).$$

10. Let a, D be positive constants and let $X(t)$ be the diffusion process on $[0, 1]$ with periodic boundary conditions and with drift and diffusion coefficients $a(x) = a$ and $b(x) = 2D$, respectively. Assume that the process starts at x_0 , $X(0) = x_0$.
- Write down the generator of the process $X(t)$ and the forward and backward Kolmogorov equations.
 - Solve the initial/boundary value problem for the forward Kolmogorov equation to calculate the transition probability density $p(x, t|x_0, 0)$.
 - Show that the process is ergodic and calculate the invariant distribution $p_s(x)$.
 - Calculate the stationary autocorrelation function

$$\mathbb{E}(X(t)X(0)) = \int_0^1 \int_0^1 x x_0 p(x, t|x_0, 0) p_s(x_0) dx dx_0.$$

11. Prove formulas (4.107), (4.110) and (4.111).
 12. Let X_t be a reversible diffusion process. Use the spectral analysis from Sect. 4.7 to obtain a spectral representation for an autocorrelation function of the form

$$\mathbb{E}(f(X_t)h(X_0)), \quad (4.149)$$

where f and h are arbitrary observables.

Chapter 5

Modeling with Stochastic Differential Equations

In this chapter, we discuss some of the issues that arise in modeling with stochastic differential equations, and we develop techniques for studying such equations. Colored noise and the derivation of the Stratonovich stochastic differential equation are discussed in Sect. 5.1. A brief overview of numerical methods for SDEs is presented in Sect. 5.2. Parameter estimation for SDEs is discussed in Sect. 5.3. Noise-induced transitions are studied in Sect. 5.4. Bibliographical remarks and exercises can be found in Sects. 5.5 and 5.6, respectively.

5.1 Colored Noise and Stratonovich SDEs

When the white noise in a stochastic differential equation is approximated by a smoother process, then in the limit as we remove the regularization, we obtain the Stratonovich stochastic equation.¹ This is usually called the Wong–Zakai theorem. In this section, we derive the limiting Stratonovich SDE for a particular class of regularization of the white noise process using singular perturbation theory for Markov processes. In particular, we consider *colored noise*, which we model as a Gaussian stationary diffusion process, i.e., the Ornstein–Uhlenbeck process.

We begin by studying a one-dimensional example. Consider the equations

$$\frac{dx_t}{dt} = h(x_t) + \frac{1}{\varepsilon} f(x_t) y_t, \quad (5.1a)$$

$$dy_t = -\frac{\alpha y_t}{\varepsilon^2} dt + \sqrt{\frac{2D}{\varepsilon^2}} dW_t, \quad (5.1b)$$

¹ At least in one dimension; in higher dimensions, there might be an additional drift term to the Stratonovich stochastic equation; see (5.15) and (5.16).

with W_t a standard one-dimensional Brownian motion. We say that the process x_t is driven by *colored noise*: the noise that appears in (5.1a) has nonzero correlation time. Assume that y_t is stationary, $y_0 \sim \mathcal{N}(0, D/\alpha)$. The correlation function of the colored noise $\eta_t := y_t/\varepsilon$ is (see Example 1.5)

$$R(t) = \mathbb{E}(\eta_t \eta_s) = \frac{1}{\varepsilon^2} \frac{D}{\alpha} e^{-\frac{\alpha}{\varepsilon^2} |t-s|}.$$

The power spectrum of η_t is

$$\begin{aligned} f^\varepsilon(x) &= \frac{1}{\varepsilon^2} \frac{D\varepsilon^{-2}}{\pi} \frac{1}{x^2 + (\alpha\varepsilon^{-2})^2} \\ &= \frac{D}{\pi} \frac{1}{\varepsilon^4 x^2 + \alpha^2} \rightarrow \frac{D}{\pi\alpha^2}, \end{aligned}$$

and consequently,

$$\lim_{\varepsilon \rightarrow 0} R(t) = \frac{2D}{\alpha^2} \delta(t).$$

Thus, in the limit as ε tends to 0, the colored noise process η_t converges to white noise. We can study the limit of (5.1) in the limit as $\varepsilon \rightarrow 0$.

Result 5.1 *Assume that y_t is stationary and that $f \in C^1(\mathbb{R})$. Then the solution of Eq. (5.1a) converges weakly, in the limit as $\varepsilon \rightarrow 0$, to the solution of the Stratonovich SDE*

$$dX_t = h(X_t) dt + \sqrt{\frac{2D}{\alpha^2}} f(X_t) \circ dW_t. \quad (5.2)$$

Proof. The generator of the process (x_t, y_t) is

$$\begin{aligned} \mathcal{L} &= \frac{1}{\varepsilon^2} (-\alpha y \partial_y + D \partial_y^2) + \frac{1}{\varepsilon} f(x) y \partial_x + h(x) \partial_x \\ &=: \frac{1}{\varepsilon^2} \mathcal{L}_0 + \frac{1}{\varepsilon} \mathcal{L}_1 + \mathcal{L}_2. \end{aligned}$$

The process y_t that models the noise is a stationary Markov process with invariant density

$$\rho(y) = \sqrt{\frac{\alpha}{2\pi D}} e^{-\frac{\alpha y^2}{2D}}. \quad (5.3)$$

The backward Kolmogorov equation is

$$\frac{\partial u^\varepsilon}{\partial t} = \left(\frac{1}{\varepsilon^2} \mathcal{L}_0 + \frac{1}{\varepsilon} \mathcal{L}_1 + \mathcal{L}_2 \right) u^\varepsilon. \quad (5.4)$$

We look for a solution to this equation in the form of a power series expansion in ε :

$$u^\varepsilon(x, y, t) = u_0 + \varepsilon u_1 + \varepsilon^2 u_2 + \dots,$$

with $u_j = u_j(x, y, t)$, $j = 0, 1, 2, \dots$. We substitute this into (5.4) and equate terms of the same power in ε to obtain the following hierarchy of equations:

$$-\mathcal{L}_0 u_0 = 0, \quad (5.5a)$$

$$-\mathcal{L}_0 u_1 = \mathcal{L}_1 u_0, \quad (5.5b)$$

$$-\mathcal{L}_0 u_2 = \mathcal{L}_1 u_1 + \mathcal{L}_2 u_0 - \frac{\partial u_0}{\partial t}. \quad (5.5c)$$

The ergodicity of the fast process implies that the null space of the generator \mathcal{L}_0 consists of constants in y . Hence

$$u_0 = u(x, t).$$

The second equation in the hierarchy becomes

$$-\mathcal{L}_0 u_1 = f(x) y \partial_x u.$$

This equation is solvable, since the right-hand side is orthogonal to the null space of the adjoint of \mathcal{L}_0 (this is the *Fredholm alternative*). We solve it using separation of variables:

$$u_1(x, y, t) = \frac{1}{\alpha} f(x) y \partial_x u + \psi_1(x, t).$$

In order for the third equation to have a solution, we need to require that the right-hand side be orthogonal to the null space of \mathcal{L}_0^* :

$$\int_{\mathbb{R}} \left(\mathcal{L}_1 u_1 + \mathcal{L}_2 u_0 - \frac{\partial u_0}{\partial t} \right) \rho(y) dy = 0.$$

We calculate

$$\int_{\mathbb{R}} \frac{\partial u_0}{\partial t} \rho(y) dy = \frac{\partial u}{\partial t}.$$

Furthermore,

$$\int_{\mathbb{R}} \mathcal{L}_2 u_0 \rho(y) dy = h(x) \partial_x u.$$

Finally,

$$\begin{aligned} \int_{\mathbb{R}} \mathcal{L}_1 u_1 \rho(y) dy &= \int_{\mathbb{R}} f(x) y \partial_x \left(\frac{1}{\alpha} f(x) \partial_x u y + \psi_1(x, t) \right) \rho(y) dy \\ &= \frac{1}{\alpha} f(x) \partial_x (f(x) \partial_x u) \langle y^2 \rangle + f(x) \partial_x \psi_1(x, t) \langle y \rangle \\ &= \frac{D}{\alpha^2} f(x) \partial_x (f(x) \partial_x u). \end{aligned}$$

Putting everything together, we obtain the limiting backward Kolmogorov equation

$$\frac{\partial u}{\partial t} = h(x) \partial_x u + \frac{D}{\alpha^2} f(x) \partial_x (f(x) \partial_x u).$$

The operator that appears on the right-hand side of the above equation is the generator of the Stratonovich SDE; see (3.62). Consequently, the limiting equation is the Stratonovich SDE

$$dX_t = h(X_t) dt + \sqrt{\frac{2D}{\alpha^2}} f(X_t) \circ dW_t.$$

□

Note that the above calculation leads naturally to the generator written in the form (3.62), i.e., in the form of the generator of the Stratonovich equation.

We can perform similar calculations at the level of paths, through a repeated application of Itô's formula. This approach enables us to prove a pathwise convergence result and to obtain error estimates. In particular, we can prove that

$$\mathbb{E}|x_t - X_t|^2 \leq C\epsilon^2,$$

for all $t \in [0, T]$ and for a constant C independent of ϵ . The Brownian motion that is driving the limiting SDE is the one that is driving the colored noise in (5.1b). See Exercise 4.

The situation is slightly different in higher dimensions. We consider a multidimensional version of (5.1):

$$\frac{dx_t}{dt} = h(x_t) + \frac{1}{\epsilon} F(x_t) y_t, \quad (5.6a)$$

$$\frac{dy_t}{dt} = -\frac{Ay_t}{\epsilon^2} + \frac{\sigma}{\epsilon} \frac{dW_t}{dt}, \quad (5.6b)$$

where $x : \mathbb{R}^+ \mapsto \mathbb{R}^d$, $y : \mathbb{R}^+ \mapsto \mathbb{R}^n$, $F \in (C^1(\mathbb{R}^d))^{d \times n}$, $h \in (C(\mathbb{R}^d))^d$, $A \in \mathbb{R}^{n \times n}$, $\sigma \in \mathbb{R}^{d \times n}$, and $W(t)$ is standard Brownian motion on \mathbb{R}^n . We assume that $-A$ is a stable matrix, i.e., all its eigenvalues have negative real parts and that the matrix $\Sigma := \sigma\sigma^T$ is positive definite. These assumptions imply that y has a unique invariant measure. From the results in Sect. 3.7, we know that the invariant measure is Gaussian with mean zero and covariance matrix Σ_∞ given by the steady-state variance equation (3.104). We will denote this Gaussian measure by $\mu(dy)$, and we will assume that $y_0 \sim \mu(dy)$, i.e., that y_t is stationary.

Our goal is to eliminate the variable y_t in (5.6) and to obtain a closed-form equation for x_t . We will proceed as in the proof of Result 5.1. The generator of the Markov process $\{x_t, y_t\}$ is

$$\begin{aligned} \mathcal{L} &= \frac{1}{\epsilon^2} \left(-Ay \cdot \nabla_y + \frac{1}{2} \Sigma : D^2 \right) + \frac{1}{\epsilon} F(x) y \cdot \nabla_x + h(x) \cdot \nabla_x \\ &=: \frac{1}{\epsilon^2} \mathcal{L}_0 + \frac{1}{\epsilon} \mathcal{L}_1 + \mathcal{L}_2. \end{aligned}$$

The backward Kolmogorov equation is

$$\frac{\partial u^\epsilon}{\partial t} = \left(\frac{1}{\epsilon^2} \mathcal{L}_0 + \frac{1}{\epsilon} \mathcal{L}_1 + \mathcal{L}_2 \right) u^\epsilon. \quad (5.7)$$

As in the one-dimensional case, we look for a solution to this equation in the form of a power series expansion in ε :

$$u^\varepsilon(x, y, t) = u_0 + \varepsilon u_1 + \varepsilon^2 u_2 + \dots$$

with $u_j = u_j(x, y, t)$, $j = 0, 1, 2, \dots$. We substitute this into (5.7) and equate terms of the same power in ε to obtain the following hierarchy of equations:

$$-\mathcal{L}_0 u_0 = 0, \tag{5.8a}$$

$$-\mathcal{L}_0 u_1 = \mathcal{L}_1 u_0, \tag{5.8b}$$

$$-\mathcal{L}_0 u_2 = \mathcal{L}_1 u_1 + \mathcal{L}_2 u_0 - \frac{\partial u_0}{\partial t}. \tag{5.8c}$$

Our assumptions on the ergodicity of y_t give

$$u_0 = u(x, t).$$

The second equation in the hierarchy becomes

$$-\mathcal{L}_0 u_1 = F(x) y \cdot \nabla_x u.$$

As in the one-dimensional case, we look for a solution of the form²

$$u_1(x, y, t) = B(x) y \cdot \nabla_x u.$$

We can check by inspection that

$$B(x) = F(x) A^{-1},$$

since by assumption, A is an invertible matrix. Now the solvability condition for the $\mathcal{O}(1)$ equation in (5.8) gives

$$\begin{aligned} \frac{\partial u}{\partial t} &= h(x) \cdot \nabla u + \int_{\mathbb{R}^n} F(x) y \cdot \nabla (F(x) A^{-1} y \cdot \nabla u) \mu(dy) \\ &= h(x) \cdot \nabla u + \sum_{i, \ell=1}^d \sum_{m=1}^n R_{im}(x) \frac{\partial}{\partial x_i} \left(B_{\ell m}(x) \frac{\partial u}{\partial x_\ell} \right) \\ &= h(x) \cdot \nabla u + R^T(x) : \nabla (B^T(x) \nabla u), \end{aligned} \tag{5.9}$$

where

$$R(x) = F(x) \Sigma_\infty.$$

In view of (3.62), the formula for the generator of the Stratonovich SDE (3.62), the backward Kolmogorov equation does not correspond, in general, to a Stratonovich

² We should also add a function $\psi_1(x, t)$, which belongs to the null space of \mathcal{L}_0 . However, as for the one-dimensional problem, we can check that this function does not affect the limiting backward Kolmogorov equation.

SDE. In order for the operator on the right-hand side of (3.62) to be the generator of a Stratonovich SDE, we need to impose the condition

$$R(x) = CB(x) \quad \forall x \in \mathbb{R}^d, \quad (5.10)$$

for an arbitrary constant C . A sufficient condition for (5.10) to hold is

$$\Sigma_\infty = CA^{-1}. \quad (5.11)$$

When (5.10) is not satisfied, the limiting SDE can be written as a Stratonovich (or Itô) SDE with an additional drift. The limiting SDE, corresponding to the backward Kolmogorov equation, written in Itô form, is

$$dX_t = (h(X_t) + b(X_t)) dt + \sqrt{2D^S(X_t)} dW_t, \quad (5.12)$$

where

$$D(x) = RB^T \quad \text{and} \quad b(x) = \nabla \cdot D^T - B\nabla R^T, \quad (5.13)$$

and $D^S = \frac{1}{2}(D + D^T)$ denotes the symmetric part of the matrix D .

Example 5.1. Consider the SDE (5.6) with

$$A = \alpha I + \gamma J, \quad \Sigma = \delta I,$$

for α, γ positive constants and with J the 2×2 antisymmetric matrix

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

The steady-state variance equation becomes

$$2\alpha\Sigma_\infty + \gamma[J, \Sigma_\infty] = \delta I,$$

where $[A, B] = AB - BA$ denotes the commutator between two matrices A and B . The solution to this equation is

$$\Sigma_\infty = \frac{\delta}{2\alpha} I.$$

On the other hand,

$$A^{-1} = \frac{1}{\rho} (\alpha I - \gamma J) = \frac{1}{\rho} A^T, \quad \rho = \alpha^2 + \gamma^2.$$

Consequently,

$$R = \frac{\delta}{2\alpha} F(x) \quad \text{and} \quad B(x) = \frac{\alpha}{\rho} F(x) - \frac{\gamma}{\rho} F(x) J.$$

We calculate

$$\begin{aligned} R^T(x) : \nabla (B^T(x) \nabla u) &= \frac{\delta}{2\rho} F^T(x) : \nabla (F^T(x) \nabla u) + \frac{\delta\gamma}{2\alpha\rho} F^T(x) : \nabla (JF^T(x) \nabla u) \\ &= \frac{\delta}{2\rho} F^T(x) : \nabla (F^T(x) \nabla u) \\ &\quad + \frac{\delta\gamma}{2\alpha\rho} (\nabla \cdot (FJF^T) - FJ^T \nabla \cdot F^T) \cdot \nabla u, \end{aligned} \quad (5.14)$$

where we have used the fact that the matrix $N := FJF^T$ is antisymmetric, and consequently, the Frobenius inner product of N and the Hessian vanishes. Notice that the first term on the right-hand side of (5.14) corresponds to the generator of the Stratonovich stochastic equation; see (3.62).

From the above calculations, we conclude that in the limit as the correlation time x_t of the noise process y_t/ε tends to 0, the solution of (5.6a) converges to the solution of the SDE

$$dX_t = (\mathbf{h}(X_t) + \mathbf{b}(X_t)) dt + \sqrt{\frac{\delta}{\rho}} \mathbf{F}(X_t) \circ dW_t, \quad (5.15)$$

with

$$\mathbf{b}(x) = \frac{\delta\gamma}{2\alpha\rho} (\nabla \cdot (FJF^T) - FJ^T \nabla \cdot F^T). \quad (5.16)$$

5.2 Numerical Solution of Stochastic Differential Equations

Just as with deterministic ordinary and partial differential equations, stochastic differential equations cannot be solved in closed form in general. Linear equations constitute the only general class of SDEs that can be treated analytically. It is necessary, therefore, to develop numerical methods for obtaining approximate solutions to stochastic differential equations and for calculating statistics.

We consider the scalar Itô SDE

$$dX_t = b(X_t) dt + \sigma(X_t) dW_t, \quad X_0 = x, \quad (5.17)$$

posed on the interval $[0, T]$. The initial condition can be either deterministic or random. Our goal is to obtain an approximate solution to this equation and to calculate quantities of the form $\mathbb{E} f(X_t)$, for a sufficiently large class of observables f , where \mathbb{E} denotes the expectation with respect to the law of the process X_t .

There are many numerical schemes that have been developed for studying deterministic differential equations, i.e., for $\sigma(x) \equiv 0$. The simplest such scheme is the explicit Euler method, which is based on a first-order Taylor expansion. A similar idea can be applied to the stochastic equation (5.17). We partition the interval $[0, T]$ into N equal subintervals of size Δt with $N\Delta t = T$. We use the notation

$$X_j := X(j\Delta t), \quad j = 0, 1, \dots, N.$$

The *Euler–Maruyama method* for the SDE (5.17) is

$$X_{j+1} = X_j + b(X_j) \Delta t + \sigma(X_j) \Delta W_j, \quad j = 0, \dots, N-1, \quad (5.18)$$

where $\Delta W_j = W_{j+1} - W_j$ denotes the j th Brownian increment. These Brownian increments are independent Gaussian random variables with mean zero and variance Δt . We can write $\Delta W_j = \xi_j \sqrt{\Delta t}$ with $\xi_j \sim \mathcal{N}(0, 1)$ iid. The Euler–Maruyama scheme can be written as

$$X_{j+1} = X_j + b(X_j) \Delta t + \sigma(X_j) \xi_j \sqrt{\Delta t}, \quad j = 0, \dots, N-1. \quad (5.19)$$

A random-number generator enables us to generate the iid Gaussian random variables $\{\xi_j\}_{j=0}^{N-1}$. Consequently, (5.18) provides us with an algorithm for solving (5.17).

Many different numerical methods for solving SDEs have been developed. The standard approaches from the numerical analysis of ordinary differential equations (Taylor expansions, corrector predictor methods, multistep methods) are in general applicable to the numerical analysis of SDEs, with appropriate modifications.³ Another standard numerical method for solving stochastic differential equations is *Milstein's method*:

$$X_{j+1} = X_j + b(X_j) \Delta t + \sigma(X_j) \xi_j \sqrt{\Delta t} + \frac{1}{2} \sigma(X_j) \sigma'(X_j) \Delta t (\xi_j^2 - 1). \quad (5.20)$$

Observe that the drift term is discretized in the same way in the Euler and Milstein schemes. On the other hand, the Milstein scheme has an additional term, which is related to the approximation of the stochastic term in (5.17). We can derive the Milstein scheme as follows. First, we write an increment of the solution to the SDE in the form

$$X_{j+1} = X_j + \int_{j\Delta t}^{(j+1)\Delta t} b(X_s) ds + \int_{j\Delta t}^{(j+1)\Delta t} \sigma(X_s) dW_s. \quad (5.21)$$

We apply Itô's formula to the drift and diffusion coefficients to obtain

$$b(X_s) = b(X_j) + \int_{j\Delta t}^s (\mathcal{L}b)(X_\ell) d\ell + \int_{j\Delta t}^s (b' \sigma)(X_\ell) dW_\ell$$

and

$$\sigma(X_s) = \sigma(X_j) + \int_{j\Delta t}^s (\mathcal{L}\sigma)(X_\ell) d\ell + \int_{j\Delta t}^s (\sigma' \sigma)(X_\ell) dW_\ell,$$

for $s \in [j\Delta t, (j+1)\Delta t]$, where \mathcal{L} denotes the generator of the process X_t .

³ For example, the Taylor expansion, which is the main tool for obtaining higher-order numerical methods for ODEs, has to be replaced by the stochastic Taylor expansion, which is based on Itô's formula.

We substitute these formulas into (5.21) to obtain

$$\begin{aligned}
X_{j+1} &= X_j + b(X_j) \Delta t + \sigma(X_j) \Delta W_j \\
&\quad + \int_{j\Delta t}^{(j+1)\Delta t} \int_{j\Delta t}^s (\mathcal{L}b)(X_\ell) d\ell ds + \int_{j\Delta t}^{(j+1)\Delta t} \int_{j\Delta t}^s (b' \sigma)(X_\ell) dW_\ell ds \\
&\quad + \int_{j\Delta t}^{(j+1)\Delta t} \int_{j\Delta t}^s (\mathcal{L}\sigma)(X_\ell) d\ell dW_s + \int_{j\Delta t}^{(j+1)\Delta t} \int_{j\Delta t}^s (\sigma' \sigma)(X_\ell) dW_\ell dW_s \\
&= X_j + b(X_j) \Delta t + \sigma(X_j) \Delta W_j + (\sigma' \sigma)(X_j) \int_{j\Delta t}^{(j+1)\Delta t} \int_{j\Delta t}^s dW_\ell dW_s + O(\Delta t^{3/2}) \\
&\approx X_j + b(X_j) \Delta t + \sigma(X_j) \Delta W_j + \frac{1}{2} (\sigma' \sigma)(X_j) (\Delta W_j^2 - \Delta t) \\
&= X_j + b(X_j) \Delta t + \sigma(X_j) \Delta W_j + \frac{1}{2} \Delta t (\sigma' \sigma)(X_j) (\xi_j^2 - 1),
\end{aligned}$$

with $\xi_j \sim \mathcal{N}(0, 1)$. In the above, we have used the fact that $(\Delta t)^\alpha (\Delta W_j)^\beta = O((\Delta t)^{\alpha+\beta/2})$ and that, in one dimension,

$$\int_{j\Delta t}^{(j+1)\Delta t} \int_{j\Delta t}^s dW_\ell dW_s = \frac{1}{2} (\Delta W_j^2 - \Delta t). \quad (5.22)$$

See Exercise 5.

The Euler–Mararyama and Milstein schemes provide us with the solution of an SDE only at the discretization times. The values at intermediate instants can be estimated using interpolation. We can use either constant interpolation,

$$\hat{X}(t) = X_{j_t}, \quad j_t = \max\{j = 0, 1, 2, \dots, N : j\Delta t \leq t\},$$

or linear interpolation,

$$\hat{X}(t) = \hat{X}_{j_t} + \frac{t - \tau_{j_t}}{\tau_{j_{t+1}} - \tau_{j_t}} (\hat{X}_{j_{t+1}} - \hat{X}_{j_t}),$$

where $\tau_j = j\Delta t$ and where we have used $\hat{X}(t)$ to denote the interpolated numerical solution of the SDE.

Of course, before using numerical methods of the form (5.18) or (5.20), it is necessary to analyze theoretically their performance and to study their convergence and stability properties. In particular, we need to be able to show that these schemes converge, in the limit as $\Delta t \rightarrow 0$, to the actual solution of the SDE and to calculate their convergence rates. In order to address the issue of convergence, we have to decide in what sense we want our numerical schemes to converge. Clearly, requiring pathwise convergence is much more stringent than requiring convergence of expectation values of observables. There are two main types of convergence that are used in the analysis of numerical methods for SDEs. The first is that of *strong convergence*. Let X_t denote the exact solution of the SDE (5.17) in the interval $[0, T]$. We consider a uniform partition of $[0, T]$ with $\Delta t = T/N$. Let $\hat{X}_t^{\Delta t}$ denote the numerical

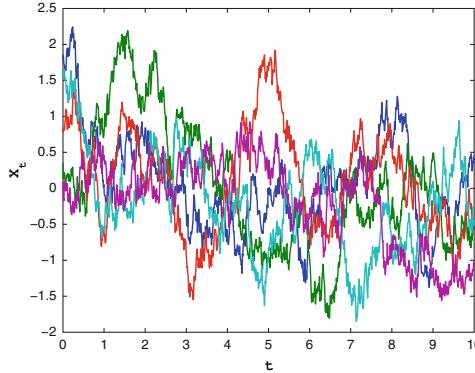


Fig. 5.1 Sample paths of the Ornstein–Uhlenbeck process

scheme; we set $X_j := X_{j\Delta t}$, and similarly for the numerical solution. We will say that a numerical scheme converges strongly if it converges in $L^1(\Omega)$, i.e.,

$$\lim_{\Delta t \rightarrow 0} \mathbb{E}|X_j - \hat{X}_j^{\Delta t}| = 0, \quad (5.23)$$

for all $j = 1, \dots, N$. Furthermore, we will say that a numerical scheme has strong order of convergence α if there exists a positive constant C independent of Δt such that

$$\mathbb{E}|X_j - \hat{X}_j^{\Delta t}| \leq C\Delta t^\alpha, \quad (5.24)$$

for every $j = 1, \dots, N$ and for Δt sufficiently small. In the above definitions, we can also consider the difference between the exact and the numerical solutions only at the final time $T = N\Delta t$. It is straightforward to extend the definition of strong convergence and of strong order of convergence to the case of a nonuniform partition of the interval $[0, T]$.

It can be shown that the Euler–Maruyama method has strong order of convergence $\frac{1}{2}$, in contrast to the deterministic case, for which the order of convergence is 1, whereas the Milstein scheme has strong order 1. The Milstein scheme has a better strong order of convergence, since it provides us with a more accurate approximation of the stochastic integral, since we are keeping the next-order term in the stochastic Taylor expansion. Not surprisingly, the strong order of convergence for the Euler–Maruyama method becomes 1 when the noise is additive. It is important to note that to study the strong convergence of a numerical scheme, the SDE and the discretized equation have to be driven by the same noise, so that we can compare between X_t and \hat{X}_t pathwise.

To demonstrate the performance of the Euler–Maruyama method, we use it to solve numerically the Ornstein–Uhlenbeck process in one dimension,

$$dX_t = -\alpha X_t dt + \sqrt{2\sigma} dW_t. \quad (5.25)$$

We solve (5.25) for $\alpha = 1$, $\sigma = \frac{1}{2}$ for $t \in [0, 10]$ with $\Delta t = 0.0098$ and initial conditions $X_0 \sim 2U(0, 1)$, where $U(0, 1)$ denotes the uniform distribution in the interval

$(0, 1)$. In Fig. 5.1, we present five sample paths of the Ornstein–Uhlenbeck process. In Fig. 5.2, we plot the first two moments of the Euler–Maruyama approximation of the Ornstein–Uhlenbeck process. We also plot the theoretical results; see (4.26).

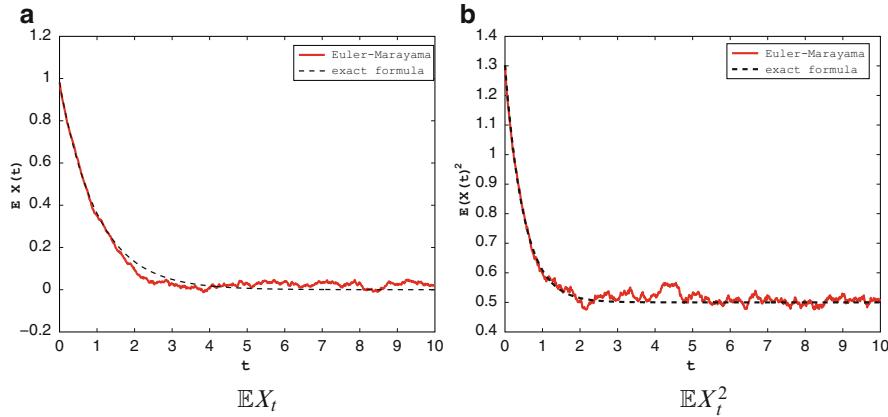


Fig. 5.2 First and second moments of the Ornstein–Uhlenbeck process using the Euler–Maruyama method. (a) $\mathbb{E}X_t$ (b) $\mathbb{E}X_t^2$

As we have already seen, we can solve (5.25) analytically:

$$X_t = e^{-\alpha t} X_0 + \sqrt{2\sigma} \int_0^t e^{-\alpha(t-s)} dW_s.$$

We use the analytical solution to check the strong order of convergence of the Euler–Maruyama scheme. We generate 1000 paths using the EM scheme as well as the exact solution, using the same Brownian paths. Since the noise in the Ornstein–Uhlenbeck SDE is additive, we expect that the strong order of convergence is 1. This is verified by our numerical experiments; see Fig. 5.3.

For the calculation of the expectation in (5.24), we need to use a Monte Carlo method for calculating the expectation. We generate N sample paths of the exact solution of the SDE and of the numerical discretization that are driven by the same noise:

$$\varepsilon = \frac{1}{M} \sum_{k=1}^M |\hat{X}_T^k - X_T^k|.$$

We can use limit theorems such as the law of large numbers and the central limit theorem to analyze the dependence of ε on M .

To estimate the strong order of convergence of a numerical scheme, we plot the error as a function of step size in log-log and use least-squares fitting:

$$\log(\text{err}) = \log C + \alpha \log \Delta t.$$

In many applications, we are interested in the calculation of statistical quantities of solutions to SDEs, rather than in pathwise properties. For such a purpose, the strong convergence (5.24) is more than what we actually need. We can introduce a different

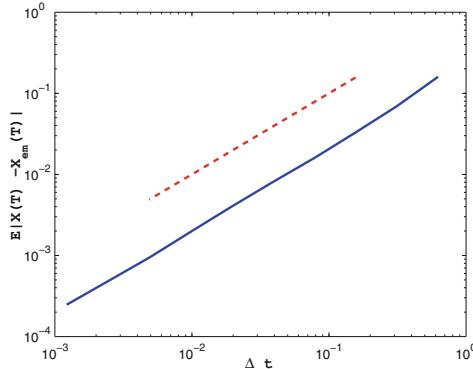


Fig. 5.3 Strong order of convergence for the Euler–Maruyama method for the Ornstein–Uhlenbeck process. The linear equation with slope 1 is plotted for comparison

concept of convergence for a numerical scheme, based on the requirement that it provide us with an accurate approximation of the expectation of a sufficiently large class of observables. Let $C_p^k(\mathbb{R})$ denote the space of all C^k functions that (together with their derivatives) grow at most polynomially at infinity. We will say that a numerical scheme $\{\hat{X}_n\}_{n=0}^N$ converges weakly to X_t if

$$\lim_{\Delta t \rightarrow 0} |\mathbb{E}f(X_j) - \mathbb{E}f(\hat{X}_j^{\Delta t})| = 0, \quad (5.26)$$

for all $j = 1, \dots, N$ and $f \in C_p^k(\mathbb{R})$ for some $k \in \mathbb{N}$. We will say that a numerical method has *weak order of convergence* β if there exists a positive constant C independent of Δt such that

$$|\mathbb{E}f(X_j) - \mathbb{E}f(\hat{X}_j^{\Delta t})| \leq C\Delta t^\beta, \quad (5.27)$$

for all $j = 1, \dots, N$, for all $f \in C_p^{2\beta+1}(\mathbb{R})$, and for Δt sufficiently small. It is possible to show that both the Euler–Maruyama and the Milstein schemes have weak order of convergence 1. If we are interested only in weak convergence, then there is no advantage in using the Milstein scheme instead of the Euler–Maruyama method.

In calculating expectation values of the form $\mathbb{E}f(X_t)$, we are introducing an additional error by approximating the expectation value. Indeed, let X be a random variable and assume that we want to calculate $Ef(X)$. To do this, we need to generate M samples and then use the formula

$$Ef(X) \approx \frac{1}{M} \sum_{j=1}^M f(X_j). \quad (5.28)$$

We will refer to this as the *Monte Carlo approximation*. An argument based on the central limit theorem provides us with the following estimate:

$$\left| Ef(X) - \frac{1}{M} \sum_{j=1}^M f(X_j) \right| \leq \frac{C}{\sqrt{M}}. \quad (5.29)$$

We note that in studying the weak order of convergence of a numerical scheme, we do not have to use the same path of Brownian motion in the SDE and in the numerical approximation. In fact, we don't even have to use a discretized Brownian motion: we can use random variables that have appropriate statistics. For example, for the Euler–Maruyama method, we can use two-point distributed random variables with $\mathbb{P}(\Delta W^j = \pm\sqrt{\Delta t}) = \frac{1}{2}$.

The weak convergence error consists of two steps, the *discretization error* and the *statistical error*, due to the Monte Carlo approximation. We estimate (\hat{X}_T denotes the solution of the numerical scheme) as follows:

$$\begin{aligned}\varepsilon &= |\mathbb{E}f(X_T) - \frac{1}{M} \sum_{m=1}^M f(\hat{X}_{N\Delta t}^m)| \\ &\leq |\mathbb{E}f(X_T) - \mathbb{E}f(\hat{X}_T)| + |\mathbb{E}f(\hat{X}_T) - \frac{1}{M} \sum_{m=1}^M f(X_{N\Delta t}^m)| \\ &\leq CN^{-\gamma} + CM^{-1/2}.\end{aligned}$$

For fixed computational resources, we can optimize over N, M .

As another example, consider the motion of a Brownian particle in a bistable potential:

$$dX_t = -V'(X_t)dt + \sqrt{2\beta^{-1}}dW_t, \quad (5.30)$$

with

$$V(x) = A\frac{x^4}{4} - B\frac{x^2}{2}. \quad (5.31)$$

Equation 5.30 will be studied in detail in Chap. 7. Here we present some illustrative numerical experiments. In Fig. 5.4a, we present a sample path of the SDE with $A = B = 1$, $\beta = 10$, obtained using the Euler–Maruyama algorithm. The transition between the two metastable states ± 1 will be studied in Chap. 7.

No analytical solution exists for this SDE. We know, however, that X_t is an ergodic Markov process with invariant measure $\mu(dx) = \frac{1}{Z}e^{-\beta V(x)}dx$, with Z denoting the normalization constant (partition function). Furthermore, from Theorem 4.4, we know that the convergence to the equilibrium distribution is exponentially fast. As a simple illustration of this fact, we start (5.30) with initial conditions distributed uniformly over $(0, 1)$, and we calculate the evolution of the second moment. We set $\alpha = \beta = 1$, $\beta = 2$. We use $\Delta t = 0.01$, and we generate 1000 paths. The results are plotted in Fig. 5.4b. The convergence of the second moment to a constant value happens very quickly. It is not difficult to calculate numerically the equilibrium value of the second moment:

$$\langle X_t^2 \rangle_{eq} = \int_{\mathbb{R}} x^2 \mu(dx) = 0.8935.$$

The agreement between the result of this calculation and the result of the Monte Carlo simulations using the Euler–Maruyama scheme is, as expected, very good.

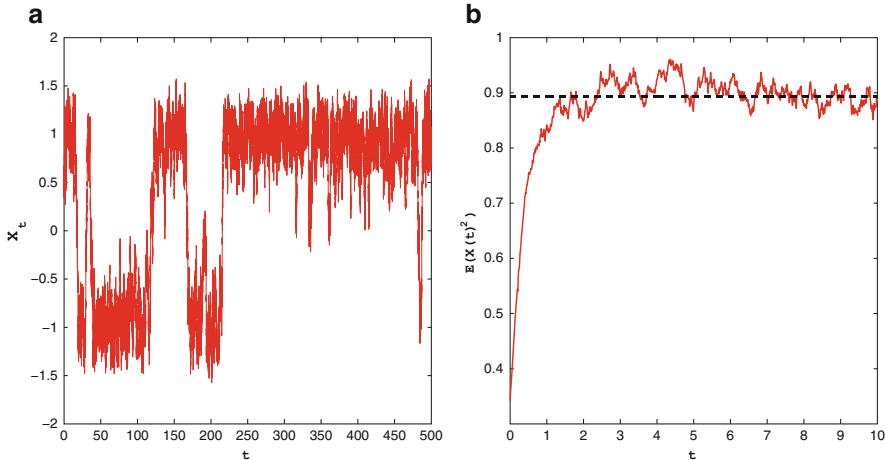


Fig. 5.4 Sample path and second moment of the bistable SDE (5.30). **(a)** Sample path. **(b)** Second moment

To recapitulate: suppose we want to compute the expectation value of an observable $f(\cdot)$ of a solution to (5.17):

$$\mathbb{E}f(X_t) = \int_{\mathbb{R}} f(x)p(x,t) dt, \quad (5.32)$$

where the distribution function $p(x,t)$ is the solution of the Fokker–Planck equation

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x} \left(b(x)p + \frac{1}{2} \frac{\partial}{\partial x} (\sigma^2(x)p) \right), \quad (5.33a)$$

$$p(x,0) = \rho_0(x). \quad (5.33b)$$

Then:

- We solve numerically (5.17) using the Euler–Maruyama, Milstein, or some other numerical method, for an ensemble of initial conditions.
- We compute expectations using the Monte Carlo approximation (5.28).

In this procedure, we introduce two sources of error: the error due to the numerical discretization of the SDE, which depends on the time step Δt , or equivalently on N ; and the Monte Carlo error, which depends on the number of trajectories M that we generate. The discretization error can be estimated from the weak order of convergence of the method, whereas the Monte Carlo error is given by (5.29). An accurate calculation of (5.32) requires both a sufficiently small time step, i.e., large N , and a sufficiently large number of samples M .

An alternative approach to computing (5.32) is based on the numerical solution of (5.33): We first use a numerical method for solving partial differential equations (finite differences, finite elements, spectral method, ...) and then we use a numerical integration technique for calculating (5.32). For low dimensional stochastic differential equations and, in particular, in bounded domains, this might be a computationally more efficient approach, in comparison to the Monte Carlo methodology. However, the numerical solution of a partial differential equation may become prohibitively expensive as the number of dimensions increases.

Consider now a stochastic differential equation in arbitrary dimensions:

$$dX_t = b(X_t) dt + \sigma(X_t) dW_t, \quad (5.34)$$

where $X_t : [0, T] \mapsto \mathbb{R}^d$, $b \in (C^\infty(\mathbb{R}^d))^d$, $\sigma \in (C^\infty(\mathbb{R}^d))^{d \times d}$, W_t is standard d -dimensional Brownian motion. It is straightforward to write the Euler–Maruyama scheme. Using the notation $b_n^i = b^i(X_n)$ for the i th component of the vector b and similarly for the diffusion matrix, we have

$$X_{n+1}^i = X_n^i + b_n^i \Delta t + \sum_{j=1}^d \sigma_n^{ij} \Delta W_n^j, \quad i = 1, \dots, d, n = 0, \dots, N-1. \quad (5.35)$$

The implementation of the Euler–Maruyama scheme in higher dimensions is similar to the one-dimensional problem.

On the other hand, for the Milstein scheme, we have

$$X_{n+1}^i = X_n^i + b_n^i \Delta t + \sum_{j=1}^d \sigma_n^{ij} \Delta W_n^j + \sum_{j_1, j_2, \ell=1}^d \sigma_n^{j\ell} \frac{\partial}{\partial x_\ell} \sigma_n^{j_1 j_2} I_{j_1 j_2}, \\ i = 1, \dots, d, n = 0, \dots, N-1, \quad (5.36)$$

where $I_{j_1 j_2}$ denotes the double Itô integral (compare with (5.22)):

$$I_{j_1 j_2} = \int_{t_n}^{t_{n+1}} \int_{s_1}^{s_2} dW_{s_2}^{j_1} dW_{s_1}^{j_2}. \quad (5.37)$$

The diagonal terms $I_{j_1 j_2}$ are given by the one-dimensional formula.

Example 5.2. Quite often, we can exploit the structure of the SDE to simplify the calculation of the Milstein correction to the EM scheme. Consider, for example, a nonlinear oscillator with friction and noise

$$\ddot{X}_t = b(X_t) - \gamma(X_t) \dot{X}_t + \sum_{j=1}^d \sigma^j(X_t) \dot{W}^j. \quad (5.38)$$

This equation is of the form of the Langevin equation (with multiplicative noise) that we will study in Chap. 6. We write it as a first-order system:

$$dX_t^1 = X_t^2 dt,$$

$$dX_t^2 = (b(X_t^1) - \gamma(X_t^1)X_t^2) dt + \sum_{j=1}^d \sigma^j(X_t^1) dW^j.$$

The Milstein discretization for this equation is

$$Y_{n+1}^1 = Y_n^1 + Y_n^2 \Delta t, \quad (5.39a)$$

$$Y_{n+1}^2 = Y_n^2 - (b(Y_n^1) - \gamma(Y_n^1)Y_n^2) \Delta t + \sum_{j=1}^d \sigma^j(Y_n^1) \Delta W^j. \quad (5.39b)$$

In fact, no Milstein correction appears in this case. This is due to the fact that noise appears only in the equation for X_t^2 , whereas the diffusion coefficient depends only on X_t^1 , which is differentiable. This scheme has strong order of convergence 1.

We can rewrite (5.39) as a multistep scheme: solve the first equation for Y_n^2 and substitute into the second to obtain a numerical method for Y_n^1 , without Y_n^2 :

$$Y_{n+2}^1 = (2 - \gamma(Y_n^1) \Delta t) Y_{n+1}^1 - (1 - \gamma(Y_n^1) \Delta t) Y_n^1 + b(Y_n^1) \Delta t^2 + \sum_{j=1}^d \sigma^j(Y_n^1) \Delta W_n^j \Delta t. \quad (5.40)$$

This is called the *Lepingle–Ribemont* method. The first equation in the Milstein scheme can be used as a starting routine and for calculating Y_n^2 .

5.2.1 Numerical Stability and Implicit Schemes

It is well known that there are many ordinary differential equations (stiff ODEs) for which explicit numerical schemes are inappropriate, since they can be unstable unless the time step is taken to be sufficiently, and perhaps computationally prohibitively, small. A similar problem arises for many types of stochastic differential equations. For such (stiff) stochastic differential equations, it is often necessary to use implicit numerical schemes.

The stability properties of numerical schemes are usually analyzed by applying them to simple test problems, in particular linear stochastic differential equations. In particular, we analyze the stability of a given numerical scheme by applying it to the SDE for geometric Brownian motion

$$dX_t = \lambda X_t dt + \sigma X_t dW_t, \quad X(0) = X_0, \quad (5.41)$$

where $\lambda, \sigma \in \mathbb{C}$. We are interested in the long-time behavior of the numerical scheme for a fixed time step Δt . In particular, we want to compare the long-time behavior of the numerical scheme to that of the solution to the SDE. A desirable

property of a numerical scheme is that for appropriate choices of the step size Δt , it has the same stability properties as the solution of the SDE.

We will say that an SDE is *stable in mean square* if for all $X_0 \neq 0$, with probability 1,

$$\lim_{t \rightarrow +\infty} \mathbb{E}|X_t|^2 = 0. \quad (5.42)$$

For the geometric Brownian motion (5.41), we have

$$\lim_{t \rightarrow +\infty} \mathbb{E}|X_t|^2 = 0 \Leftrightarrow \operatorname{Re}(\lambda) + \frac{1}{2}|\sigma|^2 < 0. \quad (5.43)$$

This follows from the calculation that leads to (4.29), but for λ and σ complex-valued.

Suppose that the parameters λ and σ are chosen so that the solution to (5.41) is mean-square stable. We want to find the values of Δt for which a numerical scheme such as the Euler or Milstein method is also mean-square stable. The numerical approximation $\{X_n^{\Delta t}\}_{n=0}^N$ is mean-square stable when for fixed Δt , we have

$$\lim_{n \rightarrow +\infty} \mathbb{E}|X_n^{\Delta t}|^2 = 0. \quad (5.44)$$

The numerical schemes that we consider in this section, when applied to the test problem (5.41), can be written in the form

$$X_{n+1}^{\Delta t} = (a(\Delta t, \lambda, \sigma) + b(\Delta t, \lambda, \sigma) \xi_n) X_n^{\Delta t}, \quad (5.45)$$

where $\xi_n \sim \mathcal{N}(0, 1)$ is iid, for appropriate coefficients a and b . By taking the square and calculating the expectation we conclude that the numerical scheme is mean-square stable, i.e., (5.44) is satisfied, if and only if

$$|a(\Delta t, \lambda, \sigma)|^2 + |b(\Delta t, \lambda, \sigma)|^2 < 1. \quad (5.46)$$

This equation will, in general, impose restrictions on the step size Δt . If the numerical scheme is mean-square stable whenever the SDE is mean-square stable for all step sizes Δt , we will say that the numerical scheme is *A-stable*.

Quite often, stability considerations impose severe restrictions on the time step Δt that we can use in the Euler–Maruyama or Milstein scheme. In such cases, it is necessary to use an *implicit scheme*. Just as with deterministic differential equations, implicit schemes have better stability properties than explicit schemes: (5.46) is satisfied for larger step sizes in comparison to the explicit scheme.

The implicit Euler scheme is

$$X_{n+1} = X_n + b(X_{n+1}) \Delta t + \sigma(X_n) \Delta W_n. \quad (5.47)$$

Notice that we are treating the drift only implicitly and the noise explicitly. Just as with deterministic differential equations, we can extend the implicit Euler scheme by considering a family of implicit Euler schemes:

$$X_{n+1} = X_n + (\theta b(X_{n+1}) + (1 - \theta)b(X_n)) \Delta t + \sigma(X_n) \Delta W_n, \quad (5.48)$$

for $\theta \in [0, 1]$. We will refer to this scheme as the *stochastic theta method*. We can analyze the mean-square stability of the stochastic theta method by checking for what step sizes (5.46) is satisfied. It is not hard to check that the stochastic theta method is A-stable for $\theta \in [\frac{1}{2}, 1]$. On the other hand, when $\theta \in [0, \frac{1}{2})$, the stochastic theta method is mean-square stable for step sizes that satisfy

$$\Delta t < \frac{-2(\operatorname{Re}(\lambda) + \frac{1}{2}|\mu|^2)}{|\lambda|^2(1 - 2\theta)}.$$

Similarly, we can introduce a family of implicit Milstein schemes:

$$X_{n+1} = X_n + (\theta b(X_{n+1}) + (1 - \theta)b(X_n)) \Delta t \quad (5.49)$$

$$+ \sigma(X_n) \Delta W_n + \frac{1}{2}(\sigma\sigma')(X_n)((\Delta W_n)^2 - \Delta t). \quad (5.50)$$

When applied to nonlinear SDEs, the implementation of an implicit scheme requires the solution of an additional algebraic equation at each time step, which can usually be done using the Newton–Raphson algorithm.

5.3 Parameter Estimation for Stochastic Differential Equations

Many of the stochastic models that are used in applications include unknown parameters that have to be determined from observations. In this section, we present some basic techniques for estimating the diffusion coefficient and parameters in the drift of SDEs. We will focus on the one-dimensional case, although the estimation techniques that we discuss can be applied to multidimensional problems.

We will consider one-dimensional Itô SDEs of the form

$$dX_t = b(X_t; \theta) dt + \sigma(X_t; \theta) dW_t, \quad X_0 = x, \quad (5.51)$$

where $\theta \in \Theta \subset \mathbb{R}^N$ is a finite set of parameters that we want to estimate from observations. The initial conditions in (5.51) can be taken to be either deterministic or random. We can consider either the case that only discrete observations are available or the case that an entire path X_t , $t \in [0, T]$, is observed. The length of the path can be fixed, or we can consider the case in which the observation interval increases, $T \rightarrow +\infty$.

We can consider the following examples of diffusion models with unknown parameters that we want to estimate from observations:

- The Ornstein–Uhlenbeck process with unknown drift coefficient α ,

$$dX_t = -\alpha X_t dt + dW_t.$$

- Brownian motion in a bistable potential, with unknown parameters A, B :

$$dX_t = (AX_t - BX_t^3) dt + dW_t.$$

- The Landau–Stuart equation with additive and multiplicative noise

$$dX_t = (AX_t - BX_t^3) dt + \sqrt{\sigma_a^2 + \sigma_b^2 X_t^2} dW_t. \quad (5.52)$$

The unknown parameters are $\theta = (A, B, \sigma_a, \sigma_b)$.

- The Heston model for option pricing with stochastic volatility:

$$\begin{aligned} dX_t &= \mu X_t dt + \sqrt{v_t} X_t dW_t^1, \\ dv_t &= \kappa(\theta - v_t) dt + \sigma \sqrt{v_t} dW_t^2, \end{aligned}$$

where W_t^1, W_t^2 are correlated with correlation coefficient ρ , and the parameters κ, θ, σ are given. The integrated stochastic volatility that we want to estimate from data is

$$\sigma(T) = \int_0^T v_t X_t^2 dt.$$

In order to estimate parameters in the diffusion coefficient, we can use the *quadratic variation* of the solution X_t to the SDE (5.51):

$$\langle X_t, X_t \rangle := \int_0^t \sigma^2(X_s; \theta) ds = \lim_{\Delta t_k \rightarrow 0} \sum_{t_k \leq t} |X_{t_{k+1}} - X_{t_k}|^2, \quad (5.53)$$

where the limit is in probability. When the diffusion coefficient is constant, $\sigma(x; \theta) \equiv \sigma$, which is the case that we will focus on, the convergence in (5.53) becomes almost sure:

$$\lim_{n \rightarrow +\infty} \sum_{i=1}^n [X_{iT2^{-n}} - X_{(i-1)T2^{-n}}]^2 = \sigma^2 T, \quad \text{a.s.} \quad (5.54)$$

In other words, if we fix the length of the observation $[0, T]$ and we let the number of observations become infinite, $n \rightarrow +\infty$, we can, in fact, *determine* (not only estimate) the diffusion coefficient. This is called the *high-frequency limit*. We also note that T can be (arbitrarily) small and that for the estimation of the diffusion coefficient, we do not need to assume that the process X_t is stationary. This is in contrast to the estimation of coefficients in the drift, where, as we will see later on, we will need to take the limit $T \rightarrow +\infty$ and also assume stationarity.

Rather than proving (5.54), we state and prove a much simpler result that will be sufficient for our purposes.

Proposition 5.2. *Let $\{X_j\}_{j=0}^J$ be a sequence of equidistant observations of*

$$dX_t = b(X_t; \theta) dt + \sigma dW_t$$

with time step $\Delta t = \delta$ and $J\delta = T$ fixed. Assume that the drift $b(x; \theta)$ is bounded, and define

$$\widehat{\sigma}_J^2 = \frac{1}{J\delta} \sum_{j=0}^{J-1} (X_{j+1} - X_j)^2. \quad (5.55)$$

Then

$$|\mathbb{E}\widehat{\sigma}_J^2 - \sigma^2| \leq C(\delta + \delta^{1/2}). \quad (5.56)$$

In particular,

$$\lim_{J \rightarrow +\infty} |\mathbb{E}\widehat{\sigma}_J^2 - \sigma^2| = 0. \quad (5.57)$$

Proof. We have

$$X_{j+1} - X_j = \int_{j\delta}^{(j+1)\delta} b(X_s; \theta) ds + \sigma \Delta W_j,$$

where $\Delta W_j = W_{(j+1)\delta} - W_{j\delta} \sim \mathcal{N}(0, \delta)$. We substitute this into (5.55) to obtain

$$\widehat{\sigma}_J^2 = \sigma^2 \frac{1}{\delta J} \sum_{j=0}^{J-1} (\Delta W_j)^2 + \frac{2}{\delta J} \sum_{j=0}^{J-1} I_j M_j + \frac{1}{\delta J} \sum_{j=0}^{J-1} I_j^2,$$

where

$$I_j := \int_{j\delta}^{(j+1)\delta} b(X_s; \theta) ds$$

and $M_j := \sigma \Delta W_j$. We note that $\mathbb{E}(\Delta W_j)^2 = \delta$. Furthermore, from the boundedness of $b(x; \theta)$ and using the Cauchy–Schwarz inequality, we get

$$\mathbb{E}I_j^2 \leq \delta \int_{j\delta}^{(j+1)\delta} \mathbb{E}(b(X_s; \theta))^2 ds \leq C\delta^2.$$

Consequently,

$$\begin{aligned} |\mathbb{E}\widehat{\sigma}_J^2 - \sigma^2| &\leq \frac{1}{\delta} \mathbb{E}I_j^2 + \frac{2}{\delta} \mathbb{E}|I_j M_j| \\ &\leq C\delta + \frac{C}{\delta} \left(\frac{1}{\alpha} \mathbb{E}I_j^2 + \alpha \mathbb{E}M_j^2 \right) \\ &\leq C(\delta + \delta^{1/2}). \end{aligned}$$

In the above, we used Cauchy's inequality with $\alpha = \delta^{1/2}$. \square

From now on, we will assume that we have already estimated the diffusion coefficient. To simplify the notation, we will set $\sigma = 1$:

$$dX_t = b(X_t; \theta) dt + dW_t. \quad (5.58)$$

Our goal now is to estimate the unknown parameters in the drift $\theta \in \Theta$ from discrete observations. We will denote the true value by θ_0 . We will use the *maximum likelihood estimator* (MLE), which is based on maximizing the *likelihood function*. To motivate the maximum likelihood approach, we consider first the case of a random variable X whose probability distribution function $f(x|\theta)$ is known up to parameters θ that we want to estimate from observations. The standard example is that of a Gaussian random variable, in which case the parameters to be estimated are the mean and variance, $\theta = (\mu, \sigma)$.

Suppose now that we have N independent observations of the random variable X . We define the *likelihood function*

$$L(\{x_i\}_{i=1}^N | \theta) = \prod_{i=1}^N f(x_i | \theta). \quad (5.59)$$

Thus, the likelihood function is essentially the probability density function of the random variable X , viewed as a function of the parameters θ . The maximum likelihood estimator (MLE) is then

$$\hat{\theta} = \operatorname{argmax} L(\mathbf{x} | \theta), \quad (5.60)$$

with $\mathbf{x} = \{x_i\}_{i=1}^N$. It should be clear that the MLE is a random variable that depends on the observations $\{x_i\}_{i=1}^N$. When X is a Gaussian random variable, $X \sim \mathcal{N}(\mu, \sigma^2)$, the likelihood function takes the form

$$L(\{x_i\}_{i=1}^N | \theta) = \left(\frac{1}{2\pi\sigma^2} \right)^{N/2} \exp \left(-\frac{\sum_{i=1}^N (x_i - \mu)^2}{2\sigma^2} \right).$$

Maximizing then with respect to μ and σ^2 , we obtain the maximum likelihood estimators (see Exercise 11)

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i, \quad \hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \hat{\mu})^2. \quad (5.61)$$

We want to use this idea to estimate the parameters in the drift of the SDE (5.58). The (independent) observations of the random variable X are now replaced by observations of the process X_t , $\{X_i\}_{i=1}^N$ with $X_i = X_{i h}$ and $hN = T$, i.e., we have N equidistant observations, or with an entire path X_t , $t \in [0, T]$.

Assume, then, that a path of the process X_t is observed. The analogue of the likelihood function (5.59) is the law of the process on the path space. From Girsanov's theorem (3.98), we know that the law of X_t , denoted by \mathbb{P}_X , is absolutely continuous with respect to the Wiener measure, the law of Brownian motion. The density of \mathbb{P}_X with respect to the Wiener measure is given by the Radon–Nikodym derivative:

$$\frac{d\mathbb{P}_X}{d\mathbb{P}_W} = \exp \left(\int_0^T b(X_s; \theta) dX_s - \frac{1}{2} \int_0^T (b(X_s; \theta))^2 ds \right) =: L(\{X_t\}_{t \in [0, T]}; \theta, T). \quad (5.62)$$

This is precisely the likelihood function. The maximum likelihood estimator MLE is defined as

$$\hat{\theta} = \operatorname{argmax}_{\theta \in \Theta} L(\{X_t\}_{t \in [0, T]}; \theta). \quad (5.63)$$

The MLE estimator is a random variable that depends on the path $\{X_t\}_{t \in [0, T]}$.

We assume that the diffusion process (5.58) is stationary. The MLE (5.63) is *asymptotically unbiased*: in the limit as the window of observation becomes infinite, $T \rightarrow +\infty$, the MLE $\hat{\theta}$ converges to the true value θ_0 . Later in this section, we will prove a particular form of this result for the case that the coefficients to be estimated appear linearly in the drift, Theorem 5.1. See also the discussion in Sect. 5.5.

Assume that there are N parameters to be estimated, $\theta = (\theta_1, \dots, \theta_N)$. The MLE is obtained by solving the (generally nonlinear) system of equations

$$\frac{\partial L}{\partial \theta_i} = 0, \quad i = 1, \dots, N. \quad (5.64)$$

We need to check, of course, that the solution of this linear system of equations corresponds to a maximum of the likelihood function. The solution of this system of equations can be expressed in terms of functionals (e.g., moments) of the observed path $\{X_t\}_{t \in [0, T]}$,

$$\hat{\theta} = \mathcal{F}(\{X_t\}_{t \in [0, T]}).$$

Example 5.3 (MLE for the stationary Ornstein–Uhlenbeck process.). Consider the stationary Ornstein–Uhlenbeck process

$$dX_t = -\alpha X_t dt + dW_t \quad (5.65)$$

with $X_0 \sim \mathcal{N}(0, \frac{1}{2\alpha})$. The *log likelihood function* is⁴

$$\log L = -\alpha \int_0^T X_t dX_t - \frac{\alpha^2}{2} \int_0^T X_t^2 dt.$$

Equation (5.64) becomes $\frac{\partial \log L}{\partial \alpha} = 0$, from which we obtain

$$\hat{\alpha} = -\frac{\int_0^T X_t dX_t}{\int_0^T X_t^2 dt} =: -\frac{B_1(\{X_t\}_{t \in [0, T]})}{M_2(\{X_t\}_{t \in [0, T]})}, \quad (5.66)$$

where we have used the notation

$$B_n(\{X_t\}_{t \in [0, T]}) := \int_0^T X_t^n dX_t, \quad M_n(\{X_t\}_{t \in [0, T]}) := \int_0^T X_t^n dt. \quad (5.67)$$

Given a set of discrete equidistant observations $\{X_j\}_{j=0}^J$, $X_j = X_{j\Delta t}$, $\Delta X_j = X_{j+1} - X_j$, formula (5.66) can be approximated by

⁴ Maximizing the likelihood function is, of course, equivalent to maximizing the log likelihood function.

$$\hat{\alpha} = -\frac{\sum_{j=0}^{J-1} X_j \Delta X_j}{\sum_{j=0}^{J-1} |X_j|^2 \Delta t}. \quad (5.68)$$

The MLE (5.66) becomes asymptotically unbiased in the *large sample limit* $J \rightarrow +\infty$, Δt fixed.

As an illustration, we estimate the drift coefficient of the Ornstein–Uhlenbeck process. We generate a long stationary trajectory using the Euler–Maruyama scheme, and then we use (5.68) to estimate α . The result of this numerical experiment is presented in Fig. 5.5. Notice that the MLE becomes unbiased after a short transient time. The results presented in this figure are obtained using only one time series. We can calculate the variance of the estimator by generating a large number of paths of the stationary Ornstein–Uhlenbeck process.

Using Itô’s formula, we can obtain an alternative formula for the maximum likelihood estimator of the drift coefficient for the Ornstein–Uhlenbeck process. First, note that the numerator in (5.66) can be written as

$$\int_0^t X_s dX_s = -\alpha \int_0^t X_s^2 ds + \int_0^t X_s dW_s.$$

We apply Itô’s formula to the function $V(x) = \frac{1}{2}x^2$ to obtain

$$dV(X_t) = -\alpha X_t^2 dt + \frac{1}{2} dt + X_t dW_t.$$

We combine the above two equations to obtain

$$\int_0^t X_s dX_s = \frac{X_t^2 - X_0^2 - t}{2}.$$

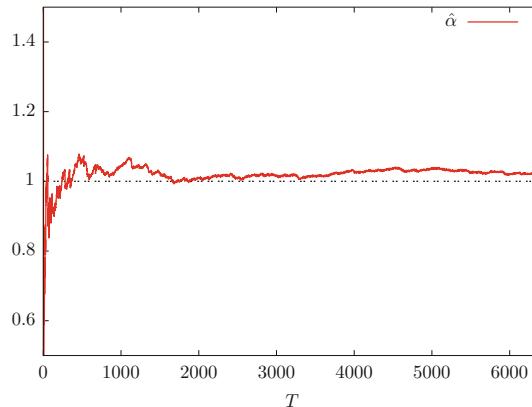


Fig. 5.5 MLE for the Ornstein–Uhlenbeck process

The formula for the MLE now becomes

$$\hat{\alpha} = -\frac{X_T^2 - X_0^2 - T}{2 \int_0^T X_t^2 dt}. \quad (5.69)$$

The advantage of (5.69) over (5.66) is that it requires the evaluation of only one integral. We can obtain a similar estimate for arbitrary potentials V that depend linearly on the parameters that we want to estimate.

Example 5.4. Consider the following generalization of the previous example:

$$dX_t = \alpha b(X_t) dt + dW_t, \quad (5.70)$$

where $b(x)$ is such that the equation has a unique ergodic solution. The log Likelihood function is

$$\log L = \alpha \int_0^T b(X_t) dX_t - \frac{\alpha^2}{2} \int_0^T b(X_t)^2 dt.$$

The MLE is

$$\hat{\alpha} = \frac{\int_0^T b(X_t) dX_t}{\int_0^T (b(X_t))^2 dt}.$$

Example 5.5 (MLE for a stationary bistable stochastic differential equation). Consider the SDE

$$dX_t = (\alpha X_t - \beta X_t^3) dt + dW_t. \quad (5.71)$$

This SDE is of the form $dX_t = -V'(X_t) dt + dW_t$ with $V(x) = -\frac{\alpha}{2}x^2 + \frac{\beta}{4}x^4$ and is ergodic with invariant distribution $\rho(x) = Z^{-1}e^{-2V(x)}$. Our goal is to estimate the coefficients α and β from observations using the maximum likelihood approach. The log likelihood functions reads

$$\begin{aligned} \log L &= \int_0^T (\alpha X_t - \beta X_t^3) dX_t - \frac{1}{2} \int_0^T (\alpha X_t - \beta X_t^3)^2 dt \\ &=: \alpha B_1 - \beta B_3 - \frac{1}{2} \alpha^2 M_2 - \frac{1}{2} \beta^2 M_6 + \alpha \beta M_4, \end{aligned}$$

using the notation (5.67). Equations (5.64) become

$$\frac{\partial \log L}{\partial \alpha}(\hat{\alpha}, \hat{\beta}) = 0, \quad \frac{\partial \log L}{\partial \beta}(\hat{\alpha}, \hat{\beta}) = 0,$$

which leads to a linear system of equations

$$\begin{pmatrix} M_2 & -M_4 \\ M_4 & -M_6 \end{pmatrix} \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \begin{pmatrix} B_1 \\ B_3 \end{pmatrix},$$

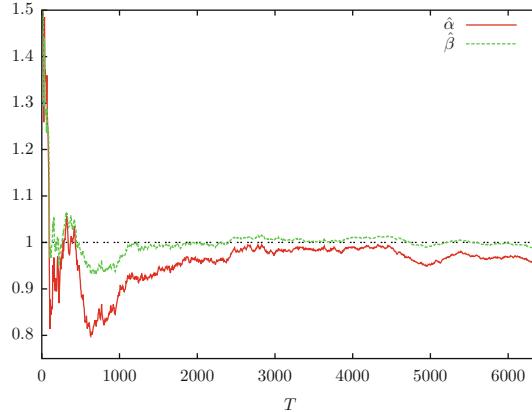


Fig. 5.6 MLE estimators for a bistable potential

the solution of which is

$$\hat{\alpha} = \frac{B_1 M_6 - B_3 M_4}{M_2 M_6 - M_4^2}, \quad \hat{\beta} = \frac{B_1 M_4 - B_3 M_2}{M_2 M_6 - M_4^2}. \quad (5.72)$$

When we have a discrete set of observations $\{X_j\}_{j=0}^{J-1}$, the integrals in (5.72) have to be approximated by sums. See Exercise 10.

As with the Ornstein–Uhlenbeck process, we estimate the coefficients in the drift by generating one long trajectory of (5.71).⁵ The results of our simulation are shown in Fig. 5.6.

As we have already mentioned, the rigorous justification of the MLE is based on Girsanov's theorem. Here we present a heuristic derivation of (5.62) that is based on the Euler–Maruyama discretization of (5.58):

$$X_{n+1} - X_n = b(X_n; \theta) \Delta t + \Delta W_n, \quad (5.73)$$

where $X_n = X(n\Delta t)$ and $\Delta W_n = W_{n+1} - W_n = \sqrt{\Delta t} \xi_n$ with $\xi_n \sim \mathcal{N}(0, 1)$. Our goal is to calculate the Radon–Nikodym derivative of the law of the discrete-time process $\{X_n\}_{n=0}^{N-1}$ and the discretized Brownian motion. In the discrete case, this derivative becomes the ratio between the distribution functions of the two processes.

We rewrite (5.73) in the form

$$\Delta X_n = b_n \Delta t + \xi_n \sqrt{\Delta t}, \quad (5.74)$$

⁵ Note, however, that this trajectory is not stationary, since the initial conditions are not distributed according to the invariant distribution $\frac{1}{Z} e^{-\beta V(x)}$. We could, in principle, sample from this distribution using the MCMC methodology that was mentioned in Sect. 4.10.

where $b_n := b(X_n; \theta)$. The distribution function of the discretized Brownian motion is

$$p_W^N = \prod_{i=0}^{N-1} \frac{1}{\sqrt{2\pi\Delta t}} \exp\left(-\frac{1}{2\Delta t}(\Delta W_i)^2\right) \quad (5.75)$$

$$= \frac{1}{(\sqrt{2\pi\Delta t})^N} \exp\left(-\frac{1}{2\Delta t} \sum_{i=0}^{N-1} (\Delta W_i)^2\right). \quad (5.76)$$

Similarly, for the law of the discretized process $\{X_n\}_{n=0}^{N-1}$, using the fact that $p(X_{i+1}|X_i) \sim \mathcal{N}(X_i + b_i\Delta t, \Delta t)$, we can write

$$p_X^N = \frac{1}{(\sqrt{2\pi\Delta t})^N} \exp\left(-\sum_{i=0}^{N-1} \left(\frac{1}{2\Delta t}(\Delta X_i)^2 + \frac{1}{2}(b_i)^2\Delta t - b_i\Delta X_i\right)\right). \quad (5.77)$$

Now we can calculate the ratio of the laws of the two processes, evaluated at the path $\{X_n\}_{n=0}^{N-1}$:

$$\frac{d\mathbb{P}_X^N}{d\mathbb{P}_W^N} = \exp\left(-\frac{1}{2} \sum_{i=0}^{N-1} (b_i)^2\Delta t + \sum_{i=0}^{N-1} b_i\Delta X_i\right).$$

Passing now (formally) to the limit as $N \rightarrow +\infty$ while keeping Δ fixed, we obtain (5.62).

It is important to note that the MLE (5.63) depends on the path $\{X_t\}_{t \in [0, T]}$ (or rather, on the discrete observations $\{X_j\}_{j=0}^{J-1}$), and consequently, it is a random variable. It is necessary to prove that in the large-sample limit $J \rightarrow +\infty$, Δt fixed, and for appropriate assumptions on the diffusion process X_t , the MLE converges to the true value θ_0 and also to obtain information about the fluctuations around θ_0 . Assuming that X_t is stationary and that the entire path $\{X_t\}_{t \in [0, T]}$ is available, we can prove that the MLE $\hat{\theta}$ converges in the limit as $T \rightarrow +\infty$ (assuming that the entire path $\{X_t\}_{t \in [0, T]}$ is available to us) to θ_0 . Furthermore, we can prove *asymptotic normality* of the maximum likelihood estimator,

$$\sqrt{T}(\hat{\theta} - \theta_0) \rightarrow \mathcal{N}(0, \sigma^2), \quad (5.78)$$

with variance

$$\sigma^2 = \left(\mathbb{E}_{\theta_0} \left(\frac{\partial b}{\partial \theta}\right)^2\right)^{-1},$$

where \mathbb{E}_{θ_0} denotes the stationary expectation of the SDE evaluated at the true value $\theta = \theta_0$. To get some intuition about the theoretical issues involved, we will prove a version of this result for a stationary Ornstein–Uhlenbeck process.

Theorem 5.1. *Let X_t be the stationary Ornstein–Uhlenbeck process*

$$dX_t = -\alpha X_t dt + dW_t, \quad X_0 \sim \mathcal{N}\left(0, \frac{1}{2\alpha}\right),$$

and let $\hat{\alpha}$ denote the MLE (5.66). Then

$$\lim_{T \rightarrow +\infty} \sqrt{T} |\hat{\alpha} - \alpha| = \mathcal{N}(0, 2\alpha) \quad (5.79)$$

in distribution.

Proof. First we observe that

$$\hat{\alpha} = -\frac{\int_0^T X_t dX_t}{\int_0^T X_t^2 dt} = \alpha - \frac{\int_0^T X_t dW_t}{\int_0^T X_t^2 dt}.$$

Consequently,

$$\begin{aligned} \hat{\alpha} - \alpha &= -\frac{\int_0^T X_t dW_t}{\int_0^T X_t^2 dt} = -\frac{1}{\sqrt{T}} \frac{\frac{1}{\sqrt{T}} \int_0^T X_t dW_t}{\frac{1}{T} \int_0^T X_t^2 dt} \\ &\stackrel{\text{Law}}{=} -\frac{1}{\sqrt{T}} \frac{W\left(\frac{1}{T} \int_0^T X_t^2 dt\right)}{\frac{1}{T} \int_0^T X_t^2 dt}, \end{aligned}$$

where the scaling property of Brownian motion was used.⁶ The process X_t is stationary. We use the ergodic theorem for stationary Markov processes to obtain

$$\lim_{T \rightarrow +\infty} \frac{1}{T} \int_0^T X_t^2 dt = \mathbb{E} X_t^2 = \frac{1}{2\alpha}, \quad (5.81)$$

almost surely. Let now $Y = \mathcal{N}(0, \sigma^2)$ with $\sigma^2 = \frac{1}{2\alpha}$. We can write $Y = W(\sigma^2)$. We use now the Hölder continuity of Brownian motion to conclude that almost surely,

$$\begin{aligned} \left| \frac{1}{\sqrt{T}} \int_0^T X_t^2 dt - Y \right| &= \left| W\left(\frac{1}{T} \int_0^T X_t^2 dt\right) - W\left(\frac{1}{2\alpha}\right) \right| \\ &\leq \text{Hö l}(W) \left| \frac{1}{T} \int_0^T X_t^2 dt - \frac{1}{2\alpha} \right|^{\frac{1}{2} - \varepsilon}, \end{aligned} \quad (5.82)$$

where $\text{Hö l}(W)$ denotes the Hölder constant of Brownian motion. We use now (5.81) to conclude that

$$\lim_{T \rightarrow +\infty} \frac{1}{T} \int_0^T X_t dW_t = \mathcal{N}\left(0, \frac{1}{2\alpha}\right), \quad (5.83)$$

⁶ We have that in law,

$$\int_0^T f(s) dW(s) = W\left(\int_0^T f^2(s) ds\right). \quad (5.80)$$

Generalizations of this formula are discussed in Sect. 5.5.

in distribution. We combine (5.81) with (5.83) and use Slutsky's theorem, Theorem B.4, to conclude that

$$\lim_{T \rightarrow +\infty} \sqrt{T} |\hat{\alpha} - \alpha| = \mathcal{N}(0, 2\alpha) \quad (5.84)$$

in distribution. \square

5.4 Noise-Induced Transitions

In this section, we study the effect of noise on the qualitative behavior of solutions of some simple one-dimensional deterministic dynamical systems. The setting that we consider is the following: We are given an ordinary differential equation

$$\frac{dX_t}{dt} = b(X_t; \lambda), \quad (5.85)$$

together with the initial condition $X_0 = x$. The ODE (5.85) depends on a parameter $\lambda \in \mathbb{R}$. The standard example is the *Landau–Stuart equation*

$$\frac{dX_t}{dt} = \lambda X_t - X_t^3, \quad X_0 = x. \quad (5.86)$$

The *bifurcation parameter* λ determines the properties of the steady-state solutions of (5.86): when $\lambda < 0$, all solutions are attracted to the single steady state $X_* = 0$; when $\lambda > 0$, the steady state $X_* = 0$ becomes unstable, and $X_t \rightarrow \sqrt{\lambda}$ if $x > 0$ and $X_t \rightarrow -\sqrt{\lambda}$ if $x < 0$.

We assume that the bifurcation diagram of (5.85) is known, and we consider the effect of stochastic perturbations on the dynamics:

$$dX_t = b(X_t; \lambda) dt + \sigma(X_t) dW_t, \quad (5.87)$$

where the stochastic integral is interpreted in the Itô sense. A similar analysis can be performed for the Stratonovich SDE. See Exercise 16. Our goal is to understand the dependence of the dynamics on the bifurcation parameter λ for the stochastic dynamics. The simplest way of doing this is by analyzing the dependence of the stationary distribution, if it exists, on λ as well as the strength of the noise σ .⁷ Our goal will be to construct the bifurcation diagram for the stationary distribution $p_s(x; \lambda)$. In analogy to the deterministic dynamics, we can think of a phase transition for the stochastic dynamics when there is a qualitative change in the stationary distribution. Consequently, the extrema of $p_s(x; \lambda)$ can be used as indicators of phase transitions. An interesting question is whether the number and position of the extrema of the

⁷ This means, of course, that we are not really looking at the dynamics but only at the dependence of the stationary state on the bifurcation parameter.

stationary distribution are the same as that of the deterministic dynamics (in other words, whether there is a one-to-one correspondence between the bifurcation diagrams of the deterministic and the stochastic dynamics) or whether the presence of noise changes the number, nature, and position of the extrema.

We begin with the case in which the noise in (5.87) is additive, corresponding to thermal fluctuations

$$dX_t = -\frac{dV}{dx}(X_t; \lambda) dt + \sqrt{2\beta^{-1}} dW_t, \quad (5.88)$$

where we have introduced the potential function

$$V(x; \lambda) = - \int^x b(x; \lambda) dx. \quad (5.89)$$

Assuming that the potential is confining (see Definition 4.2), from Proposition 4.2 we deduce that X_t is an ergodic Markov process with stationary distribution

$$p_s(x; \lambda) = \frac{1}{Z} e^{-\beta V(x; \lambda)}. \quad (5.90)$$

Thus, the presence of additive noise, even for arbitrarily small noise strength, ensures that X_t has nice ergodic properties and a unique invariant measure with a smooth density. In order to construct the bifurcation diagram, we need to analyze the critical points of the stationary distribution as a function of λ . We calculate the critical points of the stationary distribution:

$$\frac{dp_s}{dx}(x; \lambda) = 0. \quad (5.91)$$

It follows from this equation that the critical points of the stationary distribution are the same as the critical points of the deterministic dynamics:

$$-V'(x^*; \lambda) = b(x^*; \lambda) = 0. \quad (5.92)$$

Consequently, the bifurcation diagram of the stationary distribution in the presence of additive noise reflects that of the deterministic dynamics; additive noise does not induce any new bifurcation points.

As an example, consider the Landau–Stuart equation with additive noise:

$$dX_t = (\lambda X_t - X_t^3) dt + \sqrt{2\beta^{-1}} dW_t, \quad X_0 = x. \quad (5.93)$$

The process X_t is ergodic, and the stationary distribution is given by (5.90) with $V(x; \lambda) = -\frac{\lambda}{2}x^2 + \frac{1}{4}x^4$. We can read off the bifurcation diagram for the stationary distribution from that of the deterministic dynamics: when $\lambda < 0$, the stationary distribution has a single maximum centered at $x = 0$, the location of the globally stable steady state of the deterministic dynamics. On the other hand, when $\lambda > 0$, the stationary density has two maxima centered at $x = \pm\sqrt{\lambda}$, the locally stable steady

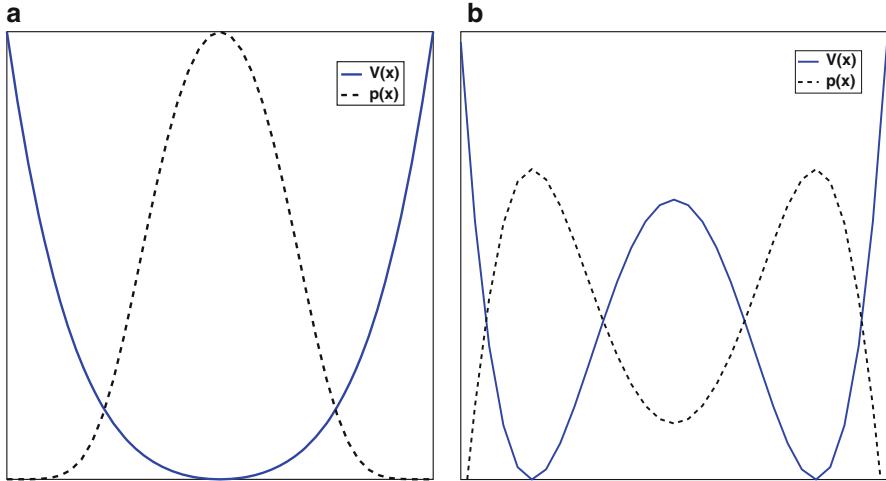


Fig. 5.7 Potential and stationary distribution of the Landau–Stuart equation for negative and positive values of the bifurcation parameter λ . (a) $\lambda < 0$. (b) $\lambda > 0$

states, and a local minimum at the unstable steady state $x = 0$. We can say that at $\lambda = 0$, we have a *phase transition*: the stationary distribution changes from unimodal to bimodal. The potential and the corresponding stationary distribution for negative and positive values of the bifurcation parameter are shown in Fig. 5.7. We observe that in the limit as the noise strength tends to 0, $\beta \rightarrow +\infty$, the invariant distribution $p_s(x; \lambda)$ converges to a δ -function centered at 0 when $\lambda < 0$, and to two δ functions centered at $\pm\sqrt{\lambda}$ when $\lambda > 0$.

From the results of Sect. 4.5, it follows that the solution of the Fokker–Planck equation corresponding to (5.93) converges (exponentially fast) to the invariant distribution.⁸

Now we consider multiplicative perturbations of the deterministic dynamics (5.85). We study Eq. (5.87) in the interval (ℓ, r) with reflecting boundary conditions. The stationary distribution, if it exists, i.e., the solution of the stationary Fokker–Planck equation with vanishing probability flux on the boundary (see (4.35)), is

$$p_s(x; \lambda) = \frac{1}{Z} \frac{1}{\sigma^2(x)} \exp \left(2 \int_{\ell}^x \frac{b(y; \lambda)}{\sigma^2(y)} dy \right),$$

$$Z = \int_{\ell}^r \left(\frac{1}{\sigma^2(x)} \exp \left(2 \int_{\ell}^x \frac{b(y; \lambda)}{\sigma^2(y)} dy \right) \right) dx. \quad (5.94)$$

We note that multiplicative noise does not necessarily lead to ergodic stochastic dynamics. This depends on whether the diffusion coefficient vanishes in the interior or the boundary of the domain $[\ell, r]$. Assuming that an invariant distribution exists,

⁸ Alternatively, we can show that $U(x) = x^2$ is a Lyapunov function; see Exercise 15.

we can study its dependence on the parameter λ and in particular, examine the number and nature of its critical points. We rewrite (5.94) in a form similar to (5.90):

$$p_s(x; \lambda) = \frac{1}{Z} \exp(-2V_{eff}(x; \lambda)), \quad V_{eff}(x; \lambda) = \ln \sigma(x) - \int_{\ell}^x \frac{b(y; \lambda)}{\sigma^2(y)} dy. \quad (5.95)$$

The critical points of the stationary distribution are given by the equation

$$b(x^*; \lambda) - \sigma(x^*)\sigma'(x^*) = 0. \quad (5.96)$$

This equation is different from (5.92). Consequently, it is possible that multiplicative noise can lead to new critical points that are not present in the deterministic dynamics. Thus, the bifurcation diagram for the stationary distribution $p_s(x; \lambda)$ can be qualitatively different from that of the deterministic dynamics. We will associate noise-induced critical points of the stationary distribution with *noise-induced transitions*.

We will now restrict attention to the case that the parameter λ appears linearly in the drift:

$$b(x; \lambda) = f(x) + \lambda g(x).$$

This is the case of the Landau–Stuart dynamics with $g(x) = x$ and $f(x) = -x^3$. We will consider the case in which λ is fluctuating:

$$\lambda \mapsto \lambda + \sigma \dot{W}.$$

We model the fluctuations as white noise. This is a modeling assumption. We can also consider, for example, noise with nonzero correlation time, for example an Ornstein–Uhlenbeck process.

The Itô SDE with multiplicative noise that we will study takes the form

$$dX_t = (f(X_t) + \lambda g(X_t)) dt + \sigma g(X_t) dW_t. \quad (5.97)$$

Equation (5.96) becomes

$$f(x^*) + \lambda g(x^*) - \sigma^2 g(x^*)g'(x^*) = 0. \quad (5.98)$$

Notice that when both f and g are polynomials, noise-induced transitions are expected to appear when the degree of g is greater than that of f .

As an example, consider the Landau–Stuart equation with a fluctuating bifurcation parameter λ :

$$dX_t = (\lambda X_t - X_t^3) dt + \sigma X_t dW_t, \quad X_0 = x. \quad (5.99)$$

Since $X_t = 0$ is always a solution to this equation, X_t will remain positive (negative) if $x_0 > 0$ ($x_0 < 0$). We will consider the case that the initial condition is positive and deterministic and study the dynamics (5.99) in $(0, +\infty)$. Using (5.94), we obtain the formula for the stationary distribution:

$$p_s(x) = \frac{1}{Z} x^\gamma e^{-\frac{x^2}{\sigma^2}}, \quad \gamma = 2 \left(\frac{\lambda}{\sigma^2} - 1 \right). \quad (5.100)$$

The normalization constant is

$$Z = \int_0^{+\infty} x^\gamma e^{-\frac{x^2}{\sigma^2}}.$$

In order for the stationary distribution to be normalizable, it is required that $\gamma > -1$, from which we deduce that

$$\sigma^2 < 2\lambda. \quad (5.101)$$

In this parameter regime, we can prove convergence to equilibrium by finding an appropriate Lyapunov function; see Exercise 15.⁹ When this condition is not satisfied, the stationary distribution becomes a delta function centered at $x = 0$:

$$p_s(x) = \delta(x), \quad \sigma^2 \geq 2\lambda.$$

We can think of a *noise-induced stabilization* phenomenon: the solution of the SDE (5.99) converges to 0 when started at an arbitrary positive initial condition x_0 .

Let us assume now that (5.101) is satisfied and let us consider the critical points of the stationary distribution. Equation (5.98) gives

$$x((\lambda - \sigma^2) - x^2) = 0. \quad (5.102)$$

The critical points are

$$x^* = 0 \quad \text{and} \quad (5.103a)$$

$$x^* = \sqrt{\lambda - \sigma^2} \quad \text{when } \lambda > \sigma^2. \quad (5.103b)$$

On the other hand, only the critical point $x^* = 0$ exists when $\lambda \leq \sigma^2$. Consequently, the behavior of the stationary distribution is qualitatively different in the different parameter regimes $\lambda \leq \frac{\sigma^2}{2}$ (the stationary distribution is a delta function), $\lambda \in \left(\frac{\sigma^2}{2}, \sigma^2 \right]$ (the stationary distribution has one critical point), and $\lambda > \sigma^2$ (the stationary distribution has two critical points). Thus, the effect of multiplicative noise is to induce an additional transition to the dynamics and also to shift the transition point from $\lambda = 0$ (additive noise) to $\lambda = \frac{\sigma^2}{2}$. The effective potential and the corresponding stationary distribution in these two regimes are shown in Fig. 5.8.

⁹ Note that the generator of (5.99) is not uniformly elliptic, and consequently, the techniques developed in Sect. 4.6 are not directly applicable.

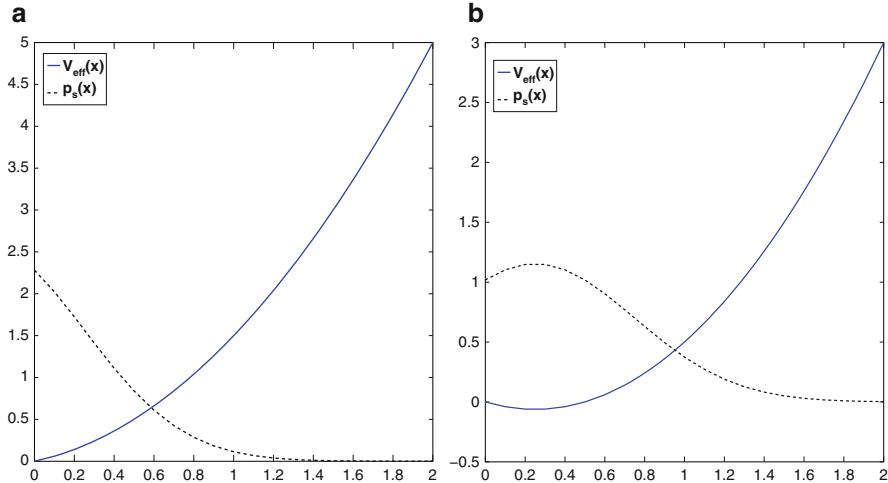


Fig. 5.8 Potential and stationary distribution of the Landau–Stuart equation for negative and positive values of the bifurcation parameter λ **(a)** $\lambda \in \left(-\frac{\sigma^2}{2}, \sigma^2\right]$. **(b)** $\lambda > \sigma^2$

A perhaps more interesting example is that of multiplicative perturbations of the ordinary differential equation

$$\frac{dx}{dt} = \frac{1}{2} - x + \lambda x(1 - x), \quad (5.104)$$

where $x \in [0, 1]$. Assuming that the bifurcation parameter λ is fluctuating, we obtain the Itô SDE

$$dX_t = \left(\frac{1}{2} - X_t + \lambda X_t(1 - X_t) \right) dt + \sigma X_t(1 - X_t) dW_t, \quad (5.105)$$

with $x \in [0, 1]$. We can check that the stationary distribution is normalizable for all values of σ ; see Exercise 14. Now the equation for the critical points of the stationary distribution (5.98) is of higher order than the equation that gives us the stationary points of the deterministic dynamics and is expected to have more roots. Indeed, consider for simplicity the case in which there are only fluctuations in the bifurcation parameter, $\lambda = 0$. The critical point of the deterministic dynamics is

$$x^* = \frac{1}{2}.$$

Multiplicative noise leads to two additional critical points:

$$x_{\pm}^* = \frac{1}{2} \left(1 \pm \frac{1}{2} \sqrt{1 - \frac{2}{\sigma^2}} \right).$$

The stationary distribution changes from unimodal to bimodal at the critical noise strength $\sigma = \sqrt{2}$. This transition is due to noise and is absent from the deterministic dynamics.

5.5 Discussion and Bibliography

SDEs driven by Gaussian noise with nonzero correlation time, *colored noise*, arise in many applications in physics, chemistry, and biology. See [81]. Many of the properties of stochastic differential equations, such as stability and ergodicity, appear to be quite robust with respect to the correlation time of the noise; see [24]. Background material on the multiscale techniques used in the proof of Proposition 5.1 can be found in [185].

The fact that smooth approximations to white noise lead to the Stratonovich stochastic integral is usually called the Wong–Zakai theorem [240, 241]. See also [229, 230] and [224]. In the multidimensional case, this limiting procedure will in general lead to a drift term, in addition to the stochastic integral. This drift term is related to the noncommutativity between the row vectors of the diffusion matrix. A very detailed study of these problems can be found in [98]. The theory of rough paths [147] provides an elegant framework for studying such limit theorems. See, for example, [16].

The correct interpretation of the stochastic integral in physical systems is a part of the overall modeling methodology and ideally, it should be chosen based on experimental data. Results of such experiments for Brownian particles have been reported in [30, 238] and the references therein. Reviews of the so-called *Itô versus Stratonovich problem* can be found in [116, 235]. In fact, there are instances in which neither the Itô nor the Stratonovich interpretation is correct. For example, when there is an additional small time scale in the problem, in addition to the correlation time of the smooth approximation to white noise, it is possible to obtain different interpretations of the stochastic integral in the limit as both time scales go to 0. Consider, for example, the SDE

$$\tau_0 \varepsilon^{\gamma} \ddot{x} = -\dot{x} + \frac{f(x) \eta^{\varepsilon}(t)}{\varepsilon}, \quad (5.106)$$

where $\eta^{\varepsilon}(t)/\varepsilon$ is the Ornstein–Uhlenbeck approximation to white noise (5.1b). When $\gamma < 2$, inertial effects dominate, and $x(t)$ is sufficiently regular that Proposition 3.1 applies, and in the limit as $\varepsilon \rightarrow 0$, we obtain the Itô SDE. On the other hand, when $\gamma > 2$, inertial effects are weak, Eq. (5.106) is effectively of first order in time, and the Wong–Zakai theorem applies: in the limit as $\varepsilon \rightarrow 0$, we arrive at

the Stratonovich SDE. When $\gamma = 2$, then the two small scales in the system are of the same magnitude, and we obtain a whole one-parameter family of stochastic integrals, parameterized by τ_0 in (5.106) and α, D in (5.1b). The limiting equation, written in Itô form, is

$$\dot{x} = \frac{D}{\alpha^2(1 + \tau_0\alpha)} f'(x)f(x) + \sqrt{\frac{2D}{\alpha^2}} f(x) \dot{W}. \quad (5.107)$$

Details can be found in [130, 182].

In fact, the coefficient α can be allowed to depend also on x [94]. Consider, for example, the SDE

$$dX_t = b(X_t) dt + \sigma(X_t) \circ^{\alpha(x)} dW_t, \quad (5.108)$$

where the stochastic integral is defined as the L^2 -limit

$$\int_0^t \sigma(X_s) \circ^{\alpha(x)} dW_s = \lim_{N \rightarrow +\infty} \sum_{n=0}^N \sigma(X_{t_n^\alpha}) \Delta W_n, \quad t_n^\alpha = \frac{n + \alpha(X_{t_n})}{N}. \quad (5.109)$$

As with the Stratonovich and Klimontovich stochastic integrals, we can rewrite the integral in (5.109) as an Itô integral with a correction term. In particular, the SDE (5.108) can be rewritten as an Itô SDE

$$dX_t = b(X_t) dt + \int_0^t \alpha(X_s) \sigma(X_s) \sigma'(X_s) ds + \int_0^t \sigma(X_s) dW_s. \quad (5.110)$$

The choices $\alpha = \frac{1}{2}$ and $\alpha = 1$ lead to the Stratonovich and the Klimontovich corrections, respectively. It turns out that for a particle coupled to a heat bath, the correct interpretation is the Klimontovich one, $\alpha = 1$. It is important to note that an SDE with multiplicative noise can have different stability properties, depending on the choice of the stochastic integral.

The stochastic integral (5.109) arises in the zero-mass limit of the Langevin equation (3.17). We consider the more general case (in one dimension) in which the fluctuation–dissipation theorem (see Eq. (6.2) with space-dependent friction) does not hold:

$$mdq_t = p_t dt, \quad (5.111a)$$

$$dp_t = F(q_t) dt - \gamma(q_t) p_t dt + \sqrt{2D(q_t)} dW_t. \quad (5.111b)$$

In the limit as $m \rightarrow 0$, we obtain the overdamped Itô SDE

$$dQ_t = \left(\frac{F(Q_t)}{\gamma(Q_t)} - \frac{D(Q_t)}{\gamma^3(Q_t)} \gamma'(Q_t) \right) dt + \sqrt{2 \frac{D(Q_t)}{\gamma^2(Q_t)}} dW_t. \quad (5.112)$$

Using the integral (5.109), we can rewrite (5.112) in the form (5.108) with

$$\alpha(q) = -\frac{\ln(\gamma(q))'}{\ln(\lambda(q))'}, \quad \lambda(q) = \frac{D(q)}{\gamma^2(q)}. \quad (5.113)$$

The reader is asked to provide the details of the above results in Exercise 2.

The standard reference on the numerical solution of stochastic differential equations is [117]. That book contains a wealth of material on numerical methods for SDEs as well as applications of stochastic differential equations to problems in the sciences, engineering, and finance. A very useful practical introduction to the numerical simulation of stochastic differential equations can be found in [89]. Section 5.2.1 is based on [88, 214]. Both the stochastic theta method and the Milstein theta scheme are semi-implicit schemes, since we treat the drift implicitly, but not the noise. Treating the noise implicitly involves reciprocals of Gaussian random variables that do not have finite absolute moments. If we are interested only in weak convergence, we can replace the noise by random variables that are accurate in the weak sense and whose reciprocals are easier to take. Then it is possible to consider fully implicit schemes. Details can be found in [117].

Statistical inference for diffusion processes is studied in [15, 23, 133]. A recent review article is [221]. The derivation of the likelihood function from the Euler discretization of the SDE is taken from [117].

Section 5.4 is based on [93], where noise-induced transitions are studied extensively. See also [134, 149, 150], where Lyapunov function techniques are used to study noise-induced stabilization and convergence to equilibrium.

The Landau–Stuart equation (5.86) is the standard example of an amplitude equation describing the amplitude of small solutions to evolution PDEs¹⁰ close to the instability threshold (bifurcation point) [132]. Similarly, the stochastic Landau–Stuart equation, either with additive or multiplicative noise, appears as the amplitude equation for stochastic partial differential equations [26, 27]. Noise-induced transitions, stabilization, and intermittent behavior appear also in stochastic PDEs, and they can often be studied through the analysis of the amplitude equation, the Landau–Stuart equation with additive and/or multiplicative noise. Such an analysis for the Kuramoto–Shivashinsky equation near the bifurcation is presented in [189].

Quite remarkably, and contrary to intuition, additive noise can stabilize solutions of deterministic systems that are not globally asymptotically stable. Such noise-induced stabilization phenomena are studied in two dimensions in [9, 28]. The analysis is based on the construction of appropriate Lyapunov functions.

5.6 Exercises

1. a. Study the stochastic integral (5.109) and the corresponding SDE (5.108). In particular, derive (5.110).
b. Define this stochastic integral in arbitrary dimensions; write it as an Itô integral and obtain a formula for the correction in the drift.

¹⁰ In bounded domains. For evolution PDEs in unbounded domains, the amplitude equation is also a PDE, the Ginzburg–Landau equation. See [39] for details.

2. Consider (5.111) and set $v = \sqrt{mp}$.
- Use the singular perturbation techniques from Sect. 5.1 to derive (5.112). See also Sect. 6.5.1 and Exercise 8 in Chap. 6.
 - Prove (5.113). Analyze this formula when

$$D(q) \propto \gamma^\lambda(q), \quad \lambda > 0.$$

- What type of stochastic integral do you obtain when the fluctuation–dissipation theorem, i.e., $\lambda = 1$, is satisfied?
3. Let X_t denote the solution of the SDE

$$dX_t = \mu X_t dt + \sigma X_t \circ^\alpha dW_t, \quad (5.114)$$

where the stochastic integral in the above equation is defined in (5.109) with $\alpha \in [0, 1]$ constant. Use the transformation (5.110), Itô’s formula, and the *law of the iterated logarithm*

$$\limsup_{t \rightarrow +\infty} \frac{W_t}{\sqrt{2t \log \log t}} = 1 \quad \text{a.s.} \quad (5.115)$$

- to study the long-time asymptotics of solutions to (5.114) as a function of the parameters μ , σ , and α .
- Consider the SDEs (3.15). Apply repeatedly Itô’s formula to appropriate functions of x and y (starting with $F(x, y) = f(x)y$) and use estimates on the solution of the Ornstein–Uhlenbeck process (see, for example, Appendix A in [181]) to obtain a pathwise version of Result 5.1, together with error estimates.
 - Prove (5.22).
 - Consider the stochastic differential equation in \mathbb{R}^2

$$d\mathbf{X}_t = \mathbf{v}(t, \mathbf{X}_t) dt + \sqrt{2\kappa} dW_t, \quad \mathbf{X}_0 = \mathbf{x}, \quad (5.116)$$

where W_t denotes standard Brownian motion on \mathbb{R}^2 . Let $\mathbf{X}_t = (X_t, Y_t)$ and $\mathbf{x} = (x, y)$.

- Solve (5.116) (analytically and numerically) for

$$\mathbf{v}(t, \mathbf{x}) = (0, \sin(x) \sin(\omega t)).$$

Calculate

$$D(t) = \frac{\text{Var}(Y_t)}{2t}.$$

Show numerically that this *effective diffusion coefficient* D becomes constant when t is sufficiently long. Study numerically the dependence of D on κ and ω . Generate plots of D as a function of κ for $\omega = 1$, for $\kappa \in [10^{-2}, 10^2]$, and of D as a function of ω for $\kappa = 0.1$ for $\omega \in [10^{-2}, 10^2]$. You can take the initial conditions for (5.116) to be either deterministic or random.

b. Repeat the same experiment for

$$\mathbf{v}(t, \mathbf{x}) = (0, \sin(x)\eta_t),$$

where η_t is the solution of the SDE

$$d\eta_t = -\alpha\eta_t dt + \sqrt{2\alpha} dW_t, \quad \eta_0 \sim \mathcal{N}(0, 1).$$

Study the dependence of the effective diffusion coefficient on κ and on α .

7. Consider the system of SDEs

$$\frac{dx_t}{dt} = \frac{x_t y_t}{\varepsilon} - y_t^2 x_t^3, \quad x_0 = 1, \quad (5.117a)$$

$$dy_t = -\frac{1}{\varepsilon^2} y_t dt + \sqrt{2} \frac{1}{\varepsilon} dW_t, \quad y_0 \sim \mathcal{N}(0, 1). \quad (5.117b)$$

In the limit as $\varepsilon \rightarrow 0$, the solution x_t of (5.117a) converges to the solution of the SDE

$$dX_t = (X_t - X_t^3) dt + \sqrt{2} X_t dW_t, \quad X_0 = 1. \quad (5.117c)$$

Investigate this limit numerically:

- Solve (5.117) for ε sufficiently small to show that x_t is close to X_t . Solve the equations over $[0, T]$ with $T = 10$ with a sufficiently small step size and an appropriate numerical method. Discuss the choice of the numerical method.
- Calculate $\text{err}(\varepsilon) = \mathbb{E}|x_T - X_T|^2$ as a function of ε . Show numerically that $\text{err}(\varepsilon) \approx C\varepsilon^\gamma$ for ε sufficiently small and estimate the exponent γ .

8. (Consult [88, 214] if necessary). Consider the scalar SDE

$$dX_t = b(X_t) dt + \sigma(X_t) dW_t, \quad X_0 = x.$$

The theta Milstein scheme is

$$X_{n+1} = X_n + (\theta b_{n+1} + (1 - \theta)b_n) \Delta t + \sigma_n \Delta W_n + \frac{1}{2} \sigma_n \sigma'_n ((\Delta W_n)^2 - \Delta t), \quad (5.118)$$

where $\theta \in [0, 1]$ and $X_n = X(n\Delta t)$, $b_n = b(X_n)$, $\sigma_n = \sigma(X_n)$. Consider the SDE

$$dX_t = \lambda X_t dt + \sigma X_t dW_t, \quad (5.119)$$

where $\lambda, \sigma \in \mathbb{C}$.

- Obtain a formula for $\mathbb{E}|X_t|^2 = \mathbb{E}X_t \bar{X}_t$, where \bar{X}_t denotes the complex conjugate. Show that X_t is mean-square stable if

$$2\text{Re}(\lambda) + |\sigma|^2 < 0.$$

- b. Apply the θ Milstein scheme to (5.119). Show that it can be written in the form

$$X_{n+1} = G(\Delta t, \lambda, \sigma, \theta)X_n$$

and obtain a formula for $G(\Delta t, \lambda, \sigma, \theta)$. Let $Z_n = \mathbb{E}X_n^2$. Use the previous calculation to obtain an equation of the form

$$Z_{n+1} = R(\Delta t, \lambda, \sigma, \theta)Z_n.$$

- c. Investigate the region of mean-square stability for the theta Milstein scheme. Compare the stability region of the numerical scheme with the stability region of the SDE. For what values of θ is the theta Milstein scheme A-stable?
d. Test your results for the SDEs

$$dX_t = -100X_t dt + 10X_t dW_t$$

and

$$dX_t = (-100 + 100i)X_t dt + 10X_t dW_t.$$

Comment on your results.

9. (Consult [88] if necessary). The following stochastic equation appears in population dynamics:

$$dX_t = -\lambda X_t(1 - X_t) dt - \mu X_t(1 - X_t) dW_t. \quad (5.120)$$

- a. Show that $X_t = 1$ is a fixed point for (5.120) and that in linearizing about this fixed point, we obtain the SDE for geometric Brownian motion.
b. Solve (5.120) numerically using the explicit Euler and Milstein schemes for $\lambda = -1$, $X_0 = 1.1$ and for $\mu = 0.5, 0.6, 0.7, 0.8, 0.9$. Calculate numerically $\mathbb{E}(X_t - 1)^2$. Comment on the mean-square stability of the explicit Euler and Milstein schemes for the nonlinear SDE (5.120).
c. Solve (5.120) using the theta scheme with $\theta = \frac{1}{2}$. Investigate the mean-square stability of this numerical scheme applied to (5.120).
10. Consider the SDE

$$dX_t = \sum_{n=1}^N \alpha_n X_t^n dt + dW_t. \quad (5.121)$$

- a. Write down the log likelihood function.
b. Obtain a system of equations for the MLE $\{\hat{\alpha}_n\}_{n=1}^N$.
c. Write a computer program for estimating the coefficients using maximum likelihood for arbitrary N .
d. Test your program (in particular, the performance of the MLE estimator as a function of the window of observation $[0, T]$) for $N = 3, 5, 7$ and $a_{2n} = 0$ (you need to consider cases in which solutions to (5.121) exist globally in time and have nice ergodic properties).

11. Derive the MLE estimators (5.61). Study their asymptotic properties in the limit as the number of observations becomes infinite, $N \rightarrow +\infty$.
12. Consider the multidimensional SDE

$$dX_t = b(X_t; \theta) dt + dW_t, \quad (5.122)$$

where $X_t \in \mathbb{R}^d$ and b is a vector field such that (5.122) has a unique solution that is ergodic.

- a. Write down the log likelihood function.
- b. Consider the case

$$b(x; \theta) = -\nabla V(x; \theta), \quad V(x; \theta) = \frac{1}{2}x^T \theta x, \quad (5.123)$$

where θ is an $n \times n$ real-valued symmetric matrix. Obtain formulas for the MLE $\hat{\theta}$.

- c. Write a computer program for calculating the maximum likelihood estimators $\hat{\theta}$ from discrete observations of (5.122) for the drift (5.123).
- d. Test your code for $d = 2$. Can you estimate the off-diagonal elements of θ ?
13. Study additive and multiplicative random perturbations of the ODE

$$\frac{dx}{dt} = x(\lambda + 2x^2 - x^4).$$

In particular, study the dependence of the stationary distribution on the bifurcation parameter λ .

14. Analyze the stationary distribution of the SDE

$$dX_t = \left(\frac{1}{2} - X_t + \lambda X_t(1 - X_t) \right) dt + \sigma X_t(1 - X_t) dW_t, \quad (5.124)$$

with $x \in [0, 1]$. Study, in particular, the dependence of the stationary distribution on λ and σ .

15. a. Show that $U(x) = x^2$ is a Lyapunov function for the Landau–Stuart equation with additive noise (5.93).
- b. Show that $U(x) = \cosh(qx)$, for an appropriately chosen constant q , is a Lyapunov function for the Landau–Stuart equation with multiplicative noise (5.99) when $\delta > \sigma^2/2$. (Hint: use Lamperti’s transformation.)
16. a. Study the stationary probability density of the Stratonovich SDE

$$dX_t = b(X_t; \lambda) dt + \sigma g(X_t) \circ dW_t, \quad (5.125)$$

posed on (ℓ, r) with reflecting boundary conditions. In particular, find conditions under which the stationary density is normalizable and investigate noise-induced transitions.

- b. Analyze the stochastic Landau–Stuart equation (5.99) with multiplicative noise when the noise is interpreted in the Stratonovich sense.

17. The following system of SDEs arises in the study of the magnetic dynamo problem [99]:

$$\begin{aligned} dx_t &= -\mu x_t dt + x_t y_t dt + \sigma dW_t^x, \\ dy_t &= -\nu y_t dt + (1 - x_t^2) dt + \sigma dW_t^y, \end{aligned}$$

where (W_t^x, W_t^y) is a standard two-dimensional Brownian motion. Construct a Lyapunov function for this system of SDEs.

18. Consider the stochastically perturbed Lorenz system [99]

$$\begin{aligned} dx_t &= ay_t dt + \sigma dW_t^x, \\ dy_t &= \left[(b - 1)x_t - (a + 1)y_t - x_t z_t \right] dt + \sigma W_t^y, \\ dz_t &= (-cz_t + x_t y_t + x_t^2) dt + \sigma dW_t^z, \end{aligned}$$

where a, b, c, σ are positive constants and (W_t^x, W_t^y, W_t^z) is a standard three-dimensional Brownian motion. Construct a Lyapunov function for this system of SDEs.

Chapter 6

The Langevin Equation

In this chapter, we study the Langevin equation and the associated Fokker–Planck equation. In Sect. 6.1, we introduce the equation and study some of the main properties of the corresponding Fokker–Planck equation. In Sect. 6.2 we give an elementary introduction to the theories of hypoellipticity and hypocoercivity. In Sect. 6.3, we calculate the spectrum of the generator and Fokker–Planck operators for the Langevin equation in a harmonic potential. In Sect. 6.4, we study Hermite polynomial expansions of solutions to the Fokker–Planck equation. In Sect. 6.5, we study the overdamped and underdamped limits for the Langevin equation. In Sect. 6.6, we study the problem of Brownian motion in a periodic potential. Bibliographical remarks and exercises can be found in Sects. 6.7 and 6.8, respectively.

6.1 The Fokker–Planck Equation in Phase Space

The Langevin equation describes the motion of a particle that is subject to friction and stochastic forcing:

$$\ddot{q} = -\nabla V(q) - \gamma \dot{q} + \sqrt{2\gamma\beta^{-1}} \dot{W}. \quad (6.1)$$

This is Newton’s equation of motion with two additional terms, a linear dissipation term $\gamma \dot{q}$ and a stochastic forcing $\sqrt{2\gamma\beta^{-1}} \dot{W}$. The noise $\xi(t) = \sqrt{2\gamma\beta^{-1}} \dot{W}$ and dissipation $-\gamma \dot{q}$ in (6.1) are controlled by two parameters, the *friction coefficient* γ and the *temperature* $\beta^{-1} = k_B T$, where k_B denotes Boltzmann’s constant and T the absolute temperature. Dissipation and noise are related through the *fluctuation–dissipation theorem*:

$$\mathbb{E}(\xi(t)\xi(s)) = \gamma\beta^{-1} \delta(t-s). \quad (6.2)$$

The Langevin equation describes the dynamics of a particle that moves according to Newton’s second law and is in contact with a thermal reservoir that is at equilibrium

at time $t = 0$ at temperature β^{-1} . In Chap. 8, we will derive the (generalized) Langevin equation, together with the fluctuation–dissipation theorem, starting from a simple model of an *open classical system*.

Introducing the momentum $p_t = \dot{q}_t$ (we write $q_t = q(t)$, $p_t = p(t)$), we can write the Langevin equation as a system of first-order stochastic differential equations in phase space $(q, p) \in \mathbb{R}^{2d}$:

$$dq_t = p_t dt, \quad (6.3a)$$

$$dp_t = -\nabla V(q_t) dt - \gamma p_t dt + \sqrt{2\gamma\beta^{-1}} dW_t. \quad (6.3b)$$

The position and momentum $\{q_t, p_t\}$ define a Markov process with generator

$$\mathcal{L} = p \cdot \nabla_q - \nabla_q \cdot V \nabla_p + \gamma(-p \nabla_p + \beta^{-1} \Delta_p). \quad (6.4)$$

Assume that initially, the position and momentum are distributed according to a distribution $\rho_0(q, p)$ (when q_0, p_0 are deterministic and the initial distribution is $\rho_0(q, p) = \delta(q - q_0)\delta(p - p_0)$). The evolution of the probability distribution function of the Markov process $\{q_t, p_t\}$ is governed by the Fokker–Planck equation

$$\frac{\partial \rho}{\partial t} = -p \cdot \nabla_q \rho + \nabla_q V \cdot \nabla_p \rho + \gamma(\nabla_p \cdot (p \rho) + \beta^{-1} \Delta_p \rho), \quad (6.5a)$$

$$\rho(q, p, 0) = \rho_0(q, p). \quad (6.5b)$$

Notice that the Fokker–Planck operator \mathcal{L}^* that appears in Eq. (6.5a) (i.e., the $L^2(\mathbb{R}^{2d})$ -adjoint of (6.4)) is not a uniformly elliptic operator: there are second-order derivatives only with respect to the momentum variable p . This is a consequence of the fact that noise in (6.3) appears only in the equation for the momentum. Consequently, Theorem 4.1 cannot be used to deduce existence and uniqueness of solutions for the initial value problem (6.5). Furthermore, the PDE argument that we can use for studying the ergodic properties of SDEs with a strictly positive definite diffusion matrix (which leads to a uniformly elliptic generator and Fokker–Planck operator) cannot be directly applied to the Langevin equation (6.3). In particular, it is not clear that a smooth probability density $\rho(q, p, t)$ exists. It turns out that noise, at the level of the trajectories, and hence regularity, at the level of probability densities, gets transmitted from the equation for the momentum to the equation for the position. The generator and the Fokker–Planck operator of the Langevin dynamics are an example of a *hypoelliptic* operator. Hypoellipticity for the generator of the Langevin dynamics will be studied in Sect. 6.2.

The first-order differential operator

$$B = -p \cdot \nabla_q + \nabla_q V \cdot \nabla_p \quad (6.6)$$

is the Liouville operator from classical (statistical) mechanics, corresponding to the Hamiltonian vector field with Hamiltonian¹

¹ Hamilton's equations of motion are $\dot{q} = \frac{\partial H}{\partial p} = p$, $\dot{p} = -\frac{\partial H}{\partial q} = -\nabla V(q)$. The Hamiltonian vector field is $b(q, p) = (p, -\nabla V)$. The corresponding Liouville operator is given by B . See Sect. 3.4.

$$H(p, q) = \frac{1}{2}|p|^2 + V(q). \quad (6.7)$$

The energy of a Hamiltonian system is conserved under the (deterministic) dynamics. More generally, arbitrary differentiable functions of the Hamiltonian belong to the null space of the Liouville operator:

$$Bf(H) = 0. \quad (6.8)$$

A consequence of this is that Hamiltonian systems have many invariant distributions. The presence of noise and dissipation in (6.3) results in “selecting” a unique invariant distribution:

Proposition 6.1. *Let $V(x)$ be a smooth confining potential. Then the Markov process with generator (6.4) is ergodic. The unique invariant distribution is*

$$\rho_\beta(p, q) = \frac{1}{Z} e^{-\beta H(p, q)}, \quad (6.9)$$

where H is the Hamiltonian (6.7), and the normalization factor Z is the partition function

$$Z = \int_{\mathbb{R}^{2d}} e^{-\beta H(p, q)} dp dq. \quad (6.10)$$

Proof. Since ρ_β is a function of the Hamiltonian, using (6.8), we conclude that it is sufficient to prove that ρ_β is in the null space of the Fokker–Planck operator $\mathcal{S} := \nabla_p \cdot (p \cdot + \beta^{-1} \nabla_p \cdot)$, i.e., the Fokker–Planck operator of the Ornstein–Uhlenbeck process. Since the Maxwell–Boltzmann distribution depends on p through $e^{-\beta|p|^2/2}$, this follows from the results of Sect. 4.4. The proof of the fact that (6.9) is the unique solution of the stationary Fokker–Planck equation is discussed in Sect. 6.7. \square

Observe that the invariant distribution (6.9) is independent of the friction coefficient: as long as $\gamma > 0$ in (6.3) (or equivalently, (6.5a)), then the dynamics q_t, p_t becomes ergodic with respect to (6.9). We will refer to this distribution as the *Gibbs*, the *Maxwell–Boltzmann*, or the *canonical* distribution. This distribution is Gaussian, and by calculating the Gaussian integrals in (6.10), it can be written as

$$\rho_\beta(q, p) = \frac{1}{(2\pi)^{d/2} \int_{\mathbb{R}^d} e^{-\beta V(q)} dq} e^{-\beta(\frac{1}{2}|p|^2 + V(q))}. \quad (6.11)$$

In particular, the marginal distribution in q is precisely the invariant distribution $\rho_\beta(q) = Z^{-1} e^{-\beta V(q)}$, $Z = \int_{\mathbb{R}^d} e^{-\beta V(q)} dq$ of the reversible dynamics

$$dq_t = -\nabla V(q_t) dt + \sqrt{2\beta^{-1}} dW_t. \quad (6.12)$$

The connection between the Langevin dynamics 6.3 and the *overdamped Langevin* dynamics (6.12) will be explored further in Sect. 6.5.1.

As with the Fokker–Planck equation corresponding to (6.12) that we studied in Sect. 4.5, it is convenient to map the Fokker–Planck equation (6.5) to the backward Kolmogorov equation. Let $\rho(q, p, t)$ denote the solution of (6.5) and write

$$\rho(q, p, t) = h(q, p, t) \rho_\beta(q, p). \quad (6.13)$$

The function $h(q, p, t)$ is the solution to the partial differential equation

$$\frac{\partial h}{\partial t} = \mathcal{L}_{kin} h \quad (6.14)$$

with $h(q, p, 0) = \rho_\beta^{-1} \rho_0(q, p)$:

$$\mathcal{L}_{kin} = -p \cdot \nabla_q + \nabla_q V \cdot \nabla_p + \gamma(-p \cdot \nabla_p + \beta^{-1} \Delta_p). \quad (6.15)$$

This is almost the generator of the Langevin dynamics: it differs from \mathcal{L} defined in (6.4) in the sign in front of the Liouville part of the generator. This is related to the fact that the dynamics (6.3) are not reversible: in considering the time-reversed dynamics, we also need to change the sign of the momentum p . Equivalently, if we write

$$\rho(q, p, t) = \hat{h}(q, -p, t) \rho_\beta(q, p) \quad (6.16)$$

instead of (6.13), then \hat{h} satisfies the backward Kolmogorov equation

$$\frac{\partial \hat{h}}{\partial t} = \mathcal{L} \hat{h}, \quad (6.17)$$

with $\hat{h}(q, p, 0) = \rho_\beta^{-1}(q, p) \rho_0(q, -p)$. To solve the Fokker–Planck equation (6.5), it is sufficient to solve either (6.14) or (6.17). As with the Smoluchowski equation, the right function space in which to study these equations is the Hilbert space (4.72):

$$L^2(\rho_\beta) := \left\{ f \left| \int_{\mathbb{R}^{2d}} |f|^2 \rho_\beta(p, q) dp dq < \infty \right. \right\}. \quad (6.18)$$

This is a Hilbert space with inner product (4.73), $\langle f, h \rangle_\rho := \int_{\mathbb{R}^{2d}} f h \rho_\beta dp dq$, and corresponding norm $\|f\|_\rho = \sqrt{\langle f, f \rangle_\rho}$. The generator \mathcal{L} has nice properties in this function space that will be useful in the study of convergence to equilibrium and in the calculation of its spectrum when the potential is quadratic.

Lemma 6.2. *The generator \mathcal{L} (6.4) can be written as*

$$\mathcal{L} = -B + \gamma \mathcal{S}, \quad (6.19)$$

where B is the Liouville operator (6.6) and $\mathcal{S} = -p \cdot \nabla_p + \beta^{-1} \Delta_p$. Here B is anti-symmetric and \mathcal{S} symmetric in $L^2(\rho_\beta)$.

Proof. The antisymmetry of B follows from the fact that it is a first-order differential operator and that $B \rho_\beta = 0$; see (6.8). The symmetry of \mathcal{S} follows from the fact that the Maxwell–Boltzmann distribution (6.11) is the product of the Gaussian distribution $e^{-\beta p^2/2}$ and $e^{-\beta V}$ and Proposition 4.1, Eq. (4.45) in particular. \square

We can now use the creation and annihilation operators that were introduced in Sect. 4.4, see (4.58) and (4.59), to write \mathcal{L} in Hörmander’s “sum of squares form”

in $L^2(\rho_\beta)$. Let $A_i := \frac{1}{\sqrt{\beta}} \frac{\partial}{\partial p_i}$. The $L^2(\rho_\beta)$ -adjoint of A_i is $A_i^* = +\sqrt{\beta} p_i - \frac{1}{\sqrt{\beta}} \frac{\partial}{\partial p_i}$. We have that the symmetric part of the generator can be written as

$$\mathcal{S} = - \sum_{i=1}^d A_i^* A_i.$$

Consequently, the generator of the Markov process $\{q_t, p_t\}$ can be written in the sum of squares form

$$\mathcal{L} = -B - \gamma \sum_{i=1}^d A_i^* A_i. \quad (6.20)$$

We calculate the commutators² between the vector fields in (6.20):

$$[B, A_i] = \frac{1}{\sqrt{\beta}} \frac{\partial}{\partial q_i}, \quad [A_i, A_j] = 0, \quad [A_i, A_j^*] = \delta_{ij}. \quad (6.21)$$

Let now $Y_i = \frac{1}{\sqrt{\beta}} \frac{\partial}{\partial q_i}$ with $L^2(\rho_\beta)$ -adjoint $Y_i^* = -\frac{1}{\sqrt{\beta}} \frac{\partial}{\partial q_i} + \sqrt{\beta} \frac{\partial V}{\partial q_i}$. We have that

$$-B = \sum_{i=1}^d (A_i^* Y_i - Y_i^* A_i) = \sum_{i=1}^d \left(p_i \frac{\partial}{\partial q_i} - \frac{\partial V}{\partial q_i} \frac{\partial}{\partial p_i} \right).$$

As a result, the generator can be written in the form

$$\mathcal{L} = \sum_{i=1}^d (A_i^* Y_i - Y_i^* A_i) - \gamma \sum_{i=1}^d A_i^* A_i. \quad (6.22)$$

Clearly, the operators A_i and Y_i as well as their adjoints commute. Furthermore,

$$[B, Y_i] = -\frac{1}{\sqrt{\beta}} \sum_{j=1}^d \frac{\partial^2 V}{\partial q_i \partial q_j} \frac{\partial}{\partial p_j}, \quad [Y_i, Y_j] = 0, \quad [Y_i, Y_j^*] = \frac{\partial^2 V}{\partial q_i \partial q_j}. \quad (6.23)$$

Notice also that the generator of the Smoluchowski dynamics can be written as

$$\mathcal{L}_V := -\nabla_q V \nabla_q + \beta^{-1} \Delta_q = - \sum_{i=1}^d Y_i^* Y_i. \quad (6.24)$$

6.2 Hypoellipticity and Hypocoercivity

³ In this section, we introduce two techniques from analysis to develop a rigorous study of the Fokker–Planck equation (6.5a), or equivalently, (6.14). As we have already mentioned, the difficulty in developing an appropriate existence, uniqueness, and regularity theory of solutions for the Fokker–Planck equation and

² Let A and B denote the first-order differential operators corresponding to the vector fields $\mathbf{A}(x)$ and $\mathbf{B}(x)$, i.e., $A = \sum_j A_j(x) \frac{\partial}{\partial q_j}$, $B = \sum_j B_j(x) \frac{\partial}{\partial q_j}$. The commutator between A and B is $[A, B] = AB - BA$.

³ This section was written in collaboration with M. Ottobre.

in proving exponentially fast convergence to equilibrium lies in the fact that noise in the Langevin equation (6.3) appears only in the equation for the momentum. Equivalently, the Fokker–Planck operator or the generator \mathcal{L} is not a uniformly elliptic operator, since it contains second-order derivatives with respect only to p and not to q . Consequently, it is not possible to use Theorem 4.1 to conclude the existence and uniqueness of solutions. Furthermore, the exponentially fast convergence to equilibrium that we proved for reversible diffusions in Sect. 4.5 relied heavily on the uniform ellipticity of the generator of the process, which enabled us to show that under appropriate assumptions on the potential, the generator had a spectral gap in $L^2(\rho_\beta)$.⁴ It should be noted that in the absence of dissipation and noise, the Fokker–Planck equation reduces to the Liouville equation from classical statistical mechanics, for which no regularity theory exists, nor does it have a one-dimensional null space (which would imply uniqueness of the invariant measure).

It is precisely the presence of dissipation and noise in the equation for the momentum that leads to both regularity at the level of the law of the process and ergodicity and exponentially fast convergence to equilibrium. Thus, the issues of regularity and convergence to equilibrium are related. The crucial observation is that even though there is no dissipation or noise acting directly in the equation for the position in the direction of q (which results in $-\mathcal{L}$ not being a coercive operator ; see Definition 6.2), there is sufficient interaction between the position and momentum variables that noise and dissipation get transmitted from the equation for p to the equation for q . We will see later in this section how to quantify this by introducing commutators between appropriate vector fields. We will treat the issues of the existence of a smooth density and exponentially fast convergence to equilibrium separately.

6.2.1 Hypoellipticity

We consider an Itô stochastic differential equation in \mathbb{R}^n ,

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad (6.25)$$

where W_t is a standard Brownian motion in \mathbb{R}^m , and $b \in \mathbb{R}^n$ and $\sigma \in \mathbb{R}^{n \times m}$ are smooth vector and matrix fields, respectively. The generator of X_t is

$$\mathcal{L} = \sum_{j=1}^n b_j(x) \frac{\partial}{\partial x_j} + \frac{1}{2} \sum_{i,j=1}^n \sum_{k=1}^m \sigma_{ik}(x) \sigma_{jk}(x) \frac{\partial^2}{\partial x_i \partial x_j}.$$

Introducing the notation

$$X_0 = \sum_{j=1}^n b_j(x) \frac{\partial}{\partial x_j}$$

⁴ In other words, the exponentially fast convergence to equilibrium for the reversible Smoluchowski dynamics follows directly from the assumption that e^{-V} satisfies a Poincaré inequality.

and

$$X_k = \sum_{j=1}^n \sigma_{jk}(x) \frac{\partial}{\partial x_j}, \quad k = 1, \dots, m,$$

we can write the generator in Hörmander's "sum of squares" form

$$\mathcal{L} = \sum_{i=1}^m X_i^2 + X_0. \quad (6.26)$$

A similar formula can be obtained for the Fokker–Planck operator. Let now $P(t, x, dy) = p(t, x, y) dy$ denote the transition function. The transition probability density $p(t, x, y)$ satisfies the forward and backward Kolmogorov equations

$$\frac{\partial}{\partial t} p(\cdot, x, \cdot) = \mathcal{L}_y^* p(\cdot, x, \cdot), \quad x \in \mathbb{R}^n \quad (6.27)$$

and

$$\frac{\partial}{\partial t} p(\cdot, \cdot, y) = \mathcal{L}_x p(\cdot, \cdot, y), \quad y \in \mathbb{R}^n, \quad (6.28)$$

see (2.53) and (2.54), where we use the subscripts y and x to clarify the variables with respect to which we are differentiating in the above equations. Since we have not made any assumptions on the diffusion matrix $\Sigma = \sigma \sigma^T$, we do not know whether a smooth transition probability density exists. Consequently, these two equations should be interpreted in a weak sense (in the sense of distributions). We have discussed in earlier chapters that when the diffusion matrix Σ is positive definite uniformly in x , then parabolic regularity theory enables us to conclude that the transition probability density is smooth. The diffusion matrix for the Langevin equation (6.3) is only positive semidefinite, and we need to use a different approach to show that the density is smooth. We can prove this using Hörmander's theory of hypoellipticity.

Definition 6.1. A linear differential operator \mathcal{L} with C^∞ coefficients is called hypoelliptic if every distribution u in an open set Ω such that $\mathcal{L}u \in C^\infty$ must itself be C^∞ in Ω .

In other words, the operator \mathcal{L} is hypoelliptic if all solutions to the equation $\mathcal{L}u = f$ are C^∞ whenever $f \in C^\infty$. An example of a hypoelliptic operator is $-\frac{\partial}{\partial t} + \frac{1}{2}\Delta$, from which it follows that Brownian motion has a smooth density. The same is true for an operator of the form $-\frac{\partial}{\partial t} + \mathcal{L}$, where \mathcal{L} is the generator of a diffusion process with a uniformly positive definite diffusion matrix.

In order to check whether an operator of the form $-\frac{\partial}{\partial t} + \mathcal{L}$, where \mathcal{L} is given by (6.26), is hypoelliptic, we need to check that the interaction between X_0 and X_1, \dots, X_m is such that noise spreads out to the directions where it does not act directly. We make this intuition precise by introducing the concept of a Lie algebra.

Given a collection of vector fields $\mathcal{G} = (X_1, \dots, X_m)$, the *Lie algebra* generated by such fields is the smallest vector space that contains all elements of \mathcal{G} and is closed under the commutator operation.

Let now X_0, \dots, X_m be the vector fields in (6.26) and define the Lie algebras

$$\begin{aligned}\mathcal{A}_0 &= \text{Lie}(X_1, \dots, X_m), \\ \mathcal{A}_1 &= \text{Lie}([X_0, X_1], [X_0, X_2], \dots, [X_0, X_m]), \\ \dots &= \dots, \\ \mathcal{A}_k &= \text{Lie}([X_0, X]; X \in \mathcal{A}_{k-1}), \quad k \geq 1.\end{aligned}$$

Define

$$\mathcal{H} = \text{Lie}(\mathcal{A}_0, \mathcal{A}_1, \dots). \quad (6.29)$$

It is important to note that X_0 is not included in \mathcal{A}_0 . In fact, it does not necessarily belong to \mathcal{H} . Now we can make precise the intuition of “spreading out of noise.”

Assumption 6.3 [Hörmander] \mathcal{H} is full in the sense that for every $x \in \mathbb{R}^d$, the vectors $\{H(x) : H \in \mathcal{H}\}$ span $T_x \mathbb{R}^n$:

$$\text{span}\{H(x) : H \in \mathcal{H}\} = T_x \mathbb{R}^n.$$

We have used the notation $T_x \mathbb{R}^n$ for the tangent space of \mathbb{R}^n at x .

Theorem 6.1 (Hörmander). Assume that the coefficients in (6.25) are smooth with bounded derivatives of all orders and that Assumption 6.3 holds. Then the process X_t with generator \mathcal{L} given by (6.26) has a smooth transition density $p(\cdot, \cdot, \cdot)$ on $(0, +\infty) \times \mathbb{R}^n \times \mathbb{R}^n$.

In particular, when Hörmander’s condition is satisfied and the drift and diffusion coefficients are smooth, the forward and backward Kolmogorov equations (6.27) and (6.28) are valid in a classical sense. From Hörmander’s theorem, it follows that when the generator (6.26) is a second-order uniformly elliptic operator (the diffusion matrix is positive definite uniformly in x), then a smooth density exists. In this case, we can take $k = 0$ in (6.29). This result is sometimes called Weyl’s lemma.

Example 6.1. Consider the stochastic differential equation

$$\ddot{X}_t = \sqrt{2} \dot{W}_t.$$

We write it in the form

$$dX_t = Y_t dt, \quad dY_t = \sqrt{2} dW_t.$$

The generator is

$$\mathcal{L} = X_0 + X_1^2, \quad X_0 = y\partial_x, \quad X_1 = \partial_y, \quad (6.30)$$

and the Fokker–Planck operator is

$$\mathcal{L}^* = X_1^2 - X_0.$$

The generator (6.30) is not a uniformly elliptic operator. We calculate the first commutator

$$[X_0, X_1] = -\partial_x.$$

Consequently,

$$\text{Lie}(X_1, [X_0, X_1]) = \text{Lie}(-\partial_x, \partial_y),$$

which spans $T_x \mathbb{R}^2$.

Consider now the Langevin equation (6.3). We first consider the equation in one dimension and set $\gamma = \beta = 1$. The generator of the process $X_t := (q_t, p_t) \in \mathbb{R}^2$ can be written in the form

$$\mathcal{L} = X_0 + X_1^2,$$

where

$$X_0 = p\partial_q - \partial_q V(q)\partial_p - p\partial_p \quad \text{and} \quad X_1 = \partial_p.$$

We calculate

$$[X_1, X_0] = (\partial_p - \partial_q).$$

Therefore, $\text{span}\{X_1, [X_1, X_0]\} = T\mathbb{R}^2$, and the Langevin dynamics has a smooth density.

A similar calculation shows that the multidimensional Langevin dynamics has a smooth density. In preparation for the calculations that we will do and for the notation that we will use in the next section, on hypocoercivity, we use (6.20) and (6.21) to conclude that

$$\text{Lie}(A_1, \dots, A_d, [B, A_1], \dots, [B, A_d]) = \text{Lie}(\nabla_p, \nabla_q),$$

which spans $\mathbb{T}_{p,q} \mathbb{R}^{2d}$ for all $p, q \in \mathbb{R}^d$. Smoothness of the transition probability density now follows from Theorem 6.1.

6.2.2 Hypocoercivity

The theory of hypocoercivity applies to evolution equations of the form

$$\partial_t h + (A^* A - B)h = 0. \quad (6.31)$$

Equation (6.14) can be written precisely in this form; see (6.20). For simplicity, we will consider the Langevin dynamics in one dimension. The theory presented in this section can be generalized to equations of the form $\partial_t h + (\sum_{i=1}^m A_i^* A_i - B)h = 0$, which in turn can be used to study convergence to equilibrium for the multidimensional Langevin dynamics. We briefly present some of the basic elements of the theory of hypocoercivity and then apply it to the Fokker–Planck equation (6.14).

We first introduce the necessary notation. Let \mathcal{H} be a Hilbert space, real and separable, $\|\cdot\|$ and (\cdot, \cdot) the norm and scalar product of \mathcal{H} , respectively. Let A and B be unbounded operators with domains $\mathcal{D}(A)$ and $\mathcal{D}(B)$ respectively, and assume that B is antisymmetric, i.e., $B^* = -B$, where * denotes the adjoint in \mathcal{H} . We will also assume that there exists a vector space $S \subset \mathcal{H}$, dense in \mathcal{H} , in which all the

operations that we will perform involving A and B are well defined. Note that from the antisymmetry of B , we deduce that

$$\frac{1}{2} \frac{d}{dt} \|h\|^2 = -\|Ah\|^2 \leq 0$$

and

$$\langle -\mathcal{L}_{kin}f, f \rangle \geq 0 \quad (6.32)$$

for all $f \in D(\mathcal{L}_{kin})$. However, the estimate (6.32) is insufficient to enable us to conclude exponentially fast convergence to equilibrium. In particular, the operator \mathcal{L}_{kin} is not coercive:

Definition 6.2 (Coercivity). Let \mathcal{T} be an unbounded operator on a Hilbert space \mathcal{H} . Denote its kernel by \mathcal{K} , and assume that there exists another Hilbert space $\tilde{\mathcal{H}}$ continuously and densely embedded in \mathcal{K}^\perp . If $\|\cdot\|_{\tilde{\mathcal{H}}}$ and $(\cdot, \cdot)_{\tilde{\mathcal{H}}}$ are the norm and scalar product on $\tilde{\mathcal{H}}$, respectively, then the operator \mathcal{T} is said to be λ -coercive on $\tilde{\mathcal{H}}$ if

$$(\mathcal{T}h, h)_{\tilde{\mathcal{H}}} \geq \lambda \|h\|_{\tilde{\mathcal{H}}}^2, \quad \forall h \in \mathcal{K}^\perp \cap D(\mathcal{T}), \quad (6.33)$$

where $D(\mathcal{T})$ is the domain of \mathcal{T} in $\tilde{\mathcal{H}}$.

The following proposition gives an equivalent definition of coercivity.

Proposition 6.4. *With the same notation as in Definition 6.2, \mathcal{T} is λ -coercive on $\tilde{\mathcal{H}}$ if and only if*

$$\|e^{-\mathcal{T}t}h\|_{\tilde{\mathcal{H}}} \leq e^{-\lambda t} \|h\|_{\tilde{\mathcal{H}}} \quad \forall h \in \tilde{\mathcal{H}} \text{ and } t \geq 0.$$

Definition 6.3 (Hypocoercivity). With the same notation of Definition 6.2, assume that \mathcal{T} generates a continuous semigroup. Then \mathcal{T} is said to be λ -hypocoercive on $\tilde{\mathcal{H}}$ if there exists a constant $\kappa > 0$ such that

$$\|e^{-\mathcal{T}t}h\|_{\tilde{\mathcal{H}}} \leq \kappa e^{-\lambda t} \|h\|_{\tilde{\mathcal{H}}}, \quad \forall h \in \tilde{\mathcal{H}} \text{ and } t \geq 0. \quad (6.34)$$

We remark that the only difference between Definitions 6.2 and 6.3 is in the constant κ on the right-hand side of (6.34), when $\kappa > 1$. Thanks to this constant, the notion of hypocoercivity is invariant under a change of equivalent norms, as opposed to the definition of coercivity, which relies on the choice of the Hilbert norm. Hence the basic idea employed in the proof of exponentially fast convergence to equilibrium for degenerate diffusions generated by operators in the form (6.31) is to construct appropriately a norm on $\tilde{\mathcal{H}}$ equivalent to the existing one and such that in this norm, the operator is coercive.

Now we carry out this program by presenting an abstract result that applies to equations of the form (6.31), and then we apply it to the Fokker–Planck equation (6.14). We need to be able to obtain qualitative information on commutators:

Definition 6.4. We say that an unbounded linear operator \mathcal{T} on \mathcal{H} is *relatively bounded* with respect to the linear operators T_1, \dots, T_n if the domain of \mathcal{T} , $\mathcal{D}(\mathcal{T})$, is contained in the intersection $\cap \mathcal{D}(T_j)$ and there exists a constant $\alpha > 0$ such that

$$\forall h \in \mathcal{D}(\mathcal{T}), \quad \|\mathcal{T}h\| \leq \alpha(\|T_1h\| + \dots + \|T_nh\|).$$

Now we are ready to present the main theorem that we will use in the study of the Fokker–Planck equation (6.14).

Theorem 6.2. *With the notation introduced so far, let \mathcal{T} be an operator of the form $\mathcal{T} = A^*A - B$, with $B^* = -B$. Let $\mathcal{K} = \text{Ker } \mathcal{T}$, define $C := [A, B]$, and consider on \mathcal{K}^\perp the norm⁵*

$$\|h\|_{\mathcal{H}^1}^2 := \|h\|^2 + \|Ah\|^2 + \|Ch\|^2.$$

Suppose the following hold:

- i. A and A^* commute with C ;
- ii. $[A, A^*]$ is relatively bounded with respect to I and A ;
- iii. $[B, C]$ is relatively bounded with respect to A , A^2 , C , and AC ;
- iv. $A^*A + C^*C$ is κ -coercive for some $\kappa > 0$.

Then \mathcal{T} is hypocoercive in $\mathcal{H}^1/\mathcal{K}$: there exist constants $c, \lambda > 0$ such that

$$\|e^{-t\mathcal{T}}\|_{\mathcal{H}^1/\mathcal{K} \rightarrow \mathcal{H}^1/\mathcal{K}} \leq ce^{-\lambda t}.$$

Remark 6.1. Let \mathcal{K} be the kernel of \mathcal{T} and note that $\text{Ker}(A^*A) = \text{Ker}(A)$ and $\mathcal{K} = \text{Ker}(A) \cap \text{Ker}(B)$. Suppose $\text{Ker}A \subset \text{Ker}B$; then $\text{Ker}\mathcal{T} = \text{Ker}A$. In this case, the coercivity of \mathcal{T} is equivalent to the coercivity of A^*A . Hypocoercivity theory becomes relevant when A^*A is coercive and \mathcal{T} is not. This is the case when the operators A^*A and B do not commute; if they did, then $e^{-t\mathcal{T}} = e^{-tA^*A}e^{-tB}$. Therefore, since e^{-tB} is norm-preserving (B is antisymmetric), we would have $\|e^{-t\mathcal{T}}\| = \|e^{-tA^*A}\|$. The operators A and B do not commute for the Langevin dynamics; see Eq. (6.21).

We can use Theorem 6.2 to prove exponentially fast convergence to equilibrium for the Langevin dynamics. As mentioned earlier, we will study the Fokker–Planck equation in the form (6.14) and consider this equation in the Hilbert space $L^2(\rho_\beta)$. The space S can be taken to be the Schwartz space of smooth rapidly decaying functions. For simplicity, we will consider the equation in one dimension and set $\gamma = \beta = 1$. We have

$$\mathcal{L}_{kin} = B - A^*A,$$

where

$$B = -p\partial_q + \partial_q V \partial_p, \quad A = \partial_p \quad \text{and} \quad A^* = -\partial_p + p. \quad (6.35)$$

⁵ One can prove that the space \mathcal{K}^\perp is the same irrespective of whether we consider the scalar product $\langle \cdot, \cdot \rangle$ of \mathcal{H} or the scalar product $\langle \cdot, \cdot \rangle_{\mathcal{H}^1}$ associated with the norm $\|\cdot\|_{\mathcal{H}^1}$.

The kernel \mathcal{K} of the operator \mathcal{L}_{kin} consists of constants: as we have already observed in Remark 6.1, $\text{Ker}\mathcal{L}_{kin} = \text{Ker}A \cap \text{Ker}B$. $\text{Ker}A = \{\text{functions of } q\}$ and $\text{Ker}B = \{\text{functions of the Hamiltonian } H(q, p)\}$. The intersection of these two sets consists of constant functions. We have $C = [A, B] = -\partial_q$; see (6.21). Consequently, the norm \mathcal{H}^1 is

$$\|h\|_{H^1}^2 = \|h\|_{L^2(\rho_\beta)}^2 + \|\partial_p h\|_{L^2(\rho_\beta)}^2 + \|\partial_q h\|_{L^2(\rho_\beta)}^2. \quad (6.36)$$

This is the norm of the weighted space $H^1(\rho_\beta)$.

The commutators needed to check the assumptions of Theorem 6.2 were calculated in (6.21) and (6.23), with $C = -Y$:

$$[A, C] = [A^*, C] = 0, \quad [A, A^*] = I, \quad (6.37)$$

where I denotes the identity operator and

$$[B, C] = \partial_q^2 V(q) \partial_p. \quad (6.38)$$

In order to proceed, we need to make assumptions on the potential V .

Assumption 6.5 *The potential $V \in C^\infty(\mathbb{R})$ satisfies a Poincaré inequality with constant λ .*

Conditions on the potential so that it satisfies a Poincaré inequality are discussed in Sect. 4.5. In particular, sufficient conditions are

$$\lim_{|x| \rightarrow +\infty} \left(\frac{|V'(x)|^2}{2} - V''(x) \right) = \infty, \quad (6.39)$$

or the Bakry–Emery criterion (4.79) $V''(x) \geq \lambda$ (written here in one dimension). Assumption 6.5 implies that for all functions $h \in H^1(e^{-V(q)})$,⁶ we have

$$\int_{\mathbb{R}} |\partial_q h|^2 e^{-V(q)} dq \geq \lambda \left[\int_{\mathbb{R}} h^2 e^{-V(q)} dq - \left(\int_{\mathbb{R}} h e^{-V(q)} dq \right)^2 \right]$$

(see (4.80)). Furthermore, we will assume that we can control the Hessian of the potential in terms of its gradient; compare also with (6.39).

Assumption 6.6 *There exists a constant $C > 0$ such that the potential $V \in C^\infty(\mathbb{R})$ satisfies*

$$|\partial_q^2 V| \leq C(1 + |\partial_q V|). \quad (6.40)$$

We will need the following estimate.

Lemma 6.7. *Let Assumptions 6.5 and 6.6 hold. Then, for all $f \in H^1(\rho_\beta)$, there exists a constant $C > 0$ such that*

⁶ We assume that $\int e^{-V} dq = 1$.

$$\|\partial_q^2 V f\|_{L^2(\rho_\beta)}^2 \leq C \left(\|f\|_{L^2(\rho_\beta)}^2 + \|\partial_q f\|_{L^2(\rho_\beta)}^2 \right). \quad (6.41)$$

Proof. We will prove that

$$\|\partial_q V f\|_{L^2(\rho_\beta)}^2 \leq C \left(\|f\|_{L^2(\rho_\beta)}^2 + \|\partial_q f\|_{L^2(\rho_\beta)}^2 \right). \quad (6.42)$$

Estimate (6.41) then follows from (6.42) and Assumption 6.6. To prove (6.42), it is sufficient to show that for all $f \in H^1(e^{-V})$, we have

$$\int |\partial_q V|^2 f^2 e^{-V} \leq C \left(\int f^2 e^{-V} + \int |\partial_q f|^2 e^{-V} \right). \quad (6.43)$$

The estimate (6.42) then follows for functions $f \in H^1(\rho_\beta)$ on multiplying by $Z^{-1} \exp(-\frac{p^2}{2})$ and integrating with respect to p . By density, it is sufficient to prove (6.43) for smooth functions that decay sufficiently fast to infinity. We use the fact that $\partial_q e^{-V} = -\partial_q V e^{-V}$ and perform an integration by parts to deduce that

$$\int |\partial_q V|^2 f^2 e^{-V} = \int f^2 \partial_q^2 V e^{-V} + 2 \int f \partial_q f \partial_q V e^{-V}. \quad (6.44)$$

Using the Cauchy–Schwarz inequality, together with Assumption 6.6 and Cauchy’s inequality, we obtain

$$\begin{aligned} \int f^2 \partial_q^2 V e^{-V} &\leq C \int f^2 (1 + \partial_q V) e^{-V} \\ &\leq C \int f^2 e^{-V} + C \int f (f \partial_q V) e^{-V} \\ &\leq C \int f^2 e^{-V} + \frac{C\alpha}{2} \int f^2 (\partial_q V)^2 e^{-V} + \frac{C}{2\alpha} \int f^2 e^{-V} \\ &\leq \tilde{C} \int f^2 e^{-V} + \frac{1}{4} \int f^2 (\partial_q V)^2 e^{-V}, \end{aligned}$$

upon choosing $\alpha = 1/(2\sqrt{2}C)$ in Cauchy’s inequality, where C is the constant in (6.40). Similarly, the Cauchy–Schwarz and Cauchy inequalities can be used on the second term on the right-hand side of (6.44) to obtain

$$2 \int f \partial_q f \partial_q V e^{-V} \leq \frac{1}{4} \int |\partial_q V|^2 f^2 e^{-V} + 4 \int |\partial_q f|^2 e^{-V}.$$

Consequently,

$$\int |\partial_q V|^2 f^2 e^{-V} \leq \frac{1}{2} \int |\partial_q V|^2 f^2 e^{-V} + \tilde{C} \left(\int f^2 e^{-V} + \int |\partial_q f|^2 e^{-V} \right),$$

from which (6.43) follows. \square

The standard example of a potential that satisfies these assumptions is the harmonic potential. We will study the Langevin equation in a harmonic potential in Sect. 6.3.

We can now use Theorem 6.2 to prove exponentially fast convergence to the Maxwell–Boltzmann distribution for the Langevin dynamics.

Theorem 6.3. *Assume that the potential satisfies Assumptions 6.5 and 6.6. Then there exist constants $C, \lambda > 0$ such that for all $h_0 \in H^1(\rho_\beta)$,*

$$\left\| e^{t\mathcal{L}_{kin}} h_0 - \int h_0 \rho_\beta \right\|_{H^1(\rho_\beta)} \leq C e^{-\lambda t} \|h_0\|_{H^1(\rho_\beta)}. \quad (6.45)$$

Proof. We will use Theorem 6.2. We need to check that conditions (i) to (iv) of the theorem are satisfied. Condition (i), that A and A^* commute with C , follows from (6.37). Condition (ii), that $[A, A^*] = I$ is relatively bounded with respect to I and $A = \partial_p$, is automatically satisfied. From (6.38), it follows that condition (iii) requires that $\partial_q^2 V \partial_p$ be relatively bounded with respect to ∂_p , ∂_p^2 , ∂_q , and ∂_{qp}^2 . For this, it is sufficient to show that for all $f \in H^2(\rho_\beta)$, there exists a constant $C > 0$ such that

$$\int |\partial_q^2 V \partial_p f|^2 \rho_\beta \leq C \left(\int |\partial_p f|^2 \rho_\beta + \int |\partial_p \partial_q f|^2 \rho_\beta \right). \quad (6.46)$$

This follows by applying Lemma 6.7 to $\partial_p f$. Now we turn to condition (iv), that the operator $\widehat{\mathcal{L}} = A^* A + C^* C$ is coercive. This operator has the form

$$\widehat{\mathcal{L}} = p \partial_p - \partial_p^2 + \partial_q V \partial_q - \partial_q^2. \quad (6.47)$$

To show that $\widehat{\mathcal{L}}$ is coercive, we need to show that (6.33) holds for all mean-zero $h \in H^1(\rho_\beta)$. equivalently, we need to show that there exists a constant $\widehat{\lambda}$ such that for all $h \in H^1(\rho_\beta)$, we have

$$\int_{\mathbb{R}^2} \left(|\partial_q h|^2 + |\partial_p h|^2 \right) \rho_\beta \geq \widehat{\lambda} \left[\int_{\mathbb{R}^2} h^2 \rho_\beta - \left(\int_{\mathbb{R}^2} h \rho_\beta \right)^2 \right]. \quad (6.48)$$

In other words, we need to show that the measure $\rho_\beta dp dq = Z^{-1} \exp(-H(p, q)) dp dq$ satisfies a Poincaré inequality with constant $\widehat{\lambda}$. This follows from the fact that this is a product measure,

$$\rho_\beta dp dq = Z^{-1} \left(e^{-\frac{p}{2}} dp \right) \left(e^{-V(q)} dq \right) dp dq,$$

the fact that the Gaussian measure $Z^{-1} \exp\left(-\frac{p}{2}\right) dp$ satisfies a Poincaré inequality—see Sect. 4.4—and Assumption 6.5. \square

Notice that for the proof of exponentially fast convergence to equilibrium for the Langevin dynamics, we have only to check conditions on the potential V . In fact,

Assumption 6.5 that e^{-V} satisfies a Poincaré inequality is precisely the assumption needed in order to prove exponentially fast convergence to equilibrium for the reversible dynamics $dX_t = -\nabla V(X_t) dt + \sqrt{2} dW_t$. See Sect. 4.5.

Note also that we can think of the term C^*C as a kind of regularization. Indeed, a standard PDE approach to dealing with the lack of second-order derivatives with respect to q in the generator of the Langevin dynamics is to add second-order derivatives in q , obtain good a priori estimates, and then pass to the limit, where we remove the regularization. Simply adding the Laplacian Δ_q with respect to q leads to an operator that does not preserve the invariant measure of the dynamics: e^{-H} is not in the null space of the adjoint. On the other hand, the addition of the term $-\nabla_q V \cdot \nabla_q + \Delta_q$ preserves the invariant measure. This leads to the modified dynamics

$$dq_t^\delta = p_t^\delta dt + \delta \left(-\nabla_q V(q_t^\delta) dt + \sqrt{2} dW_t^q \right), \quad dp_t^\delta = -\nabla_q V(q_t^\delta) dt - p_t^\delta dt + \sqrt{2} dW_t^p, \quad (6.49)$$

where W_t^q, W_t^p are independent d -dimensional Brownian motions. The generator of the modified dynamics q_t^δ, p_t^δ is

$$\mathcal{L}^\delta = \mathcal{L} + \delta(-\nabla_q V(q) \cdot \nabla_q + \Delta_q), \quad (6.50)$$

where \mathcal{L} denotes the generator of the Langevin dynamics. Note that the regularization is precisely the term $-C^*C$.

6.3 The Langevin Equation in a Harmonic Potential

The harmonic potential is an example of a potential for which the assumptions of Theorem 6.3 apply. In fact, for such a potential, we can obtain more detailed information, since we can solve the Langevin equation analytically, and we can also calculate the spectrum of the generator of the Langevin dynamics. This enables us, for example, to calculate the exponent in the exponential estimate (6.45). We consider the harmonic potential

$$V(q) = \frac{1}{2} \omega_0^2 q^2. \quad (6.51)$$

The Langevin equation is

$$\dot{q} = -\omega_0^2 q - \gamma \dot{q} + \sqrt{2\gamma\beta^{-1}} \dot{W}, \quad (6.52)$$

or

$$dq_t = p_t dt, \quad dp_t = -\omega_0^2 q_t dt - \gamma p_t dt + \sqrt{2\gamma\beta^{-1}} dW_t. \quad (6.53)$$

This is a linear stochastic differential equation that can be solved analytically. By setting $z_t = [q_t \ p_t]^T$, we can write (6.53) as

$$dz_t = Az_t dt + \sigma dW_t, \quad (6.54)$$

where $A \in \mathbb{R}^{2 \times 2}$ and $\sigma \in \mathbb{R}^{2 \times 1}$ are

$$A = \begin{bmatrix} 0 & 1 \\ -\omega_0^2 & -\gamma \end{bmatrix}, \quad \sigma = \begin{bmatrix} 0 \\ \sqrt{2\gamma\beta^{-1}} \end{bmatrix}. \quad (6.55)$$

We can use the general formulas presented in Sect. 3.7 to obtain a solution to this equation and to calculate the mean and the autocorrelation function.

The generator of the Langevin dynamics (6.53) is

$$\mathcal{L} = p\partial_q - \omega_0^2 q\partial_p + \gamma(-p\partial_p + \beta^{-1}\partial_p^2). \quad (6.56)$$

Since the potential $V(q)$ is quadratic, the invariant distribution is Gaussian in both q and p :

$$\rho_\beta(q, p) = \frac{\beta\omega_0}{2\pi} e^{-\frac{\beta}{2}p^2 - \frac{\beta\omega_0^2}{2}q^2}. \quad (6.57)$$

Using now (6.22), we can write the generator in the form

$$\mathcal{L} = -b^+a^- + b^-a^+ - \gamma a^+a^-, \quad (6.58)$$

with a^\pm, b^\pm the creation and annihilation operators corresponding to the Ornstein–Uhlenbeck process in both p and q :

$$a^- = \beta^{-1/2}\partial_p, \quad a^+ = -\beta^{-1/2}\partial_p + \beta^{1/2}p \quad (6.59)$$

and

$$b^- = \omega_0^{-1}\beta^{-1/2}\partial_q, \quad b^+ = -\omega_0^{-1}\beta^{-1/2}\partial_q + \beta^{1/2}\omega_0 q. \quad (6.60)$$

We have

$$a^+a^- = -\beta^{-1}\partial_p^2 + p\partial_p \quad \text{and} \quad \omega_0^2 b^+b^- = -\beta^{-1}\partial_q^2 + \omega_0^2 q\partial_q.$$

The operator

$$\widehat{\mathcal{L}} = -a^+a^- - \omega_0^2 b^+b^- \quad (6.61)$$

is the generator of the Ornstein–Uhlenbeck process in two dimensions. The invariant distribution of the corresponding dynamics is the Gaussian distribution (6.57). Notice that this is precisely the operator $-A^*A - C^*C$ that appears in Theorem 6.2, Eq. (6.47). The eigenvalues and eigenfunctions of (6.61) were calculated in Sect. 4.4. We will calculate the spectrum of the generator (6.58) by introducing a transformation that maps (6.58) to (6.61).

The operators a^\pm, b^\pm satisfy the commutation relations

$$[a^+, a^-] = -1, \quad [b^+, b^-] = -1, \quad [a^\pm, b^\pm] = 0; \quad (6.62)$$

see (6.21) and (6.23). We want to find first-order differential operators c^\pm and d^\pm such that the generator (6.58) takes the form (6.61)

$$\mathcal{L} = -Cc^+c^- - Dd^+d^-, \quad (6.63)$$

where the operators c^\pm and d^\pm satisfy the *canonical commutation relations*

$$[c^+, c^-] = -1, \quad [d^+, d^-] = -1, \quad [c^\pm, d^\pm] = 0. \quad (6.64)$$

The operators c^\pm and d^\pm should be given as linear combinations of the old operators a^\pm and b^\pm . From the structure of the generator \mathcal{L} (6.58), the decoupled form (6.63), and the commutation relations (6.64) and (6.62), we conclude that c^\pm and d^\pm should be of the form

$$c^+ = \alpha_{11}a^+ + \alpha_{12}b^+, \quad c^- = \alpha_{21}a^- + \alpha_{22}b^-, \quad (6.65a)$$

$$d^+ = \beta_{11}a^+ + \beta_{12}b^+, \quad d^- = \beta_{21}a^- + \beta_{22}b^-. \quad (6.65b)$$

Observe that c^- and d^- are not the adjoints of c^+ and d^+ . If we substitute now these equations into (6.63) and equate it with (6.58) and into the commutation relations (6.64), we obtain a system of equations for the coefficients $\{\alpha_{ij}\}$, $\{\beta_{ij}\}$. In order to write down the formulas for these coefficients, it is convenient to introduce the eigenvalues of the deterministic problem

$$\dot{z}_t = Az_t;$$

see (6.54). The solution of this equation, with $z_t = [q_t \ p_t]^T$, is

$$q_t = C_1 e^{-\lambda_1 t} + C_2 e^{-\lambda_2 t},$$

where λ_1, λ_2 are the eigenvalues of the matrix A defined in (6.55):

$$\lambda_{1,2} = \frac{\gamma \pm \delta}{2}, \quad \delta = \sqrt{\gamma^2 - 4\omega_0^2}. \quad (6.66)$$

The eigenvalues satisfy the relations

$$\lambda_1 + \lambda_2 = \gamma, \quad \lambda_1 - \lambda_2 = \delta, \quad \lambda_1 \lambda_2 = \omega_0^2. \quad (6.67)$$

Proposition 6.8. *Let \mathcal{L} be the generator (6.58), and let c^\pm, d^\pm be the operators*

$$c^+ = \frac{1}{\sqrt{\delta}} \left(\sqrt{\lambda_1}a^+ + \sqrt{\lambda_2}b^+ \right), \quad c^- = \frac{1}{\sqrt{\delta}} \left(\sqrt{\lambda_1}a^- - \sqrt{\lambda_2}b^- \right), \quad (6.68a)$$

$$d^+ = \frac{1}{\sqrt{\delta}} \left(\sqrt{\lambda_2}a^+ + \sqrt{\lambda_1}b^+ \right), \quad d^- = \frac{1}{\sqrt{\delta}} \left(-\sqrt{\lambda_2}a^- + \sqrt{\lambda_1}b^- \right). \quad (6.68b)$$

Then c^\pm, d^\pm satisfy the canonical commutation relations (6.64) as well as

$$[\mathcal{L}, c^\pm] = -\lambda_1 c^\pm, \quad [\mathcal{L}, d^\pm] = -\lambda_2 d^\pm. \quad (6.69)$$

Furthermore, the operator \mathcal{L} can be written in the form

$$\mathcal{L} = -\lambda_1 c^+ c^- - \lambda_2 d^+ d^-. \quad (6.70)$$

Proof. First, we check the commutation relations:

$$\begin{aligned}[c^+, c^-] &= \frac{1}{\delta} (\lambda_1[a^+, a^-] - \lambda_2[b^+, b^-]) \\ &= \frac{1}{\delta}(-\lambda_1 + \lambda_2) = -1.\end{aligned}$$

Similarly,

$$\begin{aligned}[d^+, d^-] &= \frac{1}{\delta} (-\lambda_2[a^+, a^-] + \lambda_1[b^+, b^-]) \\ &= \frac{1}{\delta}(\lambda_2 - \lambda_1) = -1.\end{aligned}$$

Clearly, we have that

$$[c^+, d^+] = [c^-, d^-] = 0.$$

Furthermore,

$$\begin{aligned}[c^+, d^-] &= \frac{1}{\delta} \left(-\sqrt{\lambda_1 \lambda_2} [a^+, a^-] + \sqrt{\lambda_1 \lambda_2} [b^+, b^-] \right) \\ &= \frac{1}{\delta} (\sqrt{\lambda_1 \lambda_2} - \sqrt{\lambda_1 \lambda_2}) = 0.\end{aligned}$$

Finally,

$$\begin{aligned}[\mathcal{L}, c^+] &= -\lambda_1 c^+ c^- c^+ + \lambda_1 c^+ c^+ c^- \\ &= -\lambda_1 c^+ (1 + c^+ c^-) + \lambda_1 c^+ c^+ c^- \\ &= -\lambda_1 c^+ (1 + c^+ c^-) + \lambda_1 c^+ c^+ c^- \\ &= -\lambda_1 c^+,\end{aligned}$$

and similarly for the other equations in (6.69). Now we calculate

$$\begin{aligned}\mathcal{L} &= -\lambda_1 c^+ c^- - \lambda_2 d^+ d^- \\ &= -\frac{\lambda_2^2 - \lambda_1^2}{\delta} a^+ a^- + 0 b^+ b^- + \frac{\sqrt{\lambda_1 \lambda_2}}{\delta} (\lambda_1 - \lambda_2) a^+ b^- + \frac{1}{\delta} \sqrt{\lambda_1 \lambda_2} (-\lambda_1 + \lambda_2) b^+ a^- \\ &= -\gamma a^+ a^- - \omega_0^2 (b^+ a^- - a^+ b^-),\end{aligned}$$

which is precisely (6.58). In the above calculation, we used (6.67). \square

Using now (6.70), we can calculate the eigenvalues and eigenfunctions of \mathcal{L} . From the results of Sect. 4.4, we expect that the eigenvalues are integer linear combinations of the eigenvalues λ_1 and λ_2 of the matrix A and that the eigenfunctions are tensor products of Hermite polynomials; see Theorem 4.2, in particular 4.55 and (4.57).

Theorem 6.4. *The eigenvalues and eigenfunctions of the generator (6.58) are*

$$\lambda_{nm} = \lambda_1 n + \lambda_2 m = \frac{1}{2} \gamma(n+m) + \frac{1}{2} \delta(n-m), \quad n, m = 0, 1, \dots, \quad (6.71)$$

and

$$\phi_{nm}(q, p) = \frac{1}{\sqrt{n!m!}} (c^+)^n (d^+)^m \mathbf{1}, \quad n, m = 0, 1, \dots \quad (6.72)$$

Proof. We have

$$\begin{aligned} [\mathcal{L}, (c^+)^2] &= \mathcal{L}(c^+)^2 - (c^+)^2 \mathcal{L} \\ &= (c^+ \mathcal{L} - \lambda_1 c^+) c^+ - c^+ (\mathcal{L} c^+ + \lambda_1 c^+) \\ &= -2\lambda_1 (c^+)^2, \end{aligned}$$

and similarly, $[\mathcal{L}, (d^+)^2] = -2\lambda_1 (d^+)^2$. An induction argument now shows that (see Exercise 6)

$$[\mathcal{L}, (c^+)^n] = -n\lambda_1 (c^+)^n \quad \text{and} \quad [\mathcal{L}, (d^+)^m] = -m\lambda_1 (d^+)^m. \quad (6.73)$$

We use (6.73) to calculate

$$\begin{aligned} \mathcal{L}(c^+)^n (d^+)^m \mathbf{1} &= (c^+)^n \mathcal{L}(d^+)^m \mathbf{1} - n\lambda_1 (c^+)^n (d^+)^m \mathbf{1} \\ &= (c^+)^n (d^+)^m \mathcal{L} \mathbf{1} - m\lambda_2 (c^+)^n (d^+)^m \mathbf{1} - n\lambda_1 (c^+)^n (d^+)^m \mathbf{1} \\ &= -n\lambda_1 (c^+)^n (d^+)^m \mathbf{1} - m\lambda_2 (c^+)^n (d^+)^m \mathbf{1}, \end{aligned}$$

from which (6.71) and (6.72) follow. \square

Remark 6.2. In terms of the operators a^\pm, b^\pm , the eigenfunctions of \mathcal{L} are

$$\phi_{nm} = \sqrt{n!m!} \delta^{-\frac{n+m}{2}} \lambda_1^{n/2} \lambda_2^{m/2} \sum_{\ell=0}^n \sum_{k=0}^m \frac{1}{k!(m-k)!\ell!(n-\ell)!} \left(\frac{\lambda_1}{\lambda_2} \right)^{\frac{k-\ell}{2}} (a^+)^{n+m-k-\ell} (b^+)^{\ell+k} \mathbf{1}.$$

The first few eigenfunctions are

$$\phi_{00} = 1, \quad \phi_{10} = \frac{\sqrt{\beta} (\sqrt{\lambda_1} p + \sqrt{\lambda_2} \omega_0 q)}{\sqrt{\delta}}, \quad \phi_{01} = \frac{\sqrt{\beta} (\sqrt{\lambda_2} p + \sqrt{\lambda_1} \omega_0 q)}{\sqrt{\delta}},$$

$$\phi_{11} = \frac{-2\sqrt{\lambda_1}\sqrt{\lambda_2} + \sqrt{\lambda_1}\beta p^2\sqrt{\lambda_2} + \beta p\lambda_1\omega_0 q + \omega_0\beta q\lambda_2 p + \sqrt{\lambda_2}\omega_0^2\beta q^2\sqrt{\lambda_1}}{\delta},$$

$$\phi_{20} = \frac{-\lambda_1 + \beta p^2\lambda_1 + 2\sqrt{\lambda_2}\beta p\sqrt{\lambda_1}\omega_0 q - \lambda_2 + \omega_0^2\beta q^2\lambda_2}{\sqrt{2}\delta},$$

$$\phi_{02} = \frac{-\lambda_2 + \beta p^2\lambda_2 + 2\sqrt{\lambda_2}\beta p\sqrt{\lambda_1}\omega_0 q - \lambda_1 + \omega_0^2\beta q^2\lambda_1}{\sqrt{2}\delta}.$$

Note that these eigenfunctions have not been orthonormalized.

As we already know, the first eigenvalue, corresponding to the constant eigenfunction, is

$$\lambda_{00} = 0.$$

The generator \mathcal{L} is not self-adjoint, and we do not expect its eigenvalues to be real. Indeed, whether the eigenvalues are real depends on the sign of the discriminant $\Delta = \gamma^2 - 4\omega_0^2$. In the *underdamped regime*, $\gamma < 2\omega_0$, the eigenvalues are complex:

$$\lambda_{nm} = \frac{1}{2}\gamma(n+m) + \frac{1}{2}i\sqrt{-\gamma^2 + 4\omega_0^2}(n-m), \quad \gamma < 2\omega_0.$$

This it to be expected, since in the underdamped regime, the dynamics are dominated by the deterministic Hamiltonian dynamics, which give rise to the antisymmetric Liouville operator. We set $\omega = \sqrt{(4\omega_0^2 - \gamma^2)}$, i.e., $\delta = 2i\omega$. The eigenvalues can be written as

$$\lambda_{nm} = \frac{\gamma}{2}(n+m) + i\omega(n-m).$$

In Fig. 6.1, we present the first few eigenvalues of \mathcal{L} in the underdamped regime. The eigenvalues are contained in a cone in the right half of the complex plane. The cone is determined by

$$\lambda_{n0} = \frac{\gamma}{2}n + i\omega n \quad \text{and} \quad \lambda_{0m} = \frac{\gamma}{2}m - i\omega m.$$

The eigenvalues along the diagonal are real:

$$\lambda_{nn} = \gamma n.$$

On the other hand, in the *overdamped regime*, $\gamma \geq 2\omega_0$, all eigenvalues are real:

$$\lambda_{nm} = \frac{1}{2}\gamma(n+m) + \frac{1}{2}\sqrt{\gamma^2 - 4\omega_0^2}(n-m), \quad \gamma \geq 2\omega_0.$$

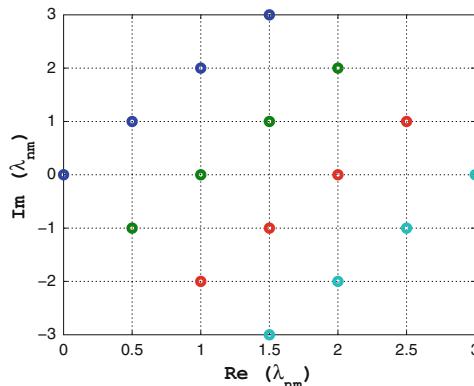


Fig. 6.1 First few eigenvalues of \mathcal{L} for $\gamma = \omega = 1$

In fact, in the overdamped limit $\gamma \rightarrow +\infty$ (which we will study in Sect. 6.5.1), the eigenvalues of the generator \mathcal{L} converge to the eigenvalues of the generator of the Ornstein–Uhlenbeck process:

$$\lambda_{nm} = \gamma n + \frac{\omega_0^2}{\gamma} (n - m) + O(\gamma^{-3}).$$

This is consistent with the fact that in this limit, the solution of the Langevin equation converges to the solution of the Ornstein–Uhlenbeck SDE. See Sect. 6.5.1 for details.

The eigenfunctions of \mathcal{L} do not form an orthonormal basis in $L^2(\rho_\beta)$, since \mathcal{L} is not a self-adjoint operator. Using the eigenfunctions/eigenvalues of \mathcal{L} , we can easily calculate the eigenfunctions/eigenvalues of the $L^2(\rho_\beta)$ adjoint of \mathcal{L} . From Lemma 6.2, we have that the adjoint operator is

$$\begin{aligned} \mathcal{L}_{kin} &:= -B + \gamma S \\ &= -\omega_0^2(b^+a^- - b^-a^+) + \gamma a^+a^- \\ &= -\lambda_1(c^-)^*(c^+)^* - \lambda_2(d^-)^*(d^+)^*, \end{aligned} \quad (6.74)$$

where

$$(c^+)^* = \frac{1}{\sqrt{\delta}} \left(\sqrt{\lambda_1}a^- + \sqrt{\lambda_2}b^- \right), \quad (6.75a)$$

$$(c^-)^* = \frac{1}{\sqrt{\delta}} \left(\sqrt{\lambda_1}a^+ - \sqrt{\lambda_2}b^+ \right), \quad (6.75b)$$

$$(d^+)^* = \frac{1}{\sqrt{\delta}} \left(\sqrt{\lambda_2}a^- + \sqrt{\lambda_1}b^- \right), \quad (6.75c)$$

$$(d^-)^* = \frac{1}{\sqrt{\delta}} \left(-\sqrt{\lambda_2}a^+ + \sqrt{\lambda_1}b^+ \right). \quad (6.75d)$$

Here \mathcal{L}_{kin} has the same eigenvalues as \mathcal{L} :

$$-\mathcal{L}_{kin}\psi_{nm} = \lambda_{nm}\psi_{nm},$$

where λ_{nm} are given by (6.71). The eigenfunctions are

$$\psi_{nm} = \frac{1}{\sqrt{n!m!}} ((c^-)^*)^n ((d^-)^*)^m \mathbf{1}. \quad (6.76)$$

Proposition 6.9. *The eigenfunctions of \mathcal{L} and \mathcal{L}_{kin} satisfy the biorthonormality relation*

$$\int \int \phi_{nm} \psi_{\ell k} \rho_\beta \, dp \, dq = \delta_{n\ell} \delta_{mk}. \quad (6.77)$$

Proof. We will use formulas (6.62). Note that from the third and fourth of these equations together with the fact that $c^- \mathbf{1} = d^- \mathbf{1} = 0$, we can conclude that (for $n \geq \ell$)

$$(c^-)^\ell (c^+)^n \mathbf{1} = n(n-1) \dots (n-\ell+1) (c^+)^{n-\ell}. \quad (6.78)$$

We have

$$\begin{aligned} \int \int \phi_{nm} \psi_{\ell k} \rho_\beta \, dp \, dq &= \frac{1}{\sqrt{n!m!\ell!k!}} \int \int ((c^+)^n ((d^+))^m \mathbf{1} ((c^-)^*)^\ell ((d^-)^*)^k \mathbf{1} \rho_\beta \, dp \, dq \\ &= \frac{n(n-1) \dots (n-\ell+1) m(m-1) \dots (m-k+1)}{\sqrt{n!m!\ell!k!}} \times \\ &\quad \int \int ((c^+)^{n-\ell} ((d^+))^{m-k} \mathbf{1} \rho_\beta \, dp \, dq \\ &= \delta_{n\ell} \delta_{mk}, \end{aligned}$$

since all eigenfunctions average to 0 with respect to ρ_β . \square

From the eigenfunctions of \mathcal{L}_{kin} , we can obtain the eigenfunctions of the Fokker–Planck operator. The Fokker–Planck operator is

$$\mathcal{L}^* \cdot = -p \partial_q \cdot + \omega_0^2 q \partial_p \cdot + \gamma (\partial_p(p \cdot) + \beta^{-1} \partial_p^2 \cdot). \quad (6.79)$$

Using now

$$\mathcal{L}^*(f \rho_\beta) = \rho \mathcal{L}_{kin} f,$$

(see (6.14)), we conclude that the Fokker–Planck operator has the same eigenvalues as those of \mathcal{L} and \mathcal{L}_{kin} . The eigenfunctions are

$$\psi_{nm}^* = \rho_\beta \phi_{nm} = \rho_\beta \frac{1}{\sqrt{n!m!}} ((c^-)^*)^n ((d^-)^*)^m \mathbf{1}. \quad (6.80)$$

6.4 Expansion in Hermite Polynomials

The Fokker–Planck equation (6.5) can be solved analytically for very few potentials, essentially only for those that are constant, linear, or quadratic. In most other cases, in order to calculate time-dependent statistical quantities of the solution to the Langevin equation (q_t, p_t), we need either to resort to Monte Carlo simulations, based on the numerical solution of the Langevin SDE, see Sect. 5.2, or use a numerical method for partial differential equations, such as finite differences or a spectral method. For low-dimensional problems, the numerical solution of these PDEs can lead to computationally tractable problems.

We can use the structure of the generator of the Langevin dynamics and the fact that the invariant distribution (6.9) of the dynamics is a product distribution to develop a spectral numerical method for solving the Fokker–Planck equation.

In particular, we will expand the solution into Hermite polynomials in the momentum, the eigenfunctions of the generator of the Ornstein–Uhlenbeck process. We can then expand the resulting equations in an appropriate basis in q and truncate. This leads to a finite system of ordinary differential equations that can be solved numerically.

Using the transformation (6.16), we can map the Fokker–Planck equation to the backward Kolmogorov equation. We will consider the initial value problem for this equation

$$\frac{\partial u}{\partial t} = \mathcal{L}u, \quad (6.81a)$$

$$u(q, p, 0) = \phi(q, p). \quad (6.81b)$$

For simplicity, we consider the problem in one dimension. Higher-dimensional problems can be handled similarly, in principle, by taking tensor products of the basis functions in which we are expanding the solution. As we have done earlier in this chapter, we will consider (6.81) in the Hilbert space \mathcal{H} defined in (6.18).

The generator \mathcal{L} is written in the form (6.22):

$$\mathcal{L} = -b^+a^- + b^-a^+ - \gamma a^+a^-, \quad (6.82)$$

where

$$a^- = \beta^{-1/2}\partial_p, \quad b^- = \beta^{-1/2}\partial_q \text{ and} \quad (6.83)$$

$$a^+ = -\beta^{-1/2}\partial_p + \beta^{1/2}p, \quad b^+ = -\beta^{-1/2}\partial_q + \beta^{1/2}\partial_q V(q). \quad (6.84)$$

We remark that since the operators a^\pm, b^\pm commute, the antisymmetric part of \mathcal{L} can be written in different forms, for example $-a^+b^- + a^-b^+$. The Maxwell distribution ρ_β is the product of two distributions in p and q , with the dependence in p being Gaussian:

$$\frac{1}{Z}e^{-\beta H(p, q)} = \left(\frac{1}{\sqrt{2\pi\beta^{-1}}}e^{-\beta\frac{1}{2}|p|^2} \right) \left(\frac{1}{\int_{\mathbb{R}} e^{-\beta V(q)} dq}e^{-\beta V(q)} \right).$$

As we have seen in Sect. 4.4, the space $L^2(Z^{-1}e^{-\beta\frac{1}{2}|p|^2})$ is spanned by the Hermite polynomials. Consequently, we can expand the solution of (6.81a) into the basis of Hermite polynomials:

$$u(p, q, t) = \sum_{n=0}^{\infty} u_n(q, t) f_n(p), \quad (6.85)$$

where $f_n(p) = 1/\sqrt{n!}H_n(\sqrt{\beta}p)$. We have (see Sect. 4.4)

$$a^+a^-f_n = nf_n, \quad a^+f_n = \sqrt{n+1}f_{n+1}, \quad a^-f_n = \sqrt{n}f_{n-1}.$$

Thus

$$\mathcal{L}u_n(q, t)f_n(p) = -\gamma n u_n f_n - \sqrt{n}b^+u_n f_{n-1} + \sqrt{n+1}b^-u_n f_{n+1}.$$

Consequently, by renaming indices in the second and third terms on the right-hand side (we are using the convention $f_{-1} = 0, u_{-1} = 0$), we obtain

$$\begin{aligned}\mathcal{L}u &= \sum_{n=0}^{+\infty} \left(-\gamma mu_n f_n - \sqrt{n} b^+ u_n f_{n-1} + \sqrt{n+1} b^- u_n f_{n+1} \right) \\ &= \sum_{n=0}^{\infty} \left(-\gamma mu_n - \sqrt{n+1} b^+ u_{n+1} + \sqrt{n} b^- u_{n-1} \right) f_n.\end{aligned}$$

Using the orthonormality of the Hermite polynomials (the eigenfunctions of the symmetric part of \mathcal{L}) in $L^2 \left(Z^{-1} e^{-\beta \frac{1}{2} |p|^2} \right)$, we obtain the following set of equations:

$$\frac{\partial u_n}{\partial t} = -\gamma mu_n - \sqrt{n+1} b^+ u_{n+1} + \sqrt{n} b^- u_{n-1}, \quad n = 0, 1, \dots, \quad (6.86)$$

together with the initial conditions $u_n(q, 0) = \phi_n(q)$, $n = 0, \dots$, obtained by expanding the initial condition $\phi(q, p) \in L^2(\rho_\beta)$ into the Hermite basis. This is an infinite system of linear hyperbolic partial differential equations for the coefficients $\{u_n(q, t)\}_{n=0}^{\infty}$. It is equivalent to the backward Kolmogorov equation (6.81). Truncating this system and expanding the coefficients into an appropriate orthonormal basis in q leads to a finite system of ordinary differential equations that can then be solved numerically. We can use this approach to develop a numerical method for solving the backward Kolmogorov equation. For this, we need to expand each coefficient u_n in an appropriate basis with respect to q . Natural choices are the Hermite basis when the potential in the Langevin equation is a polynomial, for example the quartic potential $V(q) = \frac{1}{4}q^4 - \frac{1}{2}q^2$, and the Fourier basis when we are studying the Langevin equation in a periodic potential. In Sect. 6.6, we will use this approach to solve the Poisson equation (6.133).

6.5 The Overdamped and Underdamped Limits for the Langevin Equation

In this section, we study the regimes of large and small friction for the Langevin equation (6.1). In particular, we obtain approximate equations that are valid in the *overdamped* $\gamma \rightarrow +\infty$ and *underdamped* $\gamma \rightarrow 0$ limits. It will turn out that passing to the underdamped limit in the Langevin equation is more delicate, and we will present formal calculations in only one dimension.

In order to obtain a well-defined limit, it is also necessary to look at the appropriate time scale. We consider a one-parameter family of rescalings of solutions to the Langevin equation (6.1). This will also be useful when we study the problem of Brownian motion in a periodic potential:

$$q^\gamma(t) = \lambda_\gamma q(t/\mu_\gamma). \quad (6.87)$$

This rescaled process satisfies the equation

$$\ddot{q}^\gamma = -\frac{\lambda_\gamma}{\mu_\gamma^2} \nabla_q V(q^\gamma/\lambda_\gamma) - \frac{\gamma}{\mu_\gamma} \dot{q}^\gamma + \sqrt{2\gamma\lambda_\gamma^2\mu_\gamma^{-3}\beta^{-1}} \dot{W}. \quad (6.88)$$

In deriving the rescaled Langevin equation, we have used the scaling properties of white noise, in particular the fact that $\dot{W}(ct) = \frac{1}{\sqrt{c}}\dot{W}(t)$ in law. Different choices for these two parameters lead to the overdamped and underdamped limits. Consider first

$$\lambda_\gamma = 1, \quad \mu_\gamma = \gamma^{-1}. \quad (6.89)$$

Equation (6.88) becomes

$$\gamma^{-2} \ddot{q}^\gamma = -\nabla_q V(q^\gamma) - \dot{q}^\gamma + \sqrt{2\beta^{-1}} \dot{W}. \quad (6.90)$$

This is the correct rescaling of the Langevin equation in which to study the overdamped limit $\gamma \gg 1$. We will see later in this section that in the limit as $\gamma \rightarrow +\infty$, the solution of (6.90) converges to the solution of the Smoluchowski/overdamped SDE

$$\dot{q} = -\nabla_q V(q) + \sqrt{2\beta^{-1}} \dot{W}. \quad (6.91)$$

Consider now the rescaling

$$\lambda_\gamma = 1, \quad \mu_\gamma = \gamma. \quad (6.92)$$

Now Eq. (6.88) becomes

$$\ddot{q}^\gamma = -\gamma^{-2} \nabla V(q^\gamma) - \dot{q}^\gamma + \sqrt{2\gamma^{-2}\beta^{-1}} \dot{W}. \quad (6.93)$$

This is the correct rescaling in time for studying the underdamped limit, $\gamma \ll 1$. We will study the limit $\gamma \rightarrow 0$ in one dimension and derive a stochastic differential equation for the energy (Hamiltonian).

6.5.1 The Overdamped Limit

We consider the rescaled Langevin equation (6.90) and set $\varepsilon^{-1} = \gamma$:

$$\varepsilon^2 \ddot{q}^\varepsilon = -\nabla V(q^\varepsilon) - \dot{q}^\varepsilon + \sqrt{2\beta^{-1}} \dot{W}(t). \quad (6.94)$$

We write (6.94) as a system of singularly perturbed stochastic differential equations

$$dq_t^\varepsilon = \frac{1}{\varepsilon} p_t^\varepsilon dt, \quad (6.95a)$$

$$dp_t^\varepsilon = -\frac{1}{\varepsilon} \nabla V(q_t^\varepsilon) dt - \frac{1}{\varepsilon^2} p_t^\varepsilon dt + \sqrt{\frac{2}{\beta\varepsilon^2}} dW_t. \quad (6.95b)$$

The position and momentum have different characteristic time scales. In particular, the momentum converges to equilibrium—“thermalizes”—much faster than the

position. It is natural to expect that when $\varepsilon \ll 1$, in the limit of large-scale separation between the characteristic time scales of position and momentum, we should be able to describe the dynamics of (6.95) using only the position variable while the momentum is thermalized.

The generator of (6.95) is

$$\begin{aligned}\mathcal{L}^\varepsilon &= \frac{1}{\varepsilon^2}(-p \cdot \nabla_p + \beta^{-1} \Delta_p) + \frac{1}{\varepsilon}(p \cdot \nabla_q - \nabla_q V \cdot \nabla_p) \\ &=: \frac{1}{\varepsilon^2} \mathcal{L}_0 + \frac{1}{\varepsilon} \mathcal{L}_1.\end{aligned}$$

To get some intuition on how the Smoluchowski equation (6.91) arises in the limit as $\varepsilon \rightarrow 0$, we apply Itô's formula to p_t^ε :

$$\begin{aligned}dp_t^\varepsilon &= \mathcal{L}^\varepsilon p_t^\varepsilon dt + \frac{1}{\varepsilon} \sqrt{2\beta^{-1}} \nabla_p p_t^\varepsilon dW \\ &= -\frac{1}{\varepsilon^2} p_t^\varepsilon dt - \frac{1}{\varepsilon} \nabla_q V(q_t^\varepsilon) dt + \frac{1}{\varepsilon} \sqrt{2\beta^{-1}} dW.\end{aligned}$$

Consequently,

$$\frac{1}{\varepsilon} \int_0^t p_s^\varepsilon ds = - \int_0^t \nabla_q V(q_s^\varepsilon) ds + \sqrt{2\beta^{-1}} W(t) + \mathcal{O}(\varepsilon).$$

From Eq. (6.95a), we have that

$$q_t^\varepsilon = q_0^\varepsilon + \frac{1}{\varepsilon} \int_0^t p_s^\varepsilon ds.$$

Combining the above two equations, we deduce

$$q_t^\varepsilon = q_0^\varepsilon - \int_0^t \nabla_q V(q_s^\varepsilon) ds + \sqrt{2\beta^{-1}} W(t) + \mathcal{O}(\varepsilon),$$

from which (6.91) follows.

In the above formal calculations, we assumed that the momentum is bounded, for example that we have an estimate of the form

$$\mathbb{E}|p_t^\varepsilon|^2 \leq C. \quad (6.96)$$

This estimate can be proved under appropriate assumptions on the potential $V(q)$. It is certainly true in the absence of a potential and when the initial conditions for p_t^ε are stationary, since in that case, (6.91) becomes an Ornstein–Uhlenbeck process. Using an estimate of the form (6.96), we can prove rigorously the convergence of the rescaled Langevin dynamics to the Smoluchowski SDE in the overdamped limit and also obtain an error estimate of the form

$$\left(\mathbb{E}|q_t^\varepsilon - q_t|^2 \right)^{1/2} \leq C\varepsilon,$$

where q_t^ε and q_t denote the solutions to (6.95a) and (6.91) with the same initial conditions.

Now we study the overdamped limit and derive the limiting Smoluchowski equation by studying the Fokker–Planck equation corresponding to (6.95) using singular perturbation theory. For simplicity, we will consider the problem in one dimension. The multidimensional problem can be treated in a similar way.

The Fokker–Planck equation associated with equations (6.95a) and (6.95b) is

$$\begin{aligned}\frac{\partial \rho}{\partial t} &= \mathcal{L}^* \rho \\ &= \frac{1}{\varepsilon} (-p \partial_q \rho + \partial_q V(q) \partial_p \rho) + \frac{1}{\varepsilon^2} (\partial_p (p \rho) + \beta^{-1} \partial_p^2 \rho) \\ &=: \left(\frac{1}{\varepsilon^2} \mathcal{L}_0^* + \frac{1}{\varepsilon} \mathcal{L}_1^* \right) \rho.\end{aligned}\quad (6.97)$$

We assume that V is a confining potential and that the dynamics (6.95) is ergodic with respect to the Gibbs distribution (6.9).⁷ We use the transformation (6.13) to obtain Eq. (6.14). In particular, we define the function $f(p, q, t)$ through

$$\rho(p, q, t) = f(p, q, t) \rho_\beta(p, q). \quad (6.98)$$

It satisfies the equation

$$\begin{aligned}\frac{\partial f}{\partial t} &= \left[\frac{1}{\varepsilon^2} (-p \partial_q + \beta^{-1} \partial_p^2) - \frac{1}{\varepsilon} (p \partial_q - \partial_q V(q) \partial_p) \right] f \\ &=: \left(\frac{1}{\varepsilon^2} \mathcal{L}_0 - \frac{1}{\varepsilon} \mathcal{L}_1 \right) f.\end{aligned}\quad (6.99)$$

We will assume that the initial conditions for (6.99) depend only on q :

$$f(p, q, 0) = f_{ic}(q). \quad (6.100)$$

This will make the analysis easier, since the initial conditions for the limiting Fokker–Planck equation will be the same as for (6.99).

We look for a solution to (6.99) in the form of a power series in ε :

$$f(p, q, t) \approx \sum_{n=0}^N \varepsilon^n f_n(p, q, t). \quad (6.101)$$

We substitute this expansion into (6.99) to obtain the following system of equations:

$$\mathcal{L}_0 f_0 = 0, \quad (6.102a)$$

$$-\mathcal{L}_0 f_1 = -\mathcal{L}_1 f_0, \quad (6.102b)$$

$$-\mathcal{L}_0 f_2 = -\mathcal{L}_1 f_1 - \frac{\partial f_0}{\partial t} \quad (6.102c)$$

$$-\mathcal{L}_0 f_n = -\mathcal{L}_1 f_{n-1} - \frac{\partial f_{n-2}}{\partial t}, \quad n = 3, 4 \dots N. \quad (6.102d)$$

⁷ Note that the invariant distribution of (6.95) is independent of ε .

The null space of \mathcal{L}_0 consists of constants in p . Consequently, from Eq. (6.102a), we conclude that

$$f_0 = f(q, t).$$

Now we can calculate the right-hand side of Eq. (6.102b):

$$\mathcal{L}_1 f_0 = p \partial_q f.$$

Equation (6.102b) becomes

$$\mathcal{L}_0 f_1 = p \partial_q f. \quad (6.103)$$

In order for this equation to be well posed, we need the right-hand side to be orthogonal to $\mathcal{N}(\mathcal{L}_0^*)$, i.e., that it be centered with respect to the Gaussian invariant measure of the Ornstein–Uhlenbeck process with generator \mathcal{L}_0 :

$$\int_{\mathbb{R}} p \partial_q f Z^{-1} e^{-\beta \frac{p^2}{2}} = 0.$$

This condition is satisfied, and consequently, a solution to (6.103) exists. We obtain this solution using separation of variables:

$$f_1 = -p \partial_q f + \psi_1(q, t).$$

Now we can calculate the right-hand side of Eq. (6.102c). We need to calculate $\mathcal{L}_1 f_1$:

$$\begin{aligned} -\mathcal{L}_1 f_1 &= \left(p \partial_q - \partial_q V \partial_p \right) \left(p \partial_q f - \psi_1(q, t) \right) \\ &= p^2 \partial_q^2 f - p \partial_q \psi_1 - \partial_q V \partial_q f. \end{aligned}$$

The solvability condition for (6.102c) is

$$\int_{\mathbb{R}} \left(-\mathcal{L}_1 f_1 - \frac{\partial f_0}{\partial t} \right) Z^{-1} e^{-\beta \frac{p^2}{2}} dp = 0,$$

from which we obtain the Fokker–Planck equation, written in the form 6.14 corresponding to the Smoluchowski SDE:

$$\frac{\partial f}{\partial t} = \partial_q V \partial_q f + \beta^{-1} \partial_q^2 f, \quad (6.104)$$

together with the initial condition (6.100).

We can also obtain higher-order corrections to the limiting Smoluchowski dynamics. We solve the equation for f_2 . We use (6.104) to write (6.102c) in the form

$$\mathcal{L}_0 f_2 = \left(\beta^{-1} - p^2 \right) \partial_q^2 f + p \partial_q \psi_1.$$

The solution of this equation is

$$f_2(p, q, t) = \frac{1}{2} \partial_q^2 f(p, q, t) p^2 - \partial_q \psi_1(q, t) p + \psi_2(q, t).$$

Now we calculate the right-hand side of the equation for f_3 , Eq. (6.102d) with $n = 3$. First we calculate

$$\mathcal{L}_1 f_2 = \frac{1}{2} p^3 \partial_q^3 f - p^2 \partial_q^2 \psi_1 + p \partial_q \psi_2 - \partial_q V \partial_q^2 f p - \partial_q V \partial_q \psi_1.$$

The solvability condition

$$\int_{\mathbb{R}} \left(\frac{\partial \psi_1}{\partial t} + \mathcal{L}_1 f_2 \right) Z^{-1} e^{-\beta \frac{p^2}{2}} dp = 0$$

leads to the equation

$$\frac{\partial \psi_1}{\partial t} = \partial_q V \partial_q \psi_1 + \beta^{-1} \partial_q^2 \psi_1,$$

together with the initial condition $\psi_1(q, 0) = 0$. From the calculations presented in the proof of Theorem 4.4, and using the Poincaré inequality for the density $\frac{1}{Z_V} e^{-\beta V(q)}$, $Z_V = \int e^{-\beta V(q)} dq$, we deduce that

$$\frac{1}{2} \frac{d}{dt} \|\psi_1\|^2 \leq -C \|\psi_1\|^2.$$

We use Gronwall's inequality now to conclude that

$$\psi_1 \equiv 0.$$

Putting everything together, we obtain the first two terms in the ε -expansion of the Fokker–Planck equation (6.99):

$$\rho(p, q, t) = Z^{-1} e^{-\beta H(p, q)} \left(f + \varepsilon (-p \partial_q f) + \mathcal{O}(\varepsilon^2) \right),$$

where f is the solution of (6.104). Note that we can rewrite the leading-order term to the expansion in the form

$$\rho(p, q, t) = (2\pi\beta^{-1})^{-\frac{1}{2}} e^{-\beta p^2/2} \rho_V(q, t) + \mathcal{O}(\varepsilon),$$

where $\rho_V = Z^{-1} e^{-\beta V(q)} f$ is the solution of the Smoluchowski Fokker–Planck equation

$$\frac{\partial \rho_V}{\partial t} = \partial_q (\partial_q V \rho_V) + \beta^{-1} \partial_q^2 \rho_V.$$

It is possible to expand the n th term in the expansion (6.101) in terms of Hermite functions, the eigenfunctions of the generator of the Ornstein–Uhlenbeck process,

$$f_n(p, q, t) = \sum_{k=0}^n f_{nk}(q, t) h_k(p), \quad (6.105)$$

where $h_k(p)$ is the k th eigenfunction of \mathcal{L}_0

$$-\mathcal{L}_0 h_k = \lambda_k h_k$$

(see Theorem 4.2). We can obtain the following system of equations, using the notation $Y^* = \sqrt{\beta^{-1}}\partial_q - \sqrt{\beta}\partial_q V$:

$$\begin{aligned} Y^* f_{n1} &= 0, \\ \sqrt{k+1} Y^* f_{n,k+1} + \sqrt{k\beta^{-1}} \partial_q f_{n,k-1} &= -k f_{n+1,k}, \quad k = 1, 2, \dots, n-1, \\ \sqrt{n\beta^{-1}} \partial_q f_{n,n-1} &= -n f_{n+1,n}, \\ \sqrt{(n+1)\beta^{-1}} \partial_q f_{n,n} &= -(n+1) f_{n+1,n+1}; \end{aligned}$$

Using this method, we can obtain the first three terms in the expansion:

$$\begin{aligned} \rho(x, y, t) &= \rho_0(p, q) \left(f + \varepsilon (-\sqrt{\beta^{-1}} \partial_q f h_1) + \varepsilon^2 \left(\frac{\beta^{-1}}{\sqrt{2}} \partial_q^2 f h_2 + f_{20} \right) \right. \\ &\quad \left. + \varepsilon^3 \left(-\sqrt{\frac{\beta^{-3}}{3!}} \partial_q^3 f h_3 + (-Y^* \partial_q^2 f - \sqrt{\beta^{-1}} \partial_q f_{20}) h_1 \right) \right) \\ &\quad + \mathcal{O}(\varepsilon^4). \end{aligned}$$

The interested reader is invited to provide the details.

6.5.2 The Underdamped Limit

Consider now the rescaling (6.92), which leads to Eq. (6.93). We rewrite this equation as a first-order system:

$$dq_t^\gamma = \gamma^{-1} p_t^\gamma dt, \quad dp_t^\gamma = -\gamma^{-1} \nabla V(q_t^\gamma) dt - p_t^\gamma dt + \sqrt{2\beta^{-1}} dW_t. \quad (6.106)$$

This is the equation for an $\mathcal{O}(1/\gamma)$ Hamiltonian system perturbed by $\mathcal{O}(1)$ noise. The situation is more complicated than in the case of overdamped dynamics, since it is not a priori clear what the fast and slow variables are in (6.106). We know that the energy (Hamiltonian) is conserved for Hamiltonian systems. We expect, then, that when $\gamma \ll 1$, the energy is slowly varying so that we can identify it as one of the slow variables. For multidimensional Hamiltonian systems, it is possible to have additional conserved quantities (integrals of motion), such as the angular momentum. To study the underdamped limit of (6.106), we need to make additional assumptions on the properties of the deterministic Hamiltonian dynamics. Understanding the small- γ asymptotics of (6.106) for multidimensional systems is beyond the scope of this book. We will consider (6.106) in one dimension, where the Hamiltonian is the only integral of motion.

The generator of (6.106) is of the form $\mathcal{L}^\gamma = -\frac{1}{\gamma}B + \mathcal{S}$, where B is the Liouville operator and \mathcal{S} is the generator of the Ornstein–Uhlenbeck process in p . We have already remarked that the Hamiltonian is in the null space of the Liouville operator.

We apply Itô's formula to the Hamiltonian $H^\gamma = H(q^\gamma, p^\gamma)$ of the system to obtain the equation

$$\dot{H}^\gamma = (\beta^{-1} - (p^\gamma)^2) + \sqrt{2\beta^{-1}(p^\gamma)^2} \dot{W}, \quad (6.107)$$

with $(p^\gamma)^2 = (p^\gamma)(H, q) = 2(H - V(q^\gamma))$. We can now rewrite (6.106) as a fast/slow system of stochastic differential equations for the Hamiltonian (the slow variable) and the momentum (the fast variable):

$$dH_t = (\beta^{-1} - p_t^2) dt + \sqrt{2\beta^{-1} p_t^2} dW_t, \quad (6.108a)$$

$$dp_t = -\frac{1}{\gamma} V'(q_t) dt - \frac{1}{\gamma} p_t dt + \sqrt{2\beta^{-1}} dW_t, \quad (6.108b)$$

where we have dropped the subscript γ for notational simplicity. Assuming that we can average over the Hamiltonian dynamics, we obtain the limiting stochastic differential equation for the Hamiltonian:

$$\dot{H} = (\beta^{-1} - \langle p^2 \rangle) + \sqrt{2\beta^{-1} \langle p^2 \rangle} \dot{W}, \quad (6.109)$$

where $\langle \cdot \rangle$ denotes an appropriate average with respect to the “fast” Hamiltonian dynamics. See Eq. (6.120) for the correct formulation of the limiting equation. This formal discussion does not provide us with any information on the state space of the limiting dynamics, or equivalently, on the boundary conditions for the generator of (6.109). Comments on this issue can be found in Sect. 6.7.

We will study the small- γ asymptotics by analyzing the backward Kolmogorov equation corresponding to (6.106) using singular perturbation theory. As we have already mentioned several times, this is equivalent to studying the Fokker–Planck equation; see (6.16) and (6.17). The generator of the process $\{q^\gamma, p^\gamma\}$ is

$$\begin{aligned} \mathcal{L}^\gamma &= \gamma^{-1} (p \partial_q - \partial_q V \partial_p) - p \partial_p + \beta^{-1} \partial_p^2 \\ &=: \gamma^{-1} \mathcal{L}_0 + \mathcal{L}_1. \end{aligned}$$

We have used the notation \mathcal{L}_0 instead of $-B$, and \mathcal{L}_1 instead of \mathcal{S} , for consistency with the notation used in Sect. 6.5 and in other parts of the book where singular perturbation calculations are presented. Let $u^\gamma(q, p, t) = \mathbb{E}(f(p^\gamma(p, q; t), q^\gamma(p, q; t)))$.⁸ It satisfies the backward Kolmogorov equation associated with the process $\{q^\gamma, p^\gamma\}$:

$$\frac{\partial u^\gamma}{\partial t} = \left(\frac{1}{\gamma} \mathcal{L}_0 + \mathcal{L}_1 \right) u^\gamma. \quad (6.110)$$

We look for a solution in the form of a power series expansion in ε :

$$u^\gamma = u_0 + \gamma u_1 + \gamma^2 u_2 + \dots$$

⁸ In order to avoid initial layers, we need to assume that the initial condition is a function of the Hamiltonian. We will not study the technical issue of initial layers.

We substitute this ansatz into (6.110) and equate equal powers in ε to obtain the following sequence of equations:

$$\mathcal{L}_0 u_0 = 0, \quad (6.111a)$$

$$\mathcal{L}_0 u_1 = -\mathcal{L}_1 u_1 + \frac{\partial u_0}{\partial t}, \quad (6.111b)$$

$$\mathcal{L}_0 u_2 = -\mathcal{L}_1 u_1 + \frac{\partial u_1}{\partial t}. \quad (6.111c)$$

We assume that there are no integrals of motion other than the Hamiltonian. This means that the null space of \mathcal{L}_0 consists of functions of the Hamiltonian:

$$\mathcal{N}(\mathcal{L}_0) = \{ \text{functions of } H \}. \quad (6.112)$$

Let us now analyze Equations (6.111). From the first equation and using (6.112), we conclude that u_0 depends on q, p through the Hamiltonian function H :

$$u_0 = u(H(p, q), t). \quad (6.113)$$

Now we proceed with (6.111b). For this, we need to find the solvability condition for equations of the form

$$\mathcal{L}_0 u = f. \quad (6.114)$$

We multiply this equation by an arbitrary smooth function of the Hamiltonian, integrate over \mathbb{R}^2 , and use the fact that \mathcal{L}_0 is (formally) antisymmetric in $\mathcal{L}^2(\mathbb{R}^d)$ to deduce

$$\begin{aligned} \int_{\mathbb{R}^2} \mathcal{L}_0 u F(H(p, q)) dp dq &= \int_{\mathbb{R}^2} u \mathcal{L}_0^* F(H(p, q)) dp dq \\ &= \int_{\mathbb{R}^2} u (-\mathcal{L}_0 F(H(p, q))) dp dq \\ &= 0, \end{aligned}$$

for all smooth functions F that decay sufficiently fast at infinity, where we have assumed that both u_1 and F decay to 0 as $|p| \rightarrow \infty$ to justify the integration by parts. This implies that the *solvability condition* for Eq. (2.3) is

$$\int_{\mathbb{R}^2} f(p, q) F(H(p, q)) dp dq = 0 \quad (6.115)$$

for all smooth, rapidly decaying F . We use the solvability condition in (6.111b) to obtain

$$\int_{\mathbb{R}^2} \left(\mathcal{L}_1 u_1 - \frac{\partial u_0}{\partial t} \right) F(H(p, q)) dp dq = 0. \quad (6.116)$$

To proceed, we need to understand how the operator \mathcal{L}_1 acts on functions of the Hamiltonian; see Eq. (6.107). Let $\phi = \phi(H(p, q))$ be a C^2 function of the Hamiltonian. We calculate

$$\frac{\partial \phi}{\partial p} = \frac{\partial H}{\partial p} \frac{\partial \phi}{\partial H} = p \frac{\partial \phi}{\partial H}$$

and

$$\frac{\partial^2 \phi}{\partial p^2} = \frac{\partial}{\partial p} \left(\frac{\partial \phi}{\partial H} \right) = \frac{\partial \phi}{\partial H} + p^2 \frac{\partial^2 \phi}{\partial H^2}.$$

Consequently,

$$\mathcal{L}_1 \phi = (\beta^{-1} - p^2(H, q)) \partial_H \phi + \beta^{-1} p^2(H, q) \partial_H^2 \phi, \quad (6.117)$$

where

$$p^2 = p^2(H, q) = 2(H - V(q)).$$

We want to change variables in the integral (6.116) and go from (p, q) to (p, H) . The Jacobian of the transformation is

$$\frac{\partial(p, q)}{\partial(H, q)} = \begin{vmatrix} \frac{\partial p}{\partial H} & \frac{\partial p}{\partial q} \\ \frac{\partial q}{\partial H} & \frac{\partial q}{\partial q} \end{vmatrix} = \frac{\partial p}{\partial H} = \frac{1}{p(H, q)}.$$

We use this, together with (6.117), to rewrite Eq. (6.116) as

$$\int \int \left(\frac{\partial u}{\partial t} + [(\beta^{-1} - p^2) \partial_H + \beta^{-1} p^2 \partial_H^2] u \right) F(H) p^{-1}(H, q) dH dq = 0.$$

We can first perform an integration with respect to q . Using the notation $\langle \cdot \rangle := \oint \cdot d\gamma$ for integration along a curve of constant energy on the phase plane, we obtain

$$\int \left[\frac{\partial u}{\partial t} \langle p^{-1} \rangle + ((\beta^{-1} \langle p^{-1} \rangle - \langle p \rangle) \partial_H + \beta^{-1} \langle p \rangle \partial_H^2) u \right] F(H) dH = 0. \quad (6.118)$$

This equation should be valid for every smooth function F of the Hamiltonian. Consequently,

$$\langle p^{-1} \rangle \frac{\partial u}{\partial t} = (\beta^{-1} \langle p^{-1} \rangle - \langle p \rangle) \partial_H u + \langle p \rangle \beta^{-1} \partial_H^2 u,$$

or

$$\frac{\partial u}{\partial t} = (\beta^{-1} - \langle p^{-1} \rangle^{-1} \langle p \rangle) \partial_H u + \langle p^{-1} \rangle^{-1} \langle p \rangle \beta^{-1} \partial_H^2 u. \quad (6.119)$$

From this equation, we can read off the limiting Itô stochastic differential equation for the Hamiltonian:

$$\dot{H} = \beta^{-1} - \langle p^{-1} \rangle^{-1} \langle p \rangle + \sqrt{2\beta^{-1} \langle p^{-1} \rangle^{-1} \langle p \rangle} \dot{W}. \quad (6.120)$$

Note that the noise that appears in the limiting equation (6.120) is multiplicative, in contrast to the additive noise that appears in the Langevin equation. The underdamped regime $\gamma \ll 1$ is also referred to as the *energy-diffusion-limited regime*.

We will now write the Fokker–Planck equation corresponding to (6.120) in a form that is convenient for further analysis. First we define the *action* $I(E)$ and *frequency* $\omega(E)$ at energy $E = p^2/2 + V(q)$:

$$I(E) = \oint p(q, E) dq \quad \text{and} \quad \omega(E) = 2\pi \left(\frac{dI}{dE} \right)^{-1}, \quad (6.121)$$

where $p^2(q, E) = 2(V(q) - E)$. We can write the limiting Fokker–Planck equation in terms of the action and the frequency.

Proposition 6.10. *The limiting Fokker–Planck equation for the energy distribution function $\rho(E, t)$ is*

$$\frac{\partial \rho}{\partial t} = \frac{\partial}{\partial E} \left(I(E) \left(1 + \beta^{-1} \frac{\partial}{\partial E} \right) \left(\frac{\omega(E) \rho}{2\pi} \right) \right). \quad (6.122)$$

Proof. We observe that

$$\frac{dI}{dE} = \oint \frac{\partial p}{\partial E} dq = \oint p^{-1} dq,$$

and consequently,

$$\langle p^{-1} \rangle^{-1} = \frac{\omega(E)}{2\pi}.$$

Hence, the limiting Fokker–Planck equation can be written as

$$\begin{aligned} \frac{\partial \rho}{\partial t} &= -\frac{\partial}{\partial E} \left(\left(\beta^{-1} \frac{I(E) \omega(E)}{2\pi} \right) \rho \right) + \beta^{-1} \frac{\partial^2}{\partial E^2} \left(\frac{I \omega}{2\pi} \right) \\ &= -\beta^{-1} \frac{\partial \rho}{\partial E} + \frac{\partial}{\partial E} \left(\frac{I \omega}{2\pi} \rho \right) + \beta^{-1} \frac{\partial}{\partial E} \left(\frac{dI}{dE} \frac{\omega \rho}{2\pi} \right) + \beta^{-1} \frac{\partial}{\partial E} \left(I \frac{\partial}{\partial E} \left(\frac{\omega \rho}{2\pi} \right) \right) \\ &= \frac{\partial}{\partial E} \left(\frac{I \omega}{2\pi} \rho \right) + \beta^{-1} \frac{\partial}{\partial E} \left(I \frac{\partial}{\partial E} \left(\frac{\omega \rho}{2\pi} \right) \right) \\ &= \frac{\partial}{\partial E} \left(I(E) \left(1 + \beta^{-1} \frac{\partial}{\partial E} \right) \left(\frac{\omega(E) \rho}{2\pi} \right) \right), \end{aligned}$$

which is precisely Eq. (6.122). \square

We emphasize again that the above formal procedure does not provide us with the boundary conditions for the limiting Fokker–Planck equation. Note also that if we rescale back to the original time scale, we obtain the equation

$$\frac{\partial \rho}{\partial t} = \gamma \frac{\partial}{\partial E} \left(I(E) \left(1 + \beta^{-1} \frac{\partial}{\partial E} \right) \left(\frac{\omega(E) \rho}{2\pi} \right) \right). \quad (6.123)$$

This equation can be used to calculate the diffusion coefficient of a Brownian particle in a periodic potential as well as the hopping rate of a Brownian particle moving in a double-well potential in the underdamped or energy-diffusion-limited regime. See Sect. 6.6 and Eq. (7.79).

6.6 The Langevin Equation in a Periodic Potential

In this section, we study the motion of a Brownian particle in a periodic potential. We model the dynamics of the particle using the Langevin equation (6.1) with $V(q)$ a smooth periodic potential. A typical trajectory of a Brownian particle moving in a cosine potential, with $\gamma = \beta = 1$, is shown in Fig. 6.2. A Brownian particle moving in a periodic potential performs a hopping motion between the local minima of the cosine potential. In particular, it oscillates around the local minima of the potential. Eventually, the particle acquires a sufficient amount of energy to surmount the potential barrier (see Chap. 7) and to jump over one or even several local minima until it loses a sufficient amount of energy due to friction that it is trapped at a different local minimum of the potential. The process is then repeated. Effectively, this amounts to a random walk whose sites are located at the local minima of the potential. We can describe the long-time/large-scale behavior using an effective Brownian motion with a diffusion coefficient that depends on the periodic potential, the friction coefficient, and the temperature.

The mean-square displacement of Brownian motion scales linearly in time. We can use this property of Brownian motion to define the diffusion coefficient of the effective Brownian motion. Denoting by $\langle \ell^2 \rangle$ the mean-square length of the jumps of the Brownian particle⁹ and by τ the average time it takes for the particle to perform a jump, we can define and calculate the diffusion coefficient using the formula

$$D = \frac{\langle \ell^2 \rangle}{2\tau}. \quad (6.124)$$

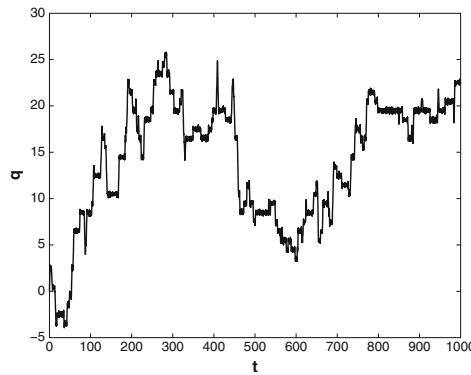


Fig. 6.2 A typical trajectory of a Brownian particle moving in a cosine potential

⁹ $\langle \ell^2 \rangle$ will not, in general, be equal to the period of the potential, with the exception of the high-friction regime.

Alternatively, and in arbitrary dimensions, we can define the diffusion tensor (covariance matrix) in terms of the long-time average of the second moment of q , divided by time:

$$D = \lim_{t \rightarrow \infty} \frac{\langle (q(t) - \langle q(t) \rangle) \otimes (q(t) - \langle q(t) \rangle) \rangle}{2t}. \quad (6.125)$$

In this section, we show that when appropriately rescaled, the solution to the Langevin equation in a periodic potential converges to a Brownian motion with a covariance matrix D that we can calculate by solving an appropriate Poisson equation. We will use techniques from singular perturbation theory and homogenization theory, similar to the techniques that we used in Sect. 6.5.

Consider the rescaling (6.87) and the rescaled Langevin equation (6.88). We set

$$\lambda = \varepsilon, \quad \mu = \varepsilon^2,$$

where we have removed the subscript γ , since the rescaling does not depend on the friction coefficient. This *diffusive rescaling* is dictated by the scaling properties of the limiting Brownian motion. The Langevin equation (6.88) becomes

$$\varepsilon^2 \ddot{q}_t^\varepsilon = -\frac{1}{\varepsilon} \nabla V \left(\frac{q_t^\varepsilon}{\varepsilon} \right) - \gamma \dot{q}_t^\varepsilon + \sqrt{2\gamma\beta^{-1}} \dot{W}. \quad (6.126)$$

The precise mathematical statement of the result that we will derive in this section is that the solution to the rescaled Langevin equation (6.126) converges weakly to a Brownian motion with covariance matrix D :

$$\lim_{\varepsilon \rightarrow 0} q_t^\varepsilon = \sqrt{2D} W_t.$$

For the calculation that follows, it is important to distinguish between the motion within a period of the potential and the motion in \mathbb{R}^d . Introducing the momentum $p = \varepsilon \dot{q}$ and the auxiliary variable $z = q/\varepsilon$, we write this equation as a first-order system:¹⁰

$$\begin{aligned} \dot{q} &= \frac{1}{\varepsilon} p, \\ \dot{p} &= -\frac{1}{\varepsilon^2} \nabla V(z) - \frac{1}{\varepsilon^2} \gamma p + \frac{1}{\varepsilon^2} \sqrt{2\gamma\beta^{-1}} \dot{W}, \\ \dot{z} &= \frac{1}{\varepsilon^2} p, \end{aligned} \quad (6.127)$$

with the understanding that $z \in \mathcal{Y} := [0, L]^d$ and $q, p \in \mathbb{R}^d$. Our goal now is to eliminate the fast variables p, z and to obtain an equation for the slow variable q .

¹⁰ This is a familiar trick from the theory of homogenization for partial differential equations with periodic coefficients. For example, to study the PDE

$$-\nabla \cdot \left(A \left(x, \frac{x}{\varepsilon} \right) \nabla u^\varepsilon \right) = f,$$

where the matrix-valued function A is periodic in its second argument, it is convenient to set $z = \frac{x}{\varepsilon}$ and to treat x and z as independent variables.

As in the previous section, we will apply singular perturbation techniques to the corresponding backward Kolmogorov equation, which reads

$$\begin{aligned}\frac{\partial u^\varepsilon}{\partial t} &= \frac{1}{\varepsilon} p \cdot \nabla_q u^\varepsilon + \frac{1}{\varepsilon^2} \left(-\nabla_z V(z) \cdot \nabla_p + p \cdot \nabla_z + \gamma(-p \cdot \nabla_p + \beta^{-1} \Delta_p) \right) u^\varepsilon. \\ &= \left(\frac{1}{\varepsilon^2} \mathcal{L}_0 + \frac{1}{\varepsilon} \mathcal{L}_1 \right) u^\varepsilon,\end{aligned}\quad (6.128)$$

where

$$\begin{aligned}\mathcal{L}_0 &= -\nabla_z V(z) \cdot \nabla_p + p \cdot \nabla_z + \gamma(-p \cdot \nabla_p + \beta^{-1} \Delta_p), \\ \mathcal{L}_1 &= p \cdot \nabla_q.\end{aligned}$$

The state space of the fast process $\{z(t), p(t)\}$ is $\mathcal{M} := \mathcal{Y} \times \mathbb{R}^d$. Since the state space is compact in z , the potential $V(z)$ satisfies the Poincaré inequality, and Theorem 6.3 applies: the fast process is an ergodic Markov process with invariant distribution

$$\rho_\beta(q, p) = Z^{-1} e^{-\beta H(z, p)}, \quad Z = \int_{\mathcal{M}} e^{-\beta H(z, p)} dz dp,$$

where $H(z, p) = \frac{1}{2} |p|^2 + V(z)$ and with exponentially fast convergence to equilibrium. This implies, in particular, that the null space of the generator \mathcal{L}_0 consists of constants in z, p ; see Sect. 2.4. Furthermore, the generator of $\{z(t), p(t)\}$ has a spectral gap, which is a sufficient condition for the well-posedness of the Poisson equation

$$-\mathcal{L}_0 f = g. \quad (6.129)$$

We consider this equation on \mathcal{M} . The boundary conditions are periodic in z , and we require that the solution be square-integrable with respect to ρ_β , $f \in L^2(\mathcal{M}; \rho_\beta)$. This equation has a unique, up to constants in z, p , solution if and only if the right-hand side is orthogonal to the null space of the L^2 -adjoint operator of \mathcal{L} , i.e., the Fokker–Planck operator:

$$\int_{\mathcal{M}} g(z, p) \rho_\beta(z, p) dz dp = 0. \quad (6.130)$$

We will refer to this as the *centering condition*.

We look for a solution to the backward Kolmogorov equation in the form of a power series expansion in ε :

$$u^\varepsilon = u_0 + \varepsilon u_1 + \varepsilon^2 u_2 + \dots, \quad (6.131)$$

with $u_i = u_i(p, q, z, t)$, $i = 1, 2, \dots$, L -periodic in z and square-integrable with respect to the equilibrium distribution ρ_β . We substitute (6.131) into (6.128) and equate equal powers in ε to obtain the following sequence of equations:

$$\mathcal{L}_0 u_0 = 0, \quad (6.132a)$$

$$\mathcal{L}_0 u_1 = -\mathcal{L}_1 u_0, \quad (6.132b)$$

$$\mathcal{L}_0 u_2 = -\mathcal{L}_1 u_1 + \frac{\partial u_0}{\partial t}. \quad (6.132c)$$

From the first equation in (6.132), we deduce that $u_0 = u_0(q, t)$, since the null space of \mathcal{L}_0 consists of functions that are constants in p and z . Now the second equation in (6.132) becomes

$$\mathcal{L}_0 u_1 = -p \cdot \nabla_z u_0.$$

Since p is of mean zero with respect to the invariant distribution of the process $\{p_t, z_t\}$, the solvability condition (6.130) is satisfied. Hence, the above equation is well posed. We solve it using separation of variables:

$$u_1 = \Phi(p, z) \cdot \nabla_q u_0$$

with

$$-\mathcal{L}_0 \Phi = p. \quad (6.133)$$

This Poisson equation is precisely of the form (6.129). Now we proceed with the third equation in (6.132). We apply the solvability condition to obtain

$$\begin{aligned} \frac{\partial u_0}{\partial t} &= \int_{\mathcal{M}} \mathcal{L}_1 u_1 \rho_\beta(p, z) dp dz \\ &= \sum_{i,j=1}^d \left(\int_{\mathcal{M}} p_i \Phi_j \rho_\beta(p, z) dp dz \right) \frac{\partial^2 u_0}{\partial q_i \partial q_j}. \end{aligned}$$

This is the backward Kolmogorov equation corresponding to a Brownian motion with diffusion tensor (covariance matrix)

$$D_{ij} = \int_{\mathcal{M}} p_i \Phi_j \rho_\beta(p, q) dp dq, \quad i, j = 1, \dots, d. \quad (6.134)$$

The calculation of the diffusion tensor requires the solution of the boundary value problem (6.133) and the calculation of the integral in (6.134).

Lemma 6.11. *The diffusion tensor (6.134) is positive definite.*

Proof. To show that the diffusion tensor is nonnegative, we use the fact that it can be expressed in terms of the Dirichlet form associated with the generator \mathcal{L}_0 ; see Eq. (4.75). Indeed, let ξ be a unit vector in \mathbb{R}^d . We calculate, using the notation $\Phi_\xi = \Phi \cdot \xi$,

$$\begin{aligned} \langle \xi, D\xi \rangle &= \int_{\mathcal{M}} (p \cdot \xi) (\Phi_\xi) \rho_\beta dp dq = \int_{\mathcal{M}} (-\mathcal{L}_0 \Phi_\xi) \Phi_\xi \rho_\beta dp dq \\ &= \gamma \beta^{-1} \int_{\mathcal{M}} |\nabla_p \Phi_\xi|^2 \rho_\beta dp dq \geq 0. \end{aligned} \quad (6.135)$$

To show that it is positive definite, we argue by contradiction: suppose that there exists a $\xi \in \mathbb{R}^d$ such that $\Phi \langle \xi, D\xi \rangle = 0$. From the previous calculation, this means that

$$\int_{\mathcal{M}} |\nabla_p \Phi_\xi|^2 \rho_\beta dp dq = 0.$$

This implies that Φ_ξ is independent of p , $\Phi_\xi = \Phi_\xi(q)$. The Poisson equation (6.133) becomes

$$-p \cdot \nabla_q \Phi^\xi = p \cdot \xi,$$

subject to periodic boundary conditions. The only solution to this equation is linear in q . Hence, no periodic solutions in q exist, and we arrive at a contradiction. \square

6.6.1 The Overdamped and Underdamped Limits of the Diffusion Coefficient

Now we study the formula for the diffusion coefficient in one dimension for potentials with period 2π . In particular, we derive approximate analytical formulas for the diffusion coefficient that are valid in the overdamped $\gamma \gg 1$ and underdamped $\gamma \ll 1$ limits. The derivation of these formulas is based on the asymptotic analysis of the Poisson equation (6.133). The calculations are similar to those that we presented in Sect. 6.5, and we will omit many of the details.

6.6.1.1 The Overdamped Limit

In this section, we study the large- γ asymptotics of the diffusion coefficient. The differential operator \mathcal{L}_0 becomes

$$\mathcal{L}_0 = \gamma \mathcal{S} - B,$$

where \mathcal{S} denotes the generator of the Ornstein–Uhlenbeck process, and B the Liouville operator. We look for a solution of (6.133) in the form of a power series expansion in γ^{-1} :

$$\Phi = \phi_0 + \frac{1}{\gamma} \phi_1 + \frac{1}{\gamma^2} \phi_2 + \dots \quad (6.136)$$

We substitute this into (6.133) and obtain a sequence of equations similar to (6.102):

$$-\mathcal{S}\phi_0 = 0, \quad -\mathcal{S}\phi_1 = p - B\phi_0, \quad -\mathcal{S}\phi_2 = -B\phi_1.$$

The analysis of these equations is similar to that of equations (6.102). The first two terms in the large- γ expansion of the solution of Eq. (6.133) are

$$\Phi(p, q) = \phi(q) + \frac{1}{\gamma} p (1 + \partial_q \phi) + \mathcal{O}\left(\frac{1}{\gamma^2}\right), \quad (6.137)$$

where $\phi(q)$ is the solution of the equation

$$\beta^{-1} \partial_q^2 \phi - \partial_q V(1 + \partial_q \phi) = 0, \quad (6.138)$$

posed on $[0, 2\pi]$ with periodic boundary conditions. We can solve this equation to obtain

$$\partial_q \phi = -1 + \frac{2\pi}{\widehat{Z}} e^{\beta V(q)}, \quad (6.139)$$

where $\widehat{Z} = \int_{-\pi}^{\pi} e^{\beta V(q)} dq$. Substituting (6.137) into the formula for the diffusion coefficient (6.134) in one dimension and using (6.139), we obtain

$$D = \int_{-\infty}^{\infty} \int_{-\pi}^{\pi} p \Phi \rho_{\beta}(p, q) dp dq = \frac{4\pi^2}{\gamma \beta Z \widehat{Z}} + \mathcal{O}\left(\frac{1}{\gamma^3}\right),$$

where $Z = \int_{-\pi}^{\pi} e^{-\beta V(q)} dq$.

Continuing in the same fashion, we can also calculate the next two terms in the expansion (6.136); see Exercise 9. From this, we can compute the next-order correction to the diffusion coefficient. The final result is

$$D = \frac{4\pi^2}{\beta \gamma Z \widehat{Z}} - \frac{4\pi^2 \beta Z_1}{\gamma^3 Z \widehat{Z}^2} + \mathcal{O}\left(\frac{1}{\gamma^5}\right), \quad (6.140)$$

where $Z_1 = \int_{-\pi}^{\pi} |V'(q)|^2 e^{\beta V(q)} dq$.

For the cosine potential $V(q) = \cos(q)$, formula (6.140) gives

$$D = \frac{1}{\gamma \beta} J_0^{-2}(\beta) - \frac{\beta}{\gamma^3} \left(\frac{J_2(\beta)}{J_0^3(\beta)} - J_0^{-2}(\beta) \right) + \mathcal{O}\left(\frac{1}{\gamma^5}\right), \quad (6.141)$$

where $J_n(\beta)$ is the modified Bessel function of the first kind.

In the multidimensional case, a similar analysis leads to the large- γ asymptotics:

$$\langle \xi, D \xi \rangle = \frac{1}{\gamma} \langle \xi, D_0 \xi \rangle + \mathcal{O}\left(\frac{1}{\gamma^3}\right),$$

where ξ is an arbitrary unit vector in \mathbb{R}^d , and D_0 is the diffusion coefficient for the Smoluchowski (overdamped Langevin) dynamics:

$$D_0 = Z^{-1} \int_{[0,L]^d} (-\mathcal{L}_V \chi) \otimes \chi e^{-\beta V(q)} dq, \quad (6.142)$$

where

$$\mathcal{L}_V = -\nabla_q V \cdot \nabla_q + \beta^{-1} \Delta_q$$

and $\chi(q)$ is the solution of the Poisson equation

$$-\mathcal{L}_V \chi = -\nabla_q V, \quad (6.143)$$

posed on $[0, L]^d$ with periodic boundary conditions.

Now we present several properties of the effective diffusion tensor in the over-damped limit.

Proposition 6.12.

1. The effective diffusion tensor D_0 (6.142) satisfies the upper and lower bounds

$$\frac{\beta^{-1}}{Z\widehat{Z}} \leq \langle \xi, D_0 \xi \rangle \leq \beta^{-1} |\xi|^2 \quad \forall \xi \in \mathbb{R}^d, \quad (6.144)$$

where

$$\widehat{Z} = \int_{\mathbb{T}^d} e^{\beta V(y)} dy.$$

In particular, diffusion is always depleted when compared to molecular diffusivity.

2. The diffusion tensor is symmetric.

3. In one dimension and for a potential of period L , the diffusion coefficient is given by the formula

$$D = \frac{L^2}{\beta Z\widehat{Z}}. \quad (6.145)$$

The proof of these results is left as an exercise. It is worth pointing out that the symmetry of the diffusion tensor follows from the self-adjointness of the generator of the overdamped dynamics in $[0, L]^d$, i.e., the reversibility of the “microscopic” dynamics. In other words, the symmetry of the diffusion tensor is a consequence of the reversibility of the microscopic dynamics.

The calculation of the diffusion coefficient in one dimension (6.145) requires the calculation of the integrals Z, \widehat{Z} . This can be done explicitly only for simple potentials, for example piecewise linear and piecewise constant potentials. Consider, for example, the piecewise linear potential

$$V(q) = \begin{cases} q & : q \in [0, \frac{1}{2}], \\ 1 - q & : q \in (\frac{1}{2}, 1]. \end{cases} \quad (6.146)$$

We can calculate the integrals in (6.145) to obtain the formula

$$D = \frac{\beta e^{\frac{\beta}{2}}}{4 \left(e^{\frac{\beta}{2}} - 1 \right)^2}. \quad (6.147)$$

In Fig. 6.3, we plot the diffusion coefficient given by (6.147) as a function of the temperature β^{-1} . We observe that D decays exponentially fast in the limit as $\beta \rightarrow +\infty$. This is true for arbitrary periodic potentials.

6.6.1.2 The Underdamped Limit

In this subsection we solve the Poisson equation (6.133) in one dimension perturbatively for small γ . The operator \mathcal{L}_0 that appears in (6.133) can be written in the form

$$\mathcal{L}_0 = -B + \gamma \mathcal{S}.$$

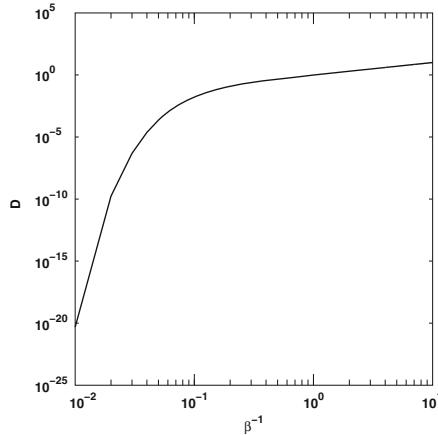


Fig. 6.3 Diffusion coefficient as a function of the temperature for the piecewise linear potential (6.146)

For the analysis of the underdamped limit that we presented in Sect. 6.5.2, we expect the solution of the Poisson equation to scale like γ^{-1} when $\gamma \ll 1$. Thus, we look for a solution of the form

$$\Phi = \frac{1}{\gamma} \phi_0 + \phi_1 + \gamma \phi_2 + \dots \quad (6.148)$$

The same analysis that leads to (6.118) enables us to obtain the integral equation for the leading term in (6.148):

$$\int_{E_{\min}}^{+\infty} \int_{-\pi}^{\pi} g(E) \left[p(q, E) + \left((\beta^{-1} - p^2) \frac{\partial}{\partial E} + \beta^{-1} p^2 \frac{\partial^2}{\partial E^2} \right) \phi_0(E) \right] \frac{1}{p(q, E)} dE dq = 0,$$

where, as in the previous section, we have made the change of variables from q, p to q, E with $p^2(E, q) = 2(E - V(q))$. Since we want to present explicit formulas for the cosine potential, we have taken V to be 2π -periodic. We will also assume that the potential has only one maximum (and consequently one minimum) per period.

Let E_0 denote the critical energy, i.e., the energy along the separatrix (homoclinic orbit). We define the action and the period (see (6.121))

$$S(E) = \int_{x_1(E)}^{x_2(E)} p(q, E) dq, \quad T(E) = \int_{x_1(E)}^{x_2(E)} \frac{1}{p(q, E)} dq,$$

where $x_1(E)$ and $x_2(E)$ denote the positions of the maximum and minimum of the potential in $[-\pi, \pi]$.

We need to consider the cases $\{E > E_0, p > 0\}$, $\{E > E_0, p < 0\}$, and $\{E_{\min} < E < E_0\}$ separately. We consider first the case $E > E_0, p > 0$. In this case, $x_1(E) = \pi$, $x_2(E) = -\pi$. We can perform the integration with respect to q to obtain

$$\int_{E_0}^{+\infty} g(E) \left[2\pi + \left((\beta^{-1}T(E) - S(E)) \frac{\partial}{\partial E} + \beta^{-1}S(E) \frac{\partial^2}{\partial E^2} \right) \phi_0(E) \right] dE = 0.$$

This equation is valid for every test function $g(E)$, from which we obtain the following differential equation for ϕ_0 :

$$-\overline{\mathcal{L}}\phi := -\beta^{-1} \frac{1}{T(E)} S(E) \phi'' + \left(\frac{1}{T(E)} S(E) - \beta^{-1} \right) \phi' = \frac{2\pi}{T(E)}, \quad (6.149)$$

where primes denote differentiation with respect to E and where the subscript 0 has been dropped for notational simplicity.

A similar calculation shows that in the regions $E > E_0$, $p < 0$, and $E_{\min} < E < E_0$, the equation for ϕ_0 is

$$-\overline{\mathcal{L}}\phi = -\frac{2\pi}{T(E)}, \quad E > E_0, p < 0, \quad (6.150)$$

and

$$-\overline{\mathcal{L}}\phi = 0, \quad E_{\min} < E < E_0. \quad (6.151)$$

The solution to equations (6.149), (6.150), (6.151) has to be square-integrable with respect to the invariant measure of the dynamics. Furthermore, it has to satisfy the following continuity condition at the critical energy:

$$2\phi'_3(E_0) = \phi'_1(E_0) + \phi'_2(E_0), \quad (6.152)$$

where ϕ_1, ϕ_2, ϕ_3 are the solutions to equations (6.149), (6.150), and (6.151), respectively.

The average of a function $h(q, p) = h(q, p(q, E))$ can be written in the form

$$\begin{aligned} \langle h(q, p) \rangle_{\beta} &:= \int_{-\infty}^{\infty} \int_{-\pi}^{\pi} h(q, p) \mu_{\beta}(q, p) dq dp \\ &= Z_{\beta}^{-1} \int_{E_{\min}}^{+\infty} \int_{x_1(E)}^{x_2(E)} \left(h(q, p(q, E)) + h(q, -p(q, E)) \right) (p(q, E))^{-1} e^{-\beta E} dE dq, \end{aligned}$$

where the partition function is

$$Z_{\beta} = \sqrt{\frac{2\pi}{\beta}} \int_{-\pi}^{\pi} e^{-\beta V(q)} dq.$$

From Eq. (6.151), we deduce that $\phi_3(E) = 0$. Furthermore, we have that $\phi_1(E) = -\phi_2(E)$. These facts, together with the above formula for the averaging with respect to the Boltzmann distribution, yield

$$\begin{aligned} D &= \langle p\Phi(p, q) \rangle_{\beta} = \langle p\phi_0 \rangle_{\beta} + \mathcal{O}(1) \\ &\approx \frac{2}{\gamma} Z_{\beta}^{-1} \int_{E_0}^{+\infty} \phi_0(z) e^{\beta E} dE + \mathcal{O}(1) \\ &= \frac{4\pi}{\gamma} Z_{\beta}^{-1} \int_{E_0}^{+\infty} \phi_0(E) e^{-\beta E} dE + \mathcal{O}(1), \end{aligned} \quad (6.153)$$

to leading order in γ , and where $\phi_0(E)$ is the solution of the two-point boundary value problem (6.149). An integration by parts yields the following formula for the leading-order term of the diffusion coefficient:

$$D = \frac{4\pi}{\gamma\beta} Z_\beta^{-1} \int_{E_0}^{+\infty} |\partial_z \phi_0(z)|^2 e^{-\beta z} dz.$$

Now we solve the equation for $\phi_0(z)$ (for notational simplicity, we will drop the subscript 0). Using the fact that $S'(E) = T(E)$, we rewrite (6.149) as

$$-\beta^{-1}(S\phi')' + S\phi' = 2\pi.$$

The solution to this equation has to be square-integrable with respect to ρ_β . The derivative of the unique (up to constants) solution of this equation, i.e., of (6.149), is

$$\phi'(E) = 2\pi S^{-1}(E).$$

We use this in (6.153), together with an integration by parts, to obtain the following formula for the diffusion coefficient:

$$D = \frac{1}{\gamma} 8\pi^2 Z_\beta^{-1} \beta^{-1} \int_{E_0}^{+\infty} \frac{e^{-\beta z}}{S(z)} dz. \quad (6.154)$$

This formula is valid for all values of the temperature and for γ sufficiently small. Consider now the case of the cosine potential $V(q) = \cos(q)$. The partition function is

$$Z_\beta = \frac{(2\pi)^{3/2}}{\beta^{1/2}} J_0(\beta),$$

where $J_0(\cdot)$ is the modified Bessel function of the first kind. Furthermore, the action S for the cosine potential can be calculated:

$$S(E) = 2^{5/2} \sqrt{E+1} I \left(\sqrt{\frac{2}{E+1}} \right),$$

where $I(\cdot)$ is the complete elliptic integral of the second kind. The formula for the diffusion coefficient becomes

$$D = \frac{1}{\gamma} \frac{\sqrt{\pi}}{2\beta^{1/2} J_0(\beta)} \int_1^{+\infty} \frac{e^{-\beta z}}{\sqrt{z+1} I(\sqrt{2/(z+1)})} dz. \quad (6.155)$$

We use now the asymptotic formula $J_0(\beta) \approx (2\pi\beta)^{-1/2} e^\beta$, $\beta \gg 1$, and the fact that $I(1) = 1$ to obtain the small-temperature asymptotics for the diffusion coefficient:

$$D = \frac{1}{\gamma} \frac{\pi}{2\beta} e^{-2\beta}, \quad \beta \gg 1. \quad (6.156)$$

Notice that the diffusion coefficient becomes exponentially small at small temperatures.

Numerical Calculation of the Diffusion Coefficient

We now have several different methods for calculating the diffusion coefficient: (a) we can perform Monte Carlo simulations (see Sect. 5.2) and use (6.125); (b) we can calculate the mean-square displacement and the hopping rate using the techniques presented in Chap. 7 and then use (6.124); (c) we can solve (6.133) and then calculate (6.134); or we can use the asymptotic formula (6.156), which is valid in the low-friction/low-temperature regime, and (6.140), which is valid in the high-friction regime, uniformly in the temperature. The Poisson equation (6.133) can be solved numerically by expanding its solution into a truncated Hermite (in p)/Fourier (in q) series:

$$\phi(p, q) = \sum_{k=-K}^K \sum_{n=0}^N \phi_{nk} e^{ikq} f_n(p), \quad (6.157)$$

where f_n denotes the n th eigenfunction of the Ornstein–Uhlenbeck generator given in Theorem 4.2. Using the fact that the $p = \beta^{-1/2} f_1$, together with the orthonormality of the eigenfunctions of the generator of the Ornstein–Uhlenbeck process, we obtain the following formula for the diffusion coefficient:

$$D = Z^{-1} \beta^{-1/2} \sum_{k=-K}^K \phi_{1k} \int_{-\pi}^{\pi} e^{ikq} e^{-\beta V(q)} dq. \quad (6.158)$$

Substituting now (6.157) and using the calculations presented in Sect. 6.4, see, e.g., Eq. (6.86), we can obtain a linear system of equations for the coefficients ϕ_{nk} . The orthonormality properties of the Hermite/Fourier basis functions enable us then to calculate the integral in (6.134). These different ways for calculating the diffusion coefficient are compared in Fig. 6.4.

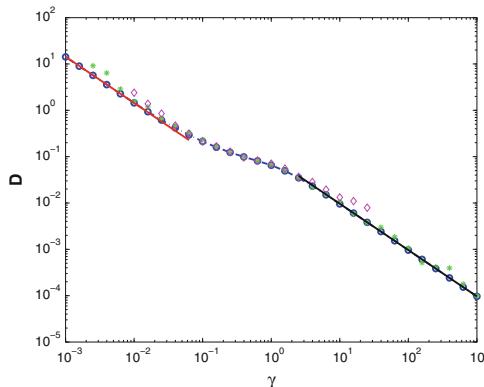


Fig. 6.4 Diffusion coefficient as a function of the friction coefficient for the cosine potential. *Dash-dot line and circles:* D obtained from the numerical solution of the Poisson equation; *stars:* D obtained from the calculation of the jump-length distribution and the hopping rate; *diamonds:* results from Monte Carlo simulations; *solid lines:* analytical approximation for $\gamma \ll 1$, $\gamma \gg 1$, equations (6.156), and (6.141).

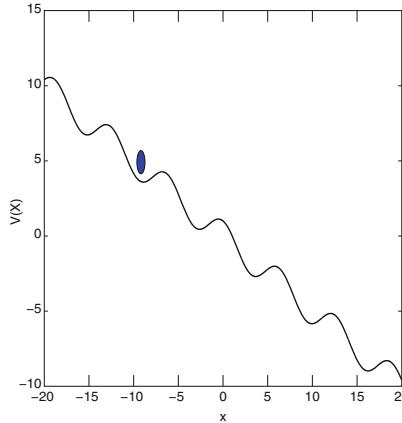


Fig. 6.5 A Brownian particle moving in the tilted periodic potential $V(x) = \cos(x) - 0.5x$

6.6.2 Brownian Motion in a Tilted Periodic Potential

As an application of the averaging and homogenization techniques used in this section, we obtain a formula for the diffusion coefficient of a Brownian particle moving in a tilted periodic potential (Fig. 6.5):

$$\dot{q} = -V'(q) + F + \sqrt{2\beta^{-1}}\dot{W}, \quad (6.159)$$

where $V(x)$ is a smooth periodic function with period L , and F is the tilt, which can be thought of as a constant external force; see Sect. 9.3. The effective potential is

$$V_{eff}(q) = V(q) - qF. \quad (6.160)$$

The long-time behavior of the dynamics (6.159) is that of a ballistic motion with effective velocity U , which is defined as

$$U = \lim_{t \rightarrow \infty} \frac{\langle q(t) \rangle}{t}, \quad (6.161)$$

with fluctuations around the mean motion described by a Brownian motion with diffusion coefficient defined in (6.125). Mathematically, we have that the rescaled process

$$q^\varepsilon := \varepsilon q(t/\varepsilon^2) - \frac{Ut}{\varepsilon} \quad (6.162)$$

converges (weakly, over the space of continuous functions) to a Brownian motion with diffusion coefficient D . Using homogenization theory (see Exercise 12), we obtain the following formulas for the drift and diffusion coefficients:

$$U = \int_0^L (-V'(q) + F) \rho(q) dq \quad (6.163)$$

and

$$D = \beta^{-1} + \int_0^L (-V'(q) + F - U)\rho(q) dq + 2\beta^{-1} \int_0^L \partial_q \chi(q)\rho(q) dq, \quad (6.164)$$

where ρ and χ are the solutions to the stationary Fokker–Planck and Poisson equations

$$\partial_q ((V'(q) - F)\rho + \beta^{-1} \partial_q \rho) = 0 \quad (6.165)$$

and

$$-((-V'(q) + F)\partial_q \chi + \beta^{-1} \partial_q^2 \chi) = -V'(q) + F - U. \quad (6.166)$$

Both equations are posed on $[0, L]$ with periodic boundary conditions.

The solution of Eq. (6.165) is

$$\rho(q) = \frac{1}{Z} \int_q^{q+L} dy Z_+(y) Z_-(q), \quad (6.167)$$

with

$$Z_{\pm}(q) := e^{\pm \beta(V(q) - Fq)}, \quad Z = \int_0^L dq \int_q^{q+L} dy Z_+(y) Z_-(q). \quad (6.168)$$

Using (6.167) in (6.163), we obtain the following formula for the effective drift:

$$U = \frac{L}{\beta Z} \left(1 - e^{-\beta F L} \right). \quad (6.169)$$

The reader is asked to provide the details of these results in Exercise 13.

Now we calculate the diffusion coefficient. The Dirichlet form associated with the generator \mathcal{L} of the dynamics (6.159), restricted on the interval $[0, L]$ with periodic boundary conditions, gives

$$\int_0^L \phi(q)(-\mathcal{L}\phi(q))\rho(q) dq = \beta^{-1} \int_0^L (\partial_q \phi(q))^2 \rho(q) dq.$$

Now we calculate

$$\begin{aligned} D &= \beta^{-1} + \int_0^L (-V'(q) + F - U)\chi(q)\rho(q) dq + 2\beta^{-1} \int_0^L \partial_q \chi(q)\rho(q) dq \\ &= \beta^{-1} + \int_0^L (-\mathcal{L}\chi(q))\chi(q)\rho(q) dq + 2\beta^{-1} \int_0^L \partial_q \chi(q)\rho(q) dq \\ &= \beta^{-1} + \beta^{-1} \int_0^L (\partial_q \chi(q))^2 \rho(q) dq + 2\beta^{-1} \int_0^L \partial_q \chi(q)\rho(q) dq \\ &= \beta^{-1} \int_0^L (1 + \partial_q \chi(q))^2 \rho(q) dq. \end{aligned} \quad (6.170)$$

Now we solve the Poisson equation (6.166) with periodic boundary conditions. We multiply the equation by $Z_-(q)$ and by β to rewrite it in the form

$$\partial_q(\partial_q \chi(x) Z_-(q)) = -\partial_q Z_-(q) + \beta U Z_-(q).$$

We integrate this equation from $q - L$ to q and use the periodicity of $\chi(q)$ and $V(q)$ together with formula (6.169) to obtain

$$\partial_q \chi(q) Z_-(q) \left(1 - e^{-\beta F L}\right) = -Z_-(q) \left(1 - e^{-\beta F L}\right) + \frac{L}{Z} \left(1 - e^{-\beta F L}\right) \int_{q-L}^q Z_-(y) dy,$$

from which we immediately get

$$\partial_q \chi(q) + 1 = \frac{1}{Z} \int_{q-L}^q Z_-(y) Z_+(q) dy.$$

Substituting this into (6.170) and using the formula for the invariant distribution (6.167), we finally obtain

$$D = \frac{1}{\beta Z^3} \int_0^L (I_+(q))^2 I_-(q) dq, \quad (6.171)$$

with

$$I_+(q) = \int_{q-L}^q Z_-(y) Z_+(q) dy \quad \text{and} \quad I_-(q) = \int_q^{q+L} Z_+(y) Z_-(x) dy.$$

The formula for the diffusion coefficient (6.171) requires the calculation of a complicated multiple integral. This integral can be calculated analytically only in very particular cases.

6.7 Discussion and Bibliography

The Fokker–Planck equation for the Langevin equation, or the *Klein–Kramers–Chandrasekhar equation* was first derived by Klein in 1923 and was studied by Kramers [122]. See [35] and [206]. Rigorous results on the Fokker–Planck equation, including existence and uniqueness of solutions, exponential convergence to equilibrium, and spectral analysis, can be found in [85].

Our presentation in Sect. 6.2.1 follows [212]. See also [177]. Hörmander’s theorem appears in [92]. C. Villani’s theory of hypocoercivity is developed in [237]. Convergence to equilibrium for the Langevin equation and the Fokker–Planck equation is studied in [43, 86]. Further information on the spectral theory of hypoelliptic operators can be found in [49, 85]. Hypoellipticity and hypocoercivity are concerned with two different problems, the study of regularity of solutions to elliptic and parabolic PDEs (hypoellipticity) and the problem of exponentially fast convergence to equilibrium for diffusion processes (hypocoercivity). It is perhaps surprising that both theories are based on the use of commutator techniques.

The phase-space Fokker–Planck equation can be written in the form

$$\frac{\partial \rho}{\partial t} + p \cdot \nabla_q \rho - \nabla_q V \cdot \nabla_p \rho = Q(\rho, \rho_\beta),$$

where the *collision operator* has the form

$$Q(\rho, \rho_\beta) = \beta^{-1} \nabla \cdot \left(\rho_\beta \nabla \left(\rho_\beta^{-1} \rho \right) \right).$$

This, the Fokker–Planck equation, has a similar structure to the Boltzmann equation, which is the fundamental equation in the kinetic theory of gases. Of course, the collision operator in the Boltzmann equation is quadratic, whereas the collision operator in the Fokker–Planck equation is linear. One might think, in principle, of the Fokker–Planck collision operator as a linearization around equilibrium, i.e., the Maxwell–Boltzmann distribution. More information can be found in [13, 199]. Convergence to equilibrium for the Fokker–Planck equation with careful control of the rate of convergence to equilibrium was proved in [43]. A similar methodology, albeit involving much more difficult computations, can also be applied to the Boltzmann equation; see [44].

The set of equations (6.86) is usually referred to in the physics literature as *Brinkman's hierarchy*. A semianalytical method for solving the Fokker–Planck equation based on this expansion, the *continued fraction expansion*, is studied extensively in [206]. A combined spectral–finite-difference method for solving the Fokker–Planck equation is developed and analyzed in [58]. The Hermite expansion of the distribution function with respect to the velocity is used in the study of various kinetic equations including the Boltzmann equation. It was initiated by Grad; see, e.g. [74]. It is quite often used in the approximate calculation of transport coefficients, e.g., the diffusion coefficient. See [13]. This expansion can be justified rigorously for the Fokker–Planck equation [166]. See also [165, 215].

The calculation of the spectrum for the Fokker–Planck operator corresponding to the Langevin equation in a harmonic potential can be found in [85, 206]. The technique of calculating the spectrum of a differential operator using appropriate creation and annihilation operators is standard in quantum mechanics [232]. It is possible to calculate the spectrum of the generator (or the Fokker–Planck operator) of arbitrary linear SDEs. Consider a linear SDE in \mathbb{R}^d ,

$$dX_t = AX_t dt + \sigma dW_t, \quad (6.172)$$

where $A, \sigma \in \mathbb{R}^{d \times d}$, and W_t is a standard d –dimensional Brownian motion. The generator of X_t is

$$\mathcal{L} = Ax \cdot \nabla + \frac{1}{2} \sum_{i,j=1}^d \Sigma_{ij} \frac{\partial^2}{\partial x_i \partial x_j}, \quad (6.173)$$

where $\Sigma = \sigma \sigma^T$, a positive semidefinite matrix, such that the generator \mathcal{L} is hypoelliptic. When A is a stable matrix, i.e., all its eigenvalues have negative real parts, X_t is an ergodic Markov process with a Gaussian invariant measure with

covariance matrix Σ_∞ ; see Eq. (3.105). An explicit formula for the spectrum of \mathcal{L} , valid in all L^p spaces weighted by the invariant measure with $p > 1$, was obtained in [164]. The spectrum consists of integer linear combinations of the eigenvalues of the drift matrix A :

$$\sigma(-\mathcal{L}) = \left\{ -\sum_{\mu \in \sigma(A)} \mu k_\mu, \quad k_\mu \in \mathbb{N} \right\}. \quad (6.174)$$

Furthermore, the spectrum of \mathcal{L} in the weighted L^1 space consists of the entire left half-plane. Note that the spectrum is independent of the diffusion matrix Σ (of course, the function space in which we calculate the spectrum does depend on Σ through the invariant measure of X_t). The generator of the Langevin dynamics in a harmonic potential $V(q) = \frac{1}{2} \omega_0^2 q^2$ can be written in the form (6.173) with

$$A = \begin{bmatrix} 0 & 1 \\ -\omega_0^2 & -\gamma \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 0 & 0 \\ 0 & 2\gamma\beta \end{bmatrix}. \quad (6.175)$$

Calculating the eigenvalues of B and using (6.174), we recover the formula obtained in Sect. 6.3. The spectral gap of the generator attains its maximum (which means that the dynamics converges to equilibrium as quickly as possible) when the friction coefficient is chosen at the critical value $\gamma = 2\omega_0$:

$$\sigma_{opt} = \frac{\gamma}{2}, \quad \text{for } \gamma = 2\omega_0.$$

Further discussion can be found in [180], where an alternative proof of the results in [164], for the case $p = 2$, can be found.

A systematic study of the overdamped limit for the Fokker–Planck equation, including the calculation of higher-order terms in the asymptotic expansion for the Fokker–Planck equation, can be found in [233, 242, 243]. See also [171, Chap. 10]. A similar approximation theorem is also valid in infinite dimensions, i.e., for stochastic partial differential equations; see [33, 34].

The state space for the limiting stochastic differential equation for the Hamiltonian in the underdamped limit, Eq. (6.109), is the graph associated with the Hamiltonian system: we construct this graph by identifying all points belonging to the same connected component of a level curve of the Hamiltonian, $\{x : H(x) = H\}$, $x = (q, p)$. Each point on the edges of the graph corresponds to a trajectory. Interior vertices correspond to separatrices. Let I_i , $i = 1, \dots, d$ be the edges of the graph. Then (i, H) defines a global coordinate system on the graph. The domain of definition of the generator of the limiting Markov process is defined through appropriate boundary conditions, the *gluing conditions*, at the interior vertices of the graph. Rigorous results on the underdamped limit of the Langevin dynamics can be found in [61, 62, 222]. The Fokker–Planck equation (6.122), valid in the energy diffusion regime, was derived in [122]. See also [82, 246].

Section 6.6 is based on [183]; see also [79] and [120, 210]. Averaging and homogenization for stochastic differential equations and for partial differential

equations can be found in [185], where the proof of Proposition 6.12 and the calculation of the diffusion coefficient for a piecewise constant potential can be found. See also [17, 108]. The fact that the diffusively rescaled Langevin dynamics in a periodic potential converges weakly to a Brownian motion is a form of the functional central limit theorem. Similar results can be proved for random potentials and for different types of dynamics. See [118]. The symmetry of the diffusion tensor when the underlying dynamics are reversible is a manifestation of *Onsager's reciprocal relations*. This is a general result in nonequilibrium thermodynamics; see [41]. A detailed study of the Fokker–Planck equation for the Langevin dynamics in a periodic potential can be found in [206, Chap. 11]. Material on special functions such as the Bessel function and elliptic integrals can be found in many books on applied mathematics. Information on asymptotic techniques for the evaluation of integrals can be found in [25].

The derivation of a formula for the diffusion coefficient by calculating the escape rate from the local minima of the periodic potential (using Kramers's theory; see Sect. 7.3) in the overdamped limit can be found in [140]. At low temperatures, we can use the method of steepest descent to obtain an approximate formula for the diffusion coefficient. This can be done rigorously and with control of the error terms, in arbitrary dimensions. Details can be found in [121].

The diffusion coefficient for a Brownian particle in a tilted periodic potential was first derived and analyzed in [197, 198] using different techniques. In these papers, it was shown that as the tilt F approaches the critical tilt $F_c = \max_{x \in [0, L]} |V'(x)|$ that separates diffusive and ballistic motion, a giant enhancement of the diffusion coefficient can take place. This phenomenon becomes more pronounced at low temperatures. Section (6.6.2) is based on [184].

The diffusion coefficient is an example of a *transport coefficient* that relates thermodynamic forces and fluxes in (linear) constitutive relations in nonequilibrium thermodynamics. Other examples of transport coefficients include the shear viscosity and the thermal conductivity. Transport coefficients can be expressed in terms of time integrals of equilibrium autocorrelation functions of appropriately chosen observables. This is the context of the *Green–Kubo formalism*. The equivalence between homogenization theory and the Green–Kubo formalism for the calculation of the diffusion coefficient is studied in Sect. 9.3.

6.8 Exercises

1. Consider the stochastic differential equation

$$dX_t = X_t \times dW_t \quad (6.176)$$

on \mathbb{R}^3 . Write the generator and the Fokker–Planck operators. Check whether Hörmander's theorem applies. Does the solution of this equation X_t have a smooth density?

2. Consider the second-order stochastic differential equation

$$\ddot{X}_t = b(X_t) - \gamma(X_t)\dot{X}_t + \dot{W}_t \quad (6.177)$$

on \mathbb{R}^d , where b and γ are a smooth vector field and scalar function, respectively.

Investigate under what assumptions on b, γ the process X_t has a smooth density.

3. Consider the system of stochastic differential equations

$$\begin{aligned}\ddot{X}_t &= b(X_t)\eta_t - \dot{X}_t + \sigma\dot{W}_t^1, \\ \dot{\eta}_t &= -\eta_t + \dot{W}_t^2,\end{aligned}$$

where X_t is a two-dimensional process, b a smooth vector field, $\sigma > 0$ a constant, and W_t^1, W_t^2 two-dimensional and one-dimensional, respectively, independent Brownian motions. Show that the process $\{X_t, \eta_t\}$ has a smooth density. Is this true if $\sigma = 0$?

4. Consider the Langevin dynamics in a weakly anharmonic potential

$$V(q) = \frac{q^2}{2} + \delta\frac{q^4}{4}. \quad (6.178)$$

Use perturbation theory and the results in Sect. 6.3 to obtain approximate formulas for the eigenvalues of the generator of the Langevin dynamics that are valid when $\delta \ll 1$. Use the expansion in Hermite polynomials presented in Sect. 6.4 to develop a numerical scheme for calculating the eigenvalues of the generator. Compare the approximate analytical results and the results of your numerical experiments.

5. Let \mathcal{L}_{kin} be the operator defined in (6.74).

- a. Show by direct substitution that $\widehat{\mathcal{L}}$ can be written in the form

$$\mathcal{L}_{kin} = -\lambda_1(c^-)^*(c^+)^* - \lambda_2(d^-)^*(d^+)^*.$$

- b. Calculate the commutators

$$[(c^+)^*, (c^-)^*], \quad [(d^+)^*, (d^-)^*], \quad [(c^\pm)^*, (d^\pm)^*], \quad [\mathcal{L}_{kin}, (c^\pm)^*], \quad [\mathcal{L}_{kin}, (d^\pm)^*].$$

6. Let \mathcal{L} denote the generator of the Langevin equation in a harmonic potential given by (6.58), and let c^\pm, d^\pm denote the operators defined in (6.68). Show that

$$[\mathcal{L}, (c^\pm)^n] = -n\lambda_1(c^\pm)^n, \quad [\mathcal{L}, (d^\pm)^n] = -n\lambda_1(d^\pm)^n, \quad (6.179a)$$

$$[c^-, (c^+)^n] = n(c^+)^{n-1}, \quad [d^-, (d^+)^n] = n(d^+)^{n-1}. \quad (6.179b)$$

7. Consider the Langevin equation

$$\ddot{q} = -Bq - \Gamma\dot{q} + \sqrt{2\beta^{-1}\Gamma}\dot{W}, \quad (6.180)$$

where $B, \Gamma \in \mathbb{R}^{d \times d}$ are positive definite matrices.

- Use the results from Sect. (3.7) to solve this equation and to calculate the mean and the covariance matrix.
 - Consider (6.180) in two dimensions. Calculate the eigenvalues of the generator as a function of the eigenvalues of B and Γ . Find the optimal choice of the matrix Γ that leads to the fastest convergence of the dynamics to equilibrium. Test your theoretical results against numerical experiments.
8. a. Study the overdamped limit for the Langevin dynamics with a space-dependent friction in arbitrary dimensions:

$$\varepsilon^{-2} \ddot{q} = F(q) - \Gamma(q) \dot{q} + \sqrt{2\beta^{-1}\Gamma(q)} \dot{W},$$

where $\Gamma(\cdot)$ is a $d \times d$ strictly positive definite matrix-valued function.

- b. Study the same problem in the case that the fluctuation–dissipation theorem does not hold:

$$\varepsilon^{-2} \ddot{q} = F(q) - \Gamma(q) \dot{q} + \sqrt{2\Sigma(q)} \dot{W},$$

with $\Sigma(q) \neq \Gamma(q)\Gamma^T(q)$, strictly positive definite.

- c. In what sense should (i.e., Itô, Stratonovich, etc.) should we interpret the stochastic integral in the limiting equation (see Eq. (5.110) and Exercise 1 in Chap. 5)?

9. Obtain the second term in the expansion (6.140).
10. Prove Proposition 6.12.
11. Use the method of steepest descent to study the low-temperature asymptotics of the diffusion coefficient for a Brownian particle moving in a periodic potential, in the overdamped limit in 1 dimension. Can you perform similar calculations in higher dimensions?
12. Use homogenization theory to study the problem of Brownian motion in a tilted periodic potential. In particular, obtain the formulas for the drift and diffusion coefficients (6.163) and (6.164), respectively. Study the dependence of the drift and diffusion coefficients on the forcing F by evaluating numerically the integrals in these formulas. Consider both symmetric and anisymmetric periodic potentials.
13. Solve the stationary Fokker–Planck equation (6.165) on $[0, L]$ with periodic boundary conditions. Calculate the integrals in (6.167) to obtain (6.163) (see [225, Chap. 9] if necessary).
14. Study the drift and diffusion coefficients of a Brownian particle moving in a tilted periodic potential for piecewise constant and piecewise linear potentials.

Chapter 7

Exit Problems for Diffusion Processes and Applications

In this chapter, we develop techniques for calculating the statistics of the time that it takes for a diffusion process in a bounded domain to reach the boundary of the domain. We then use this formalism to study the problem of Brownian motion in a bistable potential. Applications such as stochastic resonance and the modeling of Brownian motors are also presented. In Sect. 7.1, we motivate the techniques that we will develop in this chapter by looking at the problem of Brownian motion in bistable potentials. In Sect. 7.2, we obtain a boundary value problem for the mean exit time of a diffusion process from a domain. We then use this formalism in Sect. 7.3 to calculate the escape rate of a Brownian particle from a potential well. The phenomenon of stochastic resonance is investigated in Sect. 7.4. Brownian motors are studied in Sect. 7.5. Bibliographical remarks and exercises can be found in Sects. 7.6 and 7.7, respectively.

7.1 Brownian Motion in a Double-Well Potential

In this section, we study a simple dynamical stochastic system that can exist at two different (meta)stable states. Our goal is to understand how noise enables such a system to jump from one metastable state to another and to calculate how long it will take on average for this transition to occur.

We look at a Brownian particle moving in a double-well potential under the influence of thermal noise in one dimension, the problem that we studied briefly in Sect. 5.2:

$$dX_t = -V'(X_t) dt + \sqrt{2\beta^{-1}} dW_t, \quad (7.1)$$

with the bistable potential

$$V(x) = \frac{1}{4}x^4 - \frac{1}{2}x^2 + \frac{1}{4}. \quad (7.2)$$

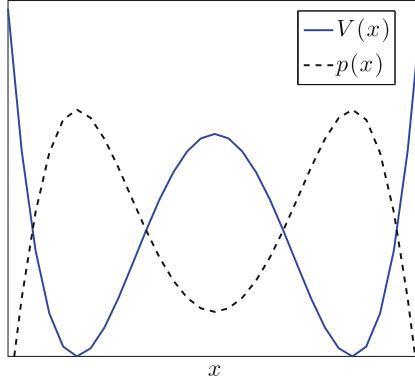


Fig. 7.1 Bistable potential (7.2) and invariant distribution (7.4)

This potential has three extrema: a local maximum at $x = 0$ and two local minima at $x = \pm 1$. The values of the potential at these three points are $V(\pm 1) = 0$, $V(0) = \frac{1}{4}$. We will say that the height of the potential barrier is $\Delta V = \frac{1}{4}$. We are interested in understanding the dynamics (7.1) in the asymptotic regime where the thermal fluctuations, whose strength is measured by the temperature β^{-1} , are weak compared to the potential barrier ΔV :

$$\frac{1}{\beta \Delta V} \ll 1. \quad (7.3)$$

As we have already seen, the dynamics (7.1) is ergodic with respect to the distribution

$$\rho_s(x) = \frac{1}{Z} e^{-\beta V(x)}. \quad (7.4)$$

At low temperatures, $\beta \gg 1$, most of mass of the invariant distribution is concentrated around the minima of the potential; see Fig. 7.1. It is expected that stationary trajectories of X_t will spend most time oscillating around the two local minima of the potential, while occasionally hopping between the two local minima of the potential. This intuition is confirmed by performing numerical simulations; see Fig. 7.2. This is a noise-assisted event: in the absence of noise, the process X_t ends up at one of the two minima of the potential, depending on its initial condition. Indeed, it is easy to check that the potential itself is a Lyapunov function for the deterministic dynamics. For the noise dynamics, we will refer to the two local minima of the potential as *metastable states*.

The time that it takes for the “particle” X_t to acquire a sufficient amount of energy from the noise so that it can surmount the potential barrier ΔV and escape from one of the metastable states depends on the strength of the noise, i.e., the temperature. When the noise is weak, the particle spends a long time at the metastable state (relative to the time scale introduced by Assumption 7.3), before being able to escape from it. This is an example of a *rare event*. The relevant time scale, the *mean exit time* or the *mean first passage time*, scales exponentially in β :

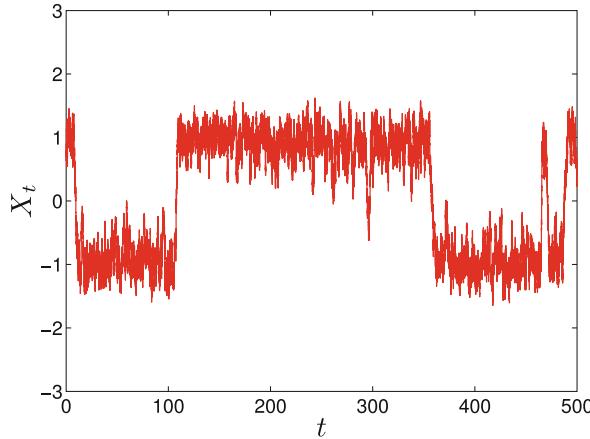


Fig. 7.2 Sample path of X_t , the solution of (7.1) with the bistable potential (7.2)

$$\tau = v^{-1} \exp(\beta \Delta V). \quad (7.5)$$

We will refer to this as the *Kramers time*. The inverse of the Kramers time is proportional to the *reaction rate* (hopping rate) $\kappa \sim \tau^{-1}$, which gives the rate at which particles escape from a local minimum of the potential:

$$\kappa \sim v \exp(-\beta \Delta V). \quad (7.6)$$

The prefactor v is called the *rate coefficient*. This hopping mechanism between metastable states becomes less pronounced at higher temperatures. For β sufficiently low, the dynamics (7.1) is dominated by noise, and transitions between the two metastable states cease to be rare events.

One of our goals in this chapter will be to obtain (7.5) in a systematic way and furthermore, to obtain a formula for the prefactor v . More generally, we want to consider stochastic dynamical systems of the form (7.1) that possess several metastable states. We want to characterize transitions between these states and to calculate transition rates. This will be done in Sects. 7.2 and 7.3. Later in this chapter, we will also consider the effect of adding a time-periodic external forcing to stochastic systems with metastable states. This will be done in Sects. 7.4 and 7.5.

7.2 The Mean Exit Time

In order to calculate the hopping rate κ for the dynamics (7.1) in a metastable potential (7.2), we need to calculate the time it takes on average for the diffusion process X_t to escape from one of the local minima of the potential, or more generally, the time it takes on average for a diffusion process to escape from a metastable state. This *mean exit time* from a metastable state is an example of a *mean first passage*

time (MFPT): we want to calculate how long it takes on average for a diffusion process to reach the boundary of a domain. When the domain is the basin of attraction of one of the local minima of the potential, the mean first passage time gives us the average time it takes for the diffusing particle to reach the local maximum of the potential.

We can calculate the mean exit time τ for a diffusion process X_t in a systematic way by showing that it is the solution of a boundary value problem that involves the generator of the process X_t . This boundary value problem, see equation (7.9), can be justified rigorously using Dynkin's formula (3.110). In this section, we present a formal derivation of this equation that will be sufficient for our purposes.

The Boundary Value Problem for the Mean Exit Time

Let X_t^x denote the solution of the stochastic differential equation

$$dX_t^x = b(X_t^x) dt + \sigma(X_t^x) dW_t, \quad X_0^x = x, \quad (7.7)$$

in \mathbb{R}^d , and let D be a bounded subset of \mathbb{R}^d with smooth boundary. We have introduced the superscript x to emphasize the dependence of the solution to the SDE on the initial point x . Given $x \in D$, we define the *first passage time* or *first exit time* to be the first time that X_t^x exits the domain D :

$$\tau_D^x = \inf \{t \geq 0 : X_t^x \notin D\}.$$

This is an example of a *stopping time* (see Sect. 3.8): the information that we have about our stochastic process up to time t is sufficient to determine whether the event $\tau \leq t$ has occurred. The average of this random variable is called the mean first passage time or the mean exit time:

$$\tau(x) := \mathbb{E} \tau_D^x = \mathbb{E} \left(\inf \{t \geq 0 : X_t^x \notin D\} \mid X_0^x = x \right).$$

We have written the second equality in the above in order to emphasize the fact that the mean first passage time is defined in terms of a conditional expectation, i.e., the mean exit time is defined as the expectation of the first time the diffusion processes X_t leaves the domain, conditioned on X_t starting at $x \in \Omega$. Consequently, the mean exit time is a function of the starting point x . Consider now an ensemble of initial conditions distributed according to a distribution $p_0(x)$. The *confinement time* is defined as

$$\bar{\tau} = \int_{\Omega} \tau(x) p_0(x) dx = \int_{\Omega} \mathbb{E} \left(\inf \{t \geq 0 : X_t^x \notin D\} \mid X_0^x = x \right) p_0(x) dx. \quad (7.8)$$

We can calculate the mean exit time by solving an appropriate boundary value problem. The calculation of the confinement time follows, then, by calculating the integral in (7.8).

Result 7.1 *The mean exit time is given by the solution of the boundary value problem*

$$-\mathcal{L}\tau = 1, \quad x \in D, \quad (7.9a)$$

$$\tau = 0, \quad x \in \partial D, \quad (7.9b)$$

where \mathcal{L} is the generator of the diffusion process 7.7.

The homogeneous Dirichlet boundary conditions correspond to an absorbing boundary: the particles are removed when they reach the boundary. Other choices of boundary conditions are also possible; see Eq. (7.11).

Derivation of Result 7.1. Let $\rho(y, t|x)$ be the probability distribution of the particles that *have not left* the domain D at time t . It satisfies the Fokker–Planck equation with absorbing boundary conditions:

$$\frac{\partial \rho}{\partial t} = \mathcal{L}^* \rho, \quad \rho(y, 0|x) = \delta(y - x), \quad \rho|_{\partial D} = 0, \quad (7.10)$$

where \mathcal{L}^* is a differential operator in y . We can write the solution to this equation in the form

$$\rho(y, t|x) = e^{\mathcal{L}^* t} \delta(y - x),$$

where the absorbing boundary conditions are included in the definition of the semi-group $e^{\mathcal{L}^* t}$. The homogeneous Dirichlet (absorbing) boundary conditions imply that

$$\lim_{t \rightarrow +\infty} \rho(y, t|x) = 0.$$

That is, all particles will eventually leave the domain. The (normalized) number of particles that are still inside D at time t is

$$S(x, t) = \int_D \rho(y, t|x) dy.$$

Note that this is a decreasing function of time. We can write

$$\frac{\partial S}{\partial t} = -f(x, t),$$

where $f(x, t)$ is the *first passage time distribution*. The mean exit time is the first moment of the distribution $f(x, t)$:

$$\begin{aligned} \tau(x) &= \int_0^{+\infty} f(x, s) s ds = \int_0^{+\infty} -\frac{dS}{ds} s ds \\ &= \int_0^{+\infty} S(s, x) ds = \int_0^{+\infty} \int_D \rho(y, s|x) dy ds \\ &= \int_0^{+\infty} \int_D e^{\mathcal{L}^* s} \delta(y - x) dy ds \\ &= \int_0^{+\infty} \int_D \delta(y - x) (e^{\mathcal{L} s} 1) dy ds = \int_0^{+\infty} (e^{\mathcal{L} s} 1)(x) ds. \end{aligned}$$

We apply \mathcal{L} to the above equation to deduce

$$\begin{aligned}\mathcal{L}\tau &= \int_0^{+\infty} (\mathcal{L}e^{\mathcal{L}t}1) dt = \int_0^{+\infty} \frac{d}{dt} (e^{\mathcal{L}t}1) dt \\ &= -1.\end{aligned}$$

□

When a part of the boundary is absorbing and a part is reflecting, then we end up with a mixed boundary value problem for the mean exit time:

$$-\mathcal{L}\tau = 1, \quad x \in D, \quad (7.11a)$$

$$\tau = 0, \quad x \in \partial D_A, \quad (7.11b)$$

$$\eta \cdot \mathbf{J} = 0, \quad x \in \partial D_R. \quad (7.11c)$$

Here $\partial D_A \cup \partial D_R = \partial D$, where $\partial D_A \neq \emptyset$ denotes the absorbing part of the boundary, ∂D_R denotes the reflecting part, and \mathbf{J} denotes the probability flux.

7.2.1 Examples

We can study now a few simple examples for which we can calculate the mean first passage time in closed form.

Brownian Motion with One Absorbing and One Reflecting Boundary

We consider the problem of Brownian motion (with diffusion coefficient 2) moving in the interval $[a, b]$. We assume that the left boundary is absorbing and the right boundary is reflecting. The boundary value problem for the mean exit time becomes

$$-\frac{d^2\tau}{dx^2} = 1, \quad \tau(a) = 0, \quad \frac{d\tau}{dx}(b) = 0. \quad (7.12)$$

The solution of this equation is

$$\tau(x) = -\frac{x^2 - a^2}{2} + b(x - a).$$

The mean exit time for Brownian motion with one absorbing and one reflecting boundary in the interval $[-1, 1]$ is plotted in Fig. 7.3a.

Brownian Motion with Two Absorbing Boundaries

Consider again the problem of Brownian motion with diffusion coefficient 2 moving in the interval $[a, b]$, but now with both boundaries absorbing. The boundary value problem for the MFPT time becomes

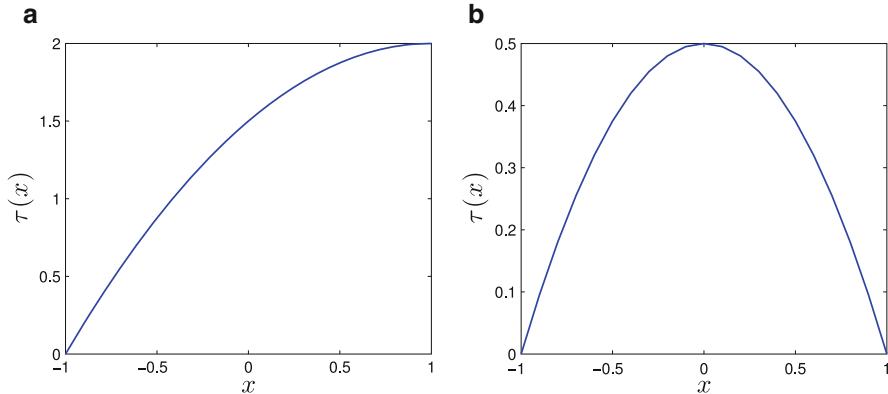


Fig. 7.3 The mean exit time for Brownian motion with one absorbing and one reflecting boundary (a) and two absorbing boundaries (b)

$$-\frac{d^2\tau}{dx^2} = 1, \quad \tau(a) = 0, \quad \tau(b) = 0. \quad (7.13)$$

The solution of this equation is

$$\tau(x) = -\frac{x^2}{2} + \frac{a+b}{2}x - \frac{ab}{2}.$$

The MFPT time for Brownian motion with two absorbing boundaries in the interval $[-1, 1]$ is plotted in Fig. 7.3b.

The Mean First Passage Time for a One-Dimensional Diffusion Process

Consider now the mean exit time problem from an interval $[a, b]$ for a general one-dimensional diffusion process with generator

$$\mathcal{L} = b(x) \frac{d}{dx} + \frac{1}{2} \Sigma(x) \frac{d^2}{dx^2},$$

where the drift and diffusion coefficients are smooth functions and where the diffusion coefficient $\Sigma(x)$ is a strictly positive function (uniform ellipticity condition). In order to calculate the mean first passage time, we need to solve the differential equation

$$-\left(b(x) \frac{d}{dx} + \frac{1}{2} \Sigma(x) \frac{d^2}{dx^2}\right) \tau = 1, \quad (7.14)$$

together with appropriate boundary conditions, depending on whether we have one absorbing and one reflecting boundary or two absorbing boundaries. To solve this equation, we first define the function $\psi(x)$ through $\psi'(x) = 2b(x)/\Sigma(x)$ to write (7.14) in the form

$$\left(e^{\psi(x)} \tau'(x)\right)' = -\frac{2}{\Sigma(x)} e^{\psi(x)}.$$

The general solution of (7.14) is obtained after two integrations:

$$\tau(x) = -2 \int_a^x e^{-\psi(z)} dz \int_a^z \frac{e^{\psi(y)}}{\Sigma(y)} dy + c_1 \int_a^x e^{-\psi(y)} dy + c_2,$$

where the constants c_1 and c_2 are to be determined from the boundary conditions. When both boundaries are absorbing, we get

$$c_1 = \frac{2 \int_a^b e^{-\psi(z)} dz \int_a^z \frac{e^{\psi(y)}}{\Sigma(y)} dy}{\int_a^b e^{-\psi(y)} dy}, \quad c_2 = 0,$$

whereas when the left boundary is absorbing and the right is reflecting, we have

$$c_1 = 2 \int_a^b \frac{e^{\psi(y)}}{\Sigma(y)} dy, \quad c_2 = 0.$$

7.3 Escape from a Potential Well

Now we can use the theory developed in the previous section to calculate the escape rate and the mean exit time from a metastable state for a particle moving in a double-well potential of the form (7.2) for the overdamped Langevin dynamics (7.1).

We assume that the left and right minima of the potential are located at $x = a$ and $x = c$, respectively; the local maximum is at $x = b$, $a < b < c$. We will calculate the rate of escape from the left minimum. For this, we need to know how long it will take, on average, for a particle starting close to the minimum a to reach the local maximum.

We assume that the particle is initially at x_0 , which is near a . The boundary value problem for the mean exit time from the interval (a, b) for the one-dimensional diffusion process (7.1) reads

$$-\beta^{-1} e^{\beta V} \frac{d}{dx} \left(e^{-\beta V} \frac{d}{dx} \tau \right) = 1. \quad (7.15)$$

In view of the fact that the particle cannot move too much to the left, since the potential is confining, we choose reflecting boundary conditions at $x = a$. We also choose absorbing boundary conditions at $x = b$, since we are assuming that the particle escapes the left minimum when it reaches the point b . We can solve (7.15) with these boundary conditions by quadratures:

$$\tau(x) = \beta \int_x^b dy e^{\beta V(y)} \int_a^y dz e^{-\beta V(z)}. \quad (7.16)$$

The potential grows sufficiently fast at infinity to allow us to replace the boundary conditions at $x = a$ by a repelling/reflecting boundary condition at $x = -\infty$:¹

¹ In other words, the integral $\int_{-\infty}^a e^{-\beta V(y)} dy$ can be neglected.

$$\tau(x) \approx \beta \int_x^b dy e^{\beta V(y)} \int_{-\infty}^y dz e^{-\beta V(z)}.$$

When $\Delta V \beta \gg 1$, the integral with respect to z is dominated by the value of the potential near a . We can use the Taylor series expansion around the minimum:

$$V(z) = V(a) + \frac{1}{2} \omega_a^2 (z - a)^2 + \dots$$

Furthermore, we can replace the upper limit of integration by $+\infty$:

$$\begin{aligned} \int_{-\infty}^y e^{-\beta V(z)} dz &\approx \int_{-\infty}^{+\infty} e^{-\beta V(a)} e^{-\frac{\beta \omega_a^2}{2} (z-a)^2} dz \\ &= e^{-\beta V(a)} \sqrt{\frac{2\pi}{\beta \omega_a^2}}. \end{aligned}$$

Similarly, the integral with respect to y is dominated by the value of the potential around the local maximum b . We use the Taylor series expansion

$$V(y) = V(b) - \frac{1}{2} \omega_b^2 (y - b)^2 + \dots$$

Assuming that x is close to the left local minimum a , we can replace the lower limit of integration by $-\infty$. We have

$$\begin{aligned} \int_x^b e^{\beta V(y)} dy &\approx \int_{-\infty}^b e^{\beta V(b)} e^{-\frac{\beta \omega_b^2}{2} (y-b)^2} dy \\ &= \frac{1}{2} e^{\beta V(b)} \sqrt{\frac{2\pi}{\beta \omega_b^2}}. \end{aligned}$$

Putting everything together, we obtain the following formula for the mean exit time:

$$\tau = \frac{\pi}{\omega_a \omega_b} e^{\beta \Delta V}. \quad (7.17)$$

This is independent of the point x , provided that it is close to the local minimum a .

The rate of arrival at the local maximum b is $1/\tau$. Once a particle has reached b , it has a 50% chance of moving to the left and a 50% of moving to the right. In other words, only half of the particles that reach b manage to escape. Consequently, the escape rate (or reaction rate) for x in the vicinity of a is given by $\frac{1}{2\tau}$:

$$\kappa = \frac{\omega_a \omega_b}{2\pi} e^{-\beta \Delta V}. \quad (7.18)$$

We will refer to (7.17) as the *Kramers time* and to (7.18) as the *Kramers rate*.

Example 7.1. We can approximate a double-well potential of the form (7.2) by either a piecewise linear or a piecewise constant potential. Consider a symmetric bistable potential with minima at $x = \pm L$ and maximum at $x = 0$. We consider the piecewise linear potential

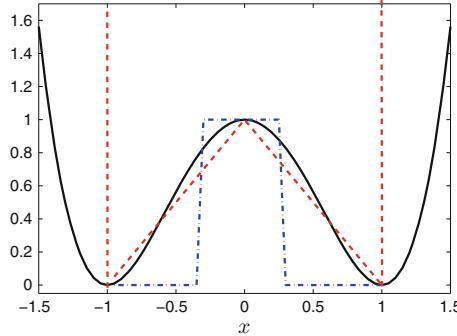


Fig. 7.4 Piecewise linear and piecewise constant approximations of a symmetric double-well potential

$$V_L(x) = \begin{cases} \frac{\delta}{L}(L-x) & x \in [0, L], \\ \frac{\delta}{L}(L+x) & x \in [-L, 0], \\ +\infty & x \notin [-L, L], \end{cases} \quad (7.19)$$

The constant $\delta > 0$ is chosen so that we get the best fit of the double-well potential. Similarly, we can also consider the piecewise constant approximation

$$V_C(x) = \begin{cases} 0 & x \in [\alpha, L], \\ \zeta & x \in [-\alpha, \alpha], \\ 0 & x \in [-L, -\alpha], \\ +\infty & x \notin [-L, L]. \end{cases} \quad (7.20)$$

Again, the constants $\zeta, \alpha > 0$ are chosen to obtain the best fit. These approximations are compared to a symmetric double-well potential in Fig. 7.4.

Using formula (7.16), we can obtain an analytical expression for the mean exit time from the left well for the piecewise linear potential (7.19). We have the following:²

$$\begin{aligned} \tau(x) &= \beta \int_x^0 dy e^{\beta \frac{\delta}{L}(L+y)} \int_{-L}^y dz e^{-\beta \frac{\delta}{L}(L+z)} \\ &= \frac{Lx}{\delta} + \frac{L^2}{\beta \delta^2} \left(e^{\beta \delta} - e^{\frac{\beta \delta}{L}(x+L)} \right), \end{aligned} \quad (7.21)$$

for $x \in [-L, 0]$.

In Fig. 7.5a, we use (7.21) (for $\delta = L = 1$) to plot the mean exit time from the left well for a particle starting at the bottom of the well as a function of the temperature. In Fig. 7.5b, we plot the mean exit time from the left well as a function of the starting point for different values of the temperature. As expected, the mean exit time decreases exponentially fast as the temperature increases. Furthermore, the mean exit time decreases rapidly in a layer close to the local maximum $x = 0$.

² The boundary conditions are reflecting at $x = -L$ and absorbing at $x = 0$, whence (7.16) is the correct formula to use.

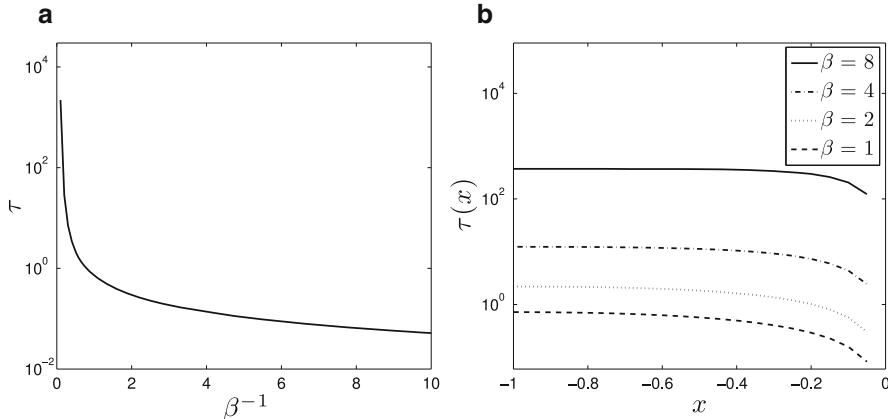


Fig. 7.5 Mean exit time for a particle starting at $x \in [-L, 0]$ for the piecewise linear potential. (a): τ as a function of temperature; (b): τ as a function of x for different values of the temperature β^{-1}

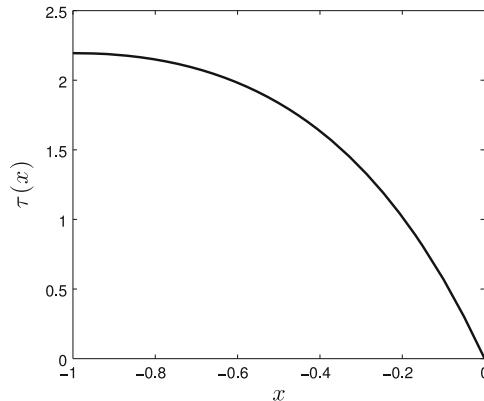


Fig. 7.6 Mean exit time for a particle starting at $x \in [-L, 0]$ for the piecewise linear potential at inverse temperature $\beta = 2$

This phenomenon becomes more pronounced for lower values of the temperature. Similar results can be obtained for the piecewise constant potential; see Exercise 1.

7.4 Stochastic Resonance

In this section, we study the effect of adding a time-periodic forcing to the dynamics (7.1). In particular, we are interested in understanding the effect that the time-dependent forcing has on the hopping rate between the two wells of the potential and on whether it can induce a coherent response of the system in the presence of noise.

We will consider the overdamped dynamics (7.1) with an external time-dependent periodic forcing:

$$dX_t = -V'(X_t) dt + A_0 \cos(\omega_0 t) dt + \sqrt{2\beta^{-1}} dW_t, \quad (7.22)$$

where the potential $V(x)$ is the symmetric bistable potential (7.2). In the absence of the time-dependent forcing, we can associate a characteristic time scale to the dynamics, the Kramers time τ_K defined in (7.17). The time-dependent forcing introduces another time scale in the problem, that of its period $T = \frac{2\pi}{\omega_0}$. The crucial observation is that when these time scales are of the same order of magnitude—see equation (7.46) for the precise relation—then a resonance-like mechanism prevails, and noise can lead to the amplification of the input signal $A_0 \cos(\omega_0 t)$. This phenomenon is referred to as *stochastic resonance*. It is important to note that since the Kramers time τ_K depends on the strength of the noise, the matching between the two time scales can be achieved only when noise is tuned appropriately. Hence the name “stochastic resonance”: this is a noise-induced phenomenon. It should also be pointed out that when the strength of the forcing is large, $A_0 \gg 1$, then it dominates the dynamics in (7.22), and the noise and bistable potential do not play an important role. On the other hand, at high temperatures $\beta^{-1} \gg 1$, the dynamics in (7.22) is dominated by noise, and the effects of the potential and the external forcing are not very strong. Consequently, we expect the dynamics to be interesting and stochastic resonance to occur for weak and moderate strengths of the noise and the periodic forcing.

We can quantify stochastic resonance for (7.22) by studying the corresponding Fokker–Planck equation

$$\frac{\partial p}{\partial t} = \frac{\partial}{\partial x} \left(V'(x)p - A_0 \cos(\omega_0 t)p + \beta^{-1} \frac{\partial p}{\partial x} \right), \quad (7.23)$$

together with the initial condition $p(x, 0) = p_0(x)$. Unlike most of the examples that we have studied so far, the stochastic differential equation (7.22) and the corresponding Fokker–Planck equation (7.23) have time-dependent coefficients. Consequently, the spectral analysis that was used in Chap. 4 to obtain formulas for the transition probability density (4.107), the stationary autocorrelation function (4.110), and the spectral density (4.111) is not directly applicable and has to be modified.

The trick to studying the spectral problem for a Fokker–Planck operator with time-periodic coefficients is, as is usually the case for time-inhomogeneous problems, to enlarge the state space by introducing an auxiliary variable accounting for the time-dependence of the coefficients. We consider then (7.22) as a time-inhomogeneous stochastic differential equation on $[0, T] \times \mathbb{R}$, where T is the period of the external forcing:

$$d\tau_t = dt, \quad (7.24a)$$

$$dX_t = -V'(X_t) dt + A_0 \cos(\omega_0 \tau_t) dt + \sqrt{2\beta^{-1}} dW_t. \quad (7.24b)$$

It is important to note that $\tau_t \in [0, T]$ is a periodic function. The generator of the Markov process $\{\tau_t, X_t\}$ is

$$\mathcal{L} = \frac{\partial}{\partial \tau} - V'(x) \frac{\partial}{\partial x} + A_0 \cos(\omega_0 \tau) \frac{\partial}{\partial x} + \beta^{-1} \frac{\partial^2}{\partial x^2}, \quad (7.25)$$

equipped with periodic boundary conditions in τ . The L^2 -adjoint of \mathcal{L} , i.e., the Fokker–Planck operator, is

$$\mathcal{L}^* = -\frac{\partial}{\partial \tau} + \frac{\partial}{\partial x} \left(V'(x) - A_0 \cos(\omega_0 \tau) + \beta^{-1} \frac{\partial}{\partial x} \right). \quad (7.26)$$

The invariant distribution of the process $\{\tau_t, X_t\}$ is the solution of the stationary Fokker–Planck equation

$$\mathcal{L}^* \rho_A = 0, \quad (7.27)$$

equipped with periodic boundary conditions in τ . We have used the subscript A in ρ_A to emphasize the dependence on the forcing $A_0 \cos(\omega_0 t)$. We can prove that there exists a unique solution to (7.27), i.e., the dynamics (7.24) is ergodic. Notice that the stationary distribution ρ_A is a periodic function of τ : the solution of the Fokker–Planck equation (7.23) becomes asymptotically a periodic function of time. In particular, the asymptotic mean and variance of X_t are periodic functions of time.

In fact, we can show that \mathcal{L} and its adjoint \mathcal{L}^* have discrete spectra and that their eigenfunctions are periodic functions of τ . This follows from general results from *Floquet theory*, the spectral theory of differential operators with periodic coefficients. Since (7.25) is not self-adjoint, in order to obtain a spectral representation for the transition probability density similar to (4.107), we need to consider the eigenvalue problems for both the generator and the Fokker–Planck operators:

$$-\mathcal{L} q_k(x, \tau) = \mu_k q_k(x, \tau), \quad (7.28a)$$

$$-\mathcal{L}^* \hat{q}_k(x, \tau) = \mu_k \hat{q}_k(x, \tau). \quad (7.28b)$$

The eigenfunctions of \mathcal{L} and \mathcal{L}^* satisfy the biorthogonality relation

$$\frac{1}{T} \int_{[0, T] \times \mathbb{R}} q_k(x, \tau) \hat{q}_\ell(x, \tau) dx d\tau = \delta_{k\ell}. \quad (7.29)$$

The *Floquet eigenmodes* are periodic functions of τ . From (7.28b), we conclude that

$$\hat{q}_0(x, \tau) = \rho_A(x, \tau),$$

the stationary distribution. Furthermore, as in the time-independent case, we have $q_0(x, \tau) = 1$.

Similarly to (4.107) and (4.110), we obtain (see Exercise 2)

$$p(x, t | y, s) = \sum_{\ell=0}^{\infty} e^{-\mu_\ell t} q_\ell(y, s) \hat{q}_\ell(x, t) \quad (7.30)$$

and

$$C_A(t) = C_A(s+t, s) = \sum_{\ell=0}^{\infty} e^{-\mu_{\ell}|t|} \alpha_{\ell}(s+t) \beta_{\ell}(s), \quad (7.31)$$

with

$$\alpha_{\ell}(u) = \int_{\mathbb{R}} x \hat{q}_{\ell}(x, u) dx \quad (7.32a)$$

$$\beta_{\ell}(u) = \int_{\mathbb{R}} x q_{\ell}(x, u) \rho_A(x, u) dx. \quad (7.32b)$$

It is clear from these formulas that the coefficients $\{\alpha_{\ell}(u), \beta_{\ell}(u)\}_{\ell=0}^{\infty}$ are periodic functions of u .

The question that we want to address now is whether noise can be tuned in such a way that an initial signal can be amplified. To study this problem, we need to find a measure for quantifying signal amplification. The simplest way is to measure the amplitude $\bar{x}(\beta)$ of the output signal, i.e., of the stationary mean of the solution to (7.22), relative to the amplitude A_0 of the input signal. We define the *spectral amplification*

$$\eta = \left| \frac{\bar{x}(\beta)}{A_0} \right|^2. \quad (7.33)$$

This formula is actually well defined under the assumption that the output signal is also monochromatic, at the same frequency as the input signal. We can justify this assumption in the weak-amplitude regime $A_0 \ll 1$ using perturbation theory. In fact, this calculation provides us with a formula for the spectral amplification η in terms of the spectral characteristics of the unperturbed problem. We will see in Chap. 9, that η can be calculated for $A_0 \ll 1$ using the general formalism of linear response theory; see Example 9.5. Here we will obtain a formula for η valid for $A_0 \ll 1$ by solving perturbatively the stationary Fokker–Planck equation (7.27).

The stationary average of X_t , which is a periodic function of time, is given by the formula

$$\langle X_t \rangle_{eq} = \int_{\mathbb{R}} x \rho_A(x, t) dx, \quad (7.34)$$

where $\rho_A(x, t)$ is the solution of (7.27). We write the Fokker–Planck operator in the form

$$\mathcal{L}^* = \mathcal{L}_0^* + A_0 \mathcal{L}_1^*, \quad (7.35)$$

where

$$\mathcal{L}_0^* = -\frac{\partial}{\partial \tau} + \frac{\partial}{\partial x} \left(V'(x) + \beta^{-1} \frac{\partial}{\partial x} \right)$$

and

$$\mathcal{L}_1^* = -\cos(\omega_0 \tau) \frac{\partial}{\partial x}.$$

We look for a solution to (7.27) in the form of a power series expansion in A_0 :

$$\rho_A = p_0 + A_0 p_1 + A_0^2 p_2 + \dots$$

We substitute this expansion into the stationary Fokker–Planck equation and use (7.35) to obtain the sequence of equations

$$\mathcal{L}_0^* p_0 = 0, \quad \mathcal{L}_0^* p_1 + \mathcal{L}_1^* p_0 = 0, \quad \dots$$

From the first equation, we obtain

$$p_0 = \frac{1}{Z} e^{-\beta V(x)} =: \rho_s(x),$$

the stationary distribution of the unperturbed dynamics. The second equation gives

$$-\mathcal{L}_0^* p_1 = \mathcal{L}_1^* \rho_s = \beta \cos(\omega_0 t) V'(x) \rho_s.$$

Setting now $p_1 = \rho_s f$, we can rewrite this equation as

$$-\left(-\frac{\partial f}{\partial t} + \mathcal{L}_0 f\right) = \beta \cos(\omega_0 t) V'(x), \quad (7.36)$$

where $\mathcal{L}_0 = -V'(x) \frac{\partial}{\partial x} + \beta^{-1} \frac{\partial^2}{\partial x^2}$ denotes the generator of the unperturbed dynamics. To solve (7.36), we use Fourier analysis: since $f = f(x, t)$ is a periodic function of time, we can expand it into a Fourier series:

$$f(x, t) = \sum_{n=-\infty}^{+\infty} f_n(x) e^{i\omega_0 n t}.$$

Substituting this into (7.36), we obtain the equations

$$(i\omega_0 - \mathcal{L}_0) f_1 = \frac{\beta V'(x)}{2}, \quad (7.37a)$$

$$(-i\omega_0 - \mathcal{L}_0) f_{-1} = \frac{\beta V'(x)}{2}, \quad (7.37b)$$

$$f_n = 0, \quad n \neq \pm 1. \quad (7.37c)$$

We can now solve (7.37a) and (7.37b) by expanding $f_{\pm 1}$ into the eigenfunctions of \mathcal{L}_0 , which form an orthonormal basis in $L^2(\mathbb{R}; \rho_s)$; see Sect. 4.7. Let $\{\phi_n, \lambda_n\}_{n=1}^{\infty}$ be the eigenvalues and eigenfunctions of \mathcal{L} . We have

$$V'(x) = \sum_{n=1}^{\infty} V_n \phi_n, \quad V_n = \langle V', \phi_n \rangle_{\rho_s}$$

and

$$f_{\pm 1} = \sum_{n=1}^{\infty} f_n^{\pm 1} \phi_n, \quad f_n^{\pm 1} = \langle f_{\pm 1}, \phi_n \rangle_{\rho_s}.$$

The solution of (7.37a) and (7.37b) is

$$f_n^{\pm 1} = \frac{\beta}{2} \frac{V_n}{\pm i\omega_0 + \lambda_n}. \quad (7.38)$$

We combine (7.38), together with (7.34), the expansion $\rho_A = \rho_s(1 + A_0 f) + \mathcal{O}(A_0^2)$, and the orthonormality of the eigenfunctions of the unperturbed generator \mathcal{L}_0 in $L^2(\mathbb{R}; \rho_s)$ to calculate³

$$\begin{aligned} \langle X_t \rangle_{eq} &= \int_{\mathbb{R}} x \rho_A(x, t) dx = A_0 \int_{\mathbb{R}} x f(x, t) \rho_s(x) dx \\ &= \frac{A_0 \beta}{2} \int_{\mathbb{R}} x \rho_s(x) \sum_{n=1}^{\infty} \left(\frac{V_n \phi_n(x)}{i\omega_0 + \lambda_n} e^{i\omega_0 t} + \frac{V_n \phi_n(x)}{-i\omega_0 + \lambda_n} e^{-i\omega_0 t} \right) dx \\ &= \frac{A_0 \beta}{2} \int_{\mathbb{R}} \sum_{\ell=1}^{\infty} \langle x, \phi_{\ell} \rangle_{\rho_s} \phi_{\ell} \rho_s(x) \sum_{n=1}^{\infty} \left(\frac{V_n \phi_n(x)}{i\omega_0 + \lambda_n} e^{i\omega_0 t} + \frac{V_n \phi_n(x)}{-i\omega_0 + \lambda_n} e^{-i\omega_0 t} \right) dx \\ &= \frac{A_0 \beta}{2} \sum_{\ell=1}^{\infty} \sum_{n=1}^{\infty} \langle x, \phi_{\ell} \rangle_{\rho_s} V_n \left(\frac{1}{i\omega_0 + \lambda_n} e^{i\omega_0 t} + \frac{1}{-i\omega_0 + \lambda_n} e^{-i\omega_0 t} \right) \langle \phi_{\ell}, \phi_n \rangle_{\rho_s} \\ &= \frac{A_0 \beta}{2} \sum_{n=1}^{\infty} g_n \left(\frac{1}{i\omega_0 + \lambda_n} e^{i\omega_0 t} + \frac{1}{-i\omega_0 + \lambda_n} e^{-i\omega_0 t} \right) \\ &= \bar{x} \cos(\omega_0 t - \bar{\phi}). \end{aligned}$$

In writing the above, we have introduced the notation

$$g_n = \langle x, \phi_n \rangle_{\rho_s} V_n = \langle x, \phi_n \rangle_{\rho_s} \langle V', \phi_n \rangle_{\rho_s} \quad (7.39)$$

and

$$\bar{x} = A_0 \beta |\chi(\omega_0)|, \quad \bar{\phi} = \arctan \left| \frac{\chi''(\omega_0)}{\chi'(\omega_0)} \right|, \quad (7.40)$$

where $\chi(\omega)$ denotes the *susceptibility*

$$\chi(\omega_0) = \chi'(\omega_0) - i\chi''(\omega_0) = \sum_{n=1}^{\infty} \frac{g_n}{i\omega_0 + \lambda_n}. \quad (7.41)$$

Now the spectral amplification becomes

$$\eta = \beta^2 |\chi(\omega_0)|^2. \quad (7.42)$$

As expected, to leading order in A_0 , the spectral amplification is independent of the amplitude of the input signal. Observe as well that in order to compute η , complete information on the spectrum of the unperturbed dynamics is needed. Since the eigenvalues and eigenfunctions of the generator of the unperturbed

³ We assume, without loss of generality, that in the absence of the time-dependent forcing $A_0 = 0$, we have that X_t is of mean zero, $\int x \rho_s(x) dx = 0$. This assumption is satisfied for the symmetric bistable potential.

dynamics \mathcal{L}_0 with the potential V given by (7.2) are not known explicitly, in order to calculate η using (7.42), we need to resort to numerical simulations or to study particular asymptotic regimes using perturbation theory. We can do this in the weak-temperature regime.

To compute η for $\beta \gg 1$, we need some information on the spectrum of the generator \mathcal{L}_0 :

1. The first nonvanishing eigenvalue is given approximately by

$$\lambda_1 \approx 2\kappa = \frac{\omega_0 \omega_b}{\pi} e^{-\beta \Delta V} = \frac{\sqrt{2}}{\pi} e^{-\beta/4}. \quad (7.43)$$

2. The second eigenfunction can be well approximated by the sign function:

$$\phi_1(x) \approx \text{sign}(x). \quad (7.44)$$

3. The eigenvalues λ_ℓ for $\ell \geq 2$ are bounded away from 0 independently of β .

From the above, it follows that in the low-temperature regime, it is sufficient to consider only the first term in (7.41) and (7.42):

$$\eta \approx \beta^2 \frac{g_1^2}{\omega_0^2 + \lambda_1^2},$$

where g_1 is defined in Eq. (7.39):

$$g_1 = \langle x, \phi_1 \rangle_{\rho_s} \langle V'(x), \phi_1 \rangle_{\rho_s}.$$

We can compute g_1 approximately for $\beta \gg 1$. For the symmetric double-well potential (7.2), we have

$$g_1 \approx \frac{\sqrt{2}}{\pi} e^{-\beta/4}.$$

We combine this with (7.43) and the formula for η to obtain

$$\eta \approx \beta^2 \frac{1}{\frac{\pi^2}{2} \omega_0^2 e^{\beta/2} + 1}. \quad (7.45)$$

In Fig. 7.7, we plot the spectral amplification as a function of the temperature for different values of the signal frequency ω_0 .⁴ The resonance effect becomes more pronounced for small values of the frequency ω_0 . The critical temperature can be calculated by solving the equation $\frac{d}{d\beta} \eta(\beta) = 0$. This leads to the equation

$$\left(\frac{\beta}{4} - 1 \right) \frac{\pi^2}{2} \omega_0^2 e^{\beta/2} = 1.$$

⁴ Note that although in this figure, we have $\beta^{-1} \in [0, 1]$, formula (7.45) is actually valid only for $\beta \gg 1$.

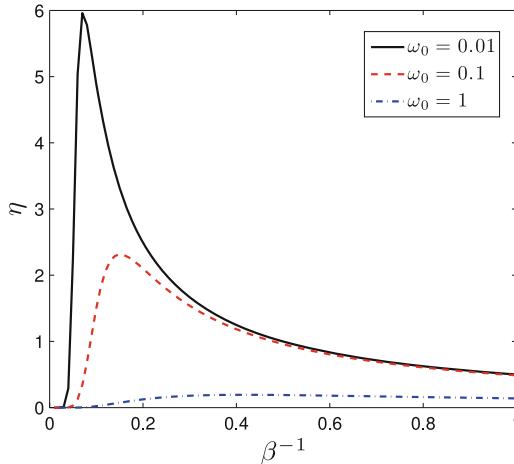


Fig. 7.7 Spectral amplification as a function of the temperature β^{-1} in the linear-response low-temperature regime $A_0 \ll 1, \beta \gg 1$

From this transcendental equation, we obtain the scaling between the period $T_0 = \frac{2\pi}{\omega_0}$ of the input signal and the Kramers time τ (7.17):

$$T_0 = \pi \sqrt{\beta} \tau. \quad (7.46)$$

7.5 Brownian Motors

Low-dimensional stochastic differential equations with time-dependent coefficients, similar to (7.22), can be used to model thermally induced directed transport in the absence of spatiotemporal symmetries. Reversibility, or equivalently, detailed balance (see Sect. 4.6), implies temporal symmetry of the dynamics. Consider, for example, the overdamped Langevin reversible dynamics

$$dX_t = -V'(X_t) dt + \sqrt{2\beta^{-1}} dW_t,$$

with a periodic potential $V(x)$ with period L in all directions. We consider the dynamics on $[0, L]^d$ with periodic boundary conditions. Then, the average position of X_t at equilibrium is

$$\mathbb{E}_{eq} X_t = \frac{1}{Z} \int_{[0,L]^d} (-\nabla V(x)) e^{-\beta V(x)} dx = 0.$$

Consequently, the effective drift, see (7.48), is

$$U = 0,$$

and there is no preferred direction in which the dynamics move.

On the other hand, while a Brownian particle in a periodic potential is kept away from equilibrium by an external, deterministic, or random force, detailed balance does not hold. Consequently, and in the absence of any spatial symmetry, a net particle current will appear, without any violation of the second law of thermodynamics.

Before we develop a general formalism for studying noise-induced directed transport, we consider a simple model for which we can calculate the effective drift explicitly. We consider a Brownian particle moving in a one-dimensional periodic potential with period L , subject to time-dependent external forcing with period T ; compare with (7.22):

$$dX_t = -V(X_t) dt + F(t) dt + \sqrt{2\beta^{-1}} dW_t. \quad (7.47)$$

Our goal is to calculate the effective drift, which is given by (see equation (7.66) below as well as Exercise 7)

$$U = \int_0^T \int_0^L \rho(x, \tau) dx d\tau, \quad (7.48)$$

where ρ is the solution of the stationary Fokker–Planck equation

$$\mathcal{L}^* \rho = \left[-\frac{\partial}{\partial \tau} + \frac{\partial}{\partial x} \left(V'(x) - F(\tau) + \beta^{-1} \frac{\partial}{\partial x} \right) \right] \rho = 0. \quad (7.49)$$

This equation is posed on $[0, L] \times [0, T]$ with periodic boundary conditions in both x and τ . For arbitrary potentials and forcings, it is not possible to solve this equation and to calculate the effective drift. We will make the following approximation: we will assume that the forcing $F(\cdot)$ is slowly varying, so that we can perform the calculation of U in two steps. First, we will use the formula for the drift of a Brownian particle in a tilted periodic potential (6.169), treating the forcing F as a constant. Then, we will average the resulting drift with respect to time. This procedure can be justified using perturbation theory: we can study (7.49) perturbatively for a slowly varying $F(\cdot)$ by introducing an appropriate small parameter. We can refer to this approximation as the *adiabatic approximation*.

Even in the adiabatic approximation, it is impossible in general to calculate the normalization constant (partition function) of the stationary distribution for a tilted periodic potential, equation (6.167), since this requires the calculation of a double integral. To obtain explicit formulas, we will consider a piecewise linear potential, similar to (6.146), which we used to calculate the diffusion coefficient of the Brownian particle in the absence of an external forcing. Now we need to consider an asymmetric potential: as we have already discussed, for the appearance of an effective drift (i.e., preferred direction of motion), it is necessary to break all spatiotemporal symmetries. We consider

$$V(x) = \begin{cases} \frac{x}{\lambda_1} & : x \in [0, \lambda_1], \\ -\frac{x}{\lambda_2} + \frac{L}{\lambda_2} & : x \in (\lambda_1, L]. \end{cases} \quad (7.50)$$

To simplify the formulas below, we have set the maximum of the potential to be 1, and we will also set $L = 1$. We have then that $\lambda_1 + \lambda_2 = 1$ and $\Delta = \lambda_1 - \lambda_2 = 2\lambda_1 - 1$. Evaluating now the partition function in (6.167)⁵

$$U = \frac{2\beta B^2 \sinh\left(\frac{F\beta}{2}\right)}{\left[\cosh\left(\beta\left(1 - \frac{F\Delta}{2}\right)\right) - \beta \cosh\left(\frac{F\beta}{2}\right)\right] - AB \sinh\left(\frac{F\beta}{2}\right)}, \quad (7.51)$$

where

$$A = \Delta + \frac{F(1 - \Delta^2)}{4}, \quad B = \left[1 - \frac{\Delta F}{2}\right]^2 - \left[\frac{F}{2}\right]^2.$$

For the slowly varying force $F(t)$, the drift (7.51) becomes a periodic function of time. To obtain the effective drift, we need to average $U(F(t))$ over a period. We will consider a piecewise constant periodic force (i.e., a *square wave*)

$$F(t) = \begin{cases} F_0 & : t \in [0, T/2], \\ -F_0 & : t \in (T/2, T]. \end{cases} \quad (7.52)$$

Averaging over a period T , we obtain

$$\bar{U} = \frac{1}{2} \left[U(F_0) + U(-F_0) \right]. \quad (7.53)$$

Now we plot the drift as a function of the temperature, for fixed strength of the amplitude F_0 as well as a function of the amplitude, at a fixed temperature (see Fig. 7.8). We take $\Delta = \frac{1}{3}$. In both figures, the “motor” effect becomes clear: by forcing the Brownian particle by a time-dependent forcing, and in the absence of parity symmetry of the potential, the particle can move in a noise-induced coherent way. We will use the term *Brownian motor* for this phenomenon.

Derivation of Formulas for the Drift and Diffusion Coefficients

It is possible to obtain general formulas for the drift and diffusion coefficients of a Brownian motor. As discussed in Sect. 6.6, the long-time behavior of a Brownian particle in a periodic potential is determined uniquely by the effective drift and the diffusion coefficient. Let X_t denote the position of a particle moving in a periodic potential under the influence of external noise and thermal fluctuations in one

⁵ Equivalently, we can solve the stationary Fokker–Planck equation (6.165) as follows: We can consider the equation separately in the intervals $[0, \lambda_1]$ and $[\lambda_1, 1]$. In these two intervals, the equation becomes a second-order differential equation with constant coefficients that we can solve. Using, then, the periodic boundary conditions, the normalization condition, and a continuity condition at $x = \lambda_1$, we can calculate the invariant distribution and then substitute it into (6.169) to calculate the effective drift.

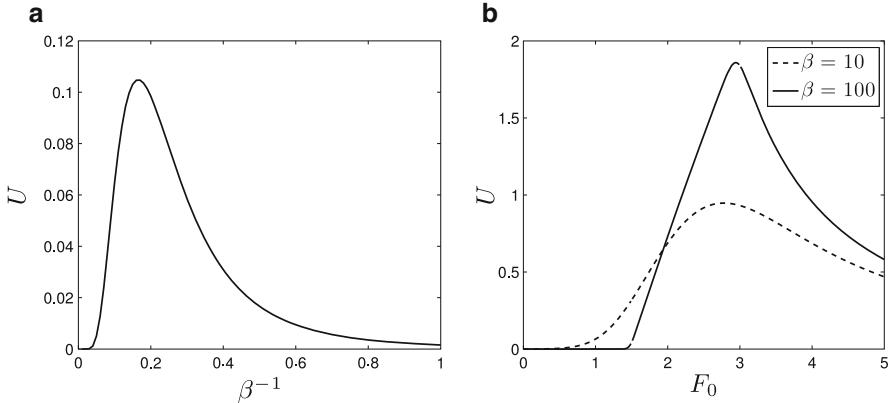


Fig. 7.8 (a) Effective drift as a function of the temperature for the piecewise linear potential and piecewise constant periodic force. (b) Effective drift as a function of the amplitude for a piecewise linear potential and piecewise constant periodic force, for two different values of the temperature

dimension. We define the effective drift and diffusion coefficients in terms of the long-time averages of the mean and variance of X_t :

$$U = \lim_{t \rightarrow \infty} \frac{\mathbb{E} X_t}{t} \quad (7.54)$$

and

$$D = \lim_{t \rightarrow \infty} \frac{\text{Var}(X_t)}{2t}, \quad (7.55)$$

where \mathbb{E} denotes the expectation with respect to the noise (Brownian motion, external random forcing, ...) that is driving the dynamics. Similar formulas can be written in arbitrary dimensions.

We can use now the multiscale techniques that we have already used in Sect. 6.6 to obtain formulas for the effective drift and diffusion coefficients for a quite general class of models for Brownian motors. We consider the overdamped 1-dimensional stochastic dynamics for a state variable $X_t \in \mathbb{R}$:⁶

$$dX_t = (-V'(X_t, f(t)) + y(t)) dt + \sqrt{2\beta^{-1}} dW_t. \quad (7.56)$$

We take $f(t)$ and $y(t)$ to be Markov processes with respective state spaces E_f, E_y and generators $\mathcal{L}_f, \mathcal{L}_y$. The potential $V(x, f)$ is periodic in x for every f , with period L :

$$V(x + L, f) = V(x, f).$$

Here $f(t)$ and $y(t)$ can be taken to be diffusion processes that are constructed as solutions of stochastic differential equations or discrete- or continuous-time Markov

⁶ The stochastic process X_t does not necessarily denote the position of a Brownian particle. We will, however, refer to X_t as the particle position in the sequel.

chains. We can also consider deterministic, periodic time-dependent functions $f(t)$ and $y(t)$. We can think of $y(t)$ as an external forcing and of $f(t)$ as a modulation of the potential. For example, we can take $y(t) = A_0 \cos(\omega_0 t)$, as we did in Sect. 7.4, or we can consider an Ornstein–Uhlenbeck process; see Example 7.2. Furthermore, we can consider a time-dependent, periodic-in-space potential of the form

$$f(t)V(x),$$

where $f(t)$ is a piecewise constant function with period T :

$$f(t) = \begin{cases} 1 & : t \in [0, T/2], \\ 0 & : t \in (T/2, T]. \end{cases}$$

In this way, we can switch the potential on and off. In any event, we can hope for noise-induced directed transport only when $f(t)$, $y(t)$, and the potential V are chosen in such a way that all spatiotemporal symmetries in (7.56) are broken.

The process $\{X_t, f(t), y(t)\}$ in the extended phase space $\mathbb{R} \times E_f \times E_y$ is a Markov process with generator

$$\mathcal{L} = F(x, f, y) \frac{\partial}{\partial x} + \beta^{-1} \frac{\partial^2}{\partial x^2} + \mathcal{L}_f + \mathcal{L}_y,$$

where

$$F(x, f, y) = (-V'(x, f) + y).$$

Similarly to the analysis presented in Sect. 6.6, we will obtain formulas for the drift and diffusion coefficients by studying the corresponding backward Kolmogorov equation using singular perturbation theory. We are interested in the long-time, large-scale behavior of X_t . For this, we will rescale space and time appropriately in order to study the problem at the right length and time scales.

We begin with the calculation of the effective drift. The backward Kolmogorov equation for the expectation value of an observable⁷

$$u(x, y, f, t) = \mathbb{E}(\phi(X_t^{x, y, f}) | X_0^{x, y, f} = x, y(0) = y, f(0) = f)$$

reads

$$\frac{\partial u}{\partial t} = \left(F(x, f, y) \frac{\partial}{\partial x} + \beta^{-1} \frac{\partial^2}{\partial x^2} + \mathcal{L}_f + \mathcal{L}_y \right) u. \quad (7.57)$$

We rescale space and time in (7.57) according to⁸

$$x \rightarrow \varepsilon x, \quad t \rightarrow \varepsilon t \quad (7.58)$$

⁷ We use the notation $X_t^{x, y, f}$ for the solution of (7.56) to emphasize its dependence on the initial conditions for $X_t, f(t), y(t)$. Also, in order to avoid problems with initial layers, we are assuming that ϕ depends explicitly only on X_t ; the dependence on $y(t), f(t)$ comes through (7.56).

⁸ Compare this with the rescaling 6.162. In contrast to Sect. 6.6.2, here we will calculate the drift and diffusion coefficients in two steps.

and divide through by ε to obtain

$$\frac{\partial u^\varepsilon}{\partial t} = \frac{1}{\varepsilon} \left(\varepsilon F \left(\left(\frac{x}{\varepsilon}, f, y \right) \right) \partial_x + \varepsilon^2 \beta^{-1} \partial_x^2 + \mathcal{L}_f + \mathcal{L}_y \right) u^\varepsilon. \quad (7.59)$$

We solve (7.59) perturbatively by looking for a solution in the form of a two-scale expansion⁹

$$u^\varepsilon(x, f, y, t) = u_0 \left(x, \frac{x}{\varepsilon}, f, y, t \right) + \varepsilon u_1 \left(x, \frac{x}{\varepsilon}, f, y, t \right) + \varepsilon^2 u_2 \left(x, \frac{x}{\varepsilon}, f, y, t \right) + \dots \quad (7.60)$$

All terms in the expansion (7.60) are periodic functions of $z = x/\varepsilon$. From the chain rule, we have

$$\frac{\partial}{\partial x} \rightarrow \frac{\partial}{\partial x} + \frac{1}{\varepsilon} \frac{\partial}{\partial z}. \quad (7.61)$$

We substitute now (7.60) into (7.57), use (7.61), and treat x and z as independent variables. On equating the coefficients of equal powers in ε , we obtain the following sequence of equations:

$$\mathcal{L}_0 u_0 = 0, \quad (7.62a)$$

$$\begin{aligned} \mathcal{L}_0 u_1 &= -\mathcal{L}_1 u_0 + \frac{\partial u_0}{\partial t}, \\ \dots &= \dots, \end{aligned} \quad (7.62b)$$

where

$$\mathcal{L}_0 = F(z, f, y) \frac{\partial}{\partial z} + \beta^{-1} \frac{\partial^2}{\partial z^2} + \mathcal{L}_y + \mathcal{L}_f \quad (7.63a)$$

$$\mathcal{L}_1 = F(z, f, y) \frac{\partial}{\partial x} + 2\beta^{-1} \frac{\partial^2}{\partial x \partial z}. \quad (7.63b)$$

The operator \mathcal{L}_0 is the generator of a Markov process in $\mathcal{M} := [0, L] \times E_y \times E_f$. In order to proceed, we need to assume that this process is ergodic: there exists a unique stationary solution of the Fokker–Planck equation

$$\mathcal{L}_0^* \rho(z, y, f) = 0, \quad (7.64)$$

with

$$\int_{\mathcal{M}} \rho(z, y, f) dz dy df = 1$$

and

$$\mathcal{L}_0^* \rho = \frac{\partial}{\partial z} (F(z, f, y) \rho) + \beta^{-1} \frac{\partial^2}{\partial z^2} \rho + \mathcal{L}_y^* \rho + \mathcal{L}_f^* \rho.$$

⁹ Alternatively, we could introduce the auxiliary variable $z = \frac{x}{\varepsilon}$, write the SDE (7.56) in terms of x , y , z , and f , and then proceed with an appropriate power series expansion. This is what we did in Sect. 6.6.

In the above, \mathcal{L}_f^* and \mathcal{L}_y^* are the Fokker–Planck operators of f and y , respectively. The stationary density $\rho(z, y, f)$ satisfies periodic boundary conditions in z and appropriate boundary conditions in f and y .¹⁰

Under the assumption that (7.64) has a unique solution, equation (7.62a) implies, by the Fredholm alternative, that u_0 is independent of the fast scales:

$$u_0 = u(x, t).$$

Equation (7.62b) now becomes

$$\mathcal{L}_0 u_1 = \frac{\partial u(x, t)}{\partial t} - F(z, y, f) \partial_x u(x, t).$$

In order for this equation to be well posed, it is necessary that the right-hand side average to 0 with respect to the invariant distribution $\rho(z, f, y)$. This leads to the backward Liouville equation

$$\frac{\partial u(x, t)}{\partial t} = U \frac{\partial u(x, t)}{\partial x}, \quad (7.65)$$

together with the initial condition $u(x, 0) = \phi(x)$. The effective drift is given by

$$\begin{aligned} U &= \int_{\mathcal{M}} F(z, y, f) \rho(z, y, f) dz dy df \\ &= \int_{\mathcal{M}} (-V'(x, f) + y) \rho(z, y, f) dz dy df. \end{aligned} \quad (7.66)$$

Thus, to leading order, the Brownian particle moves ballistically with speed U . Of course, the effective drift U can be zero; this will be the case when detailed balance is satisfied or when symmetries in the problem preclude the existence of a preferred direction of motion. Note also that the limiting transport equation (7.65), which is a first-order partial differential equation in both space and time, justifies a posteriori the “hyperbolic” rescaling (7.58).

Now we proceed with the calculation of the diffusion coefficient D . The diffusion coefficient provides information on the strength of fluctuations around the mean deterministic motion. We need to look at the rescaled process given by (6.162), $X_t^\varepsilon = \varepsilon X_t/\varepsilon^2 - Ut/\varepsilon$. To simplify the calculation, we will assume for the moment that the effective drift vanishes, $U = 0$. Then the rescaling for the backward Kolmogorov equation becomes

$$x \rightarrow \varepsilon x, \quad t \rightarrow \varepsilon^2 t,$$

¹⁰ In order to be more specific on the boundary conditions with respect to f and y , we need to specify their generator. For example, when $y(t)$ is the Ornstein–Uhlenbeck process, the boundary conditions in y are that ρ should decay sufficiently fast as $|y| \rightarrow +\infty$.

and after dividing through by ε^2 , (7.57) becomes

$$\begin{aligned}\frac{\partial u^\varepsilon}{\partial t} &= \frac{1}{\varepsilon^2} \left(\varepsilon F \left(\frac{x}{\varepsilon}, f, y \right) \frac{\partial}{\partial x} + \varepsilon^2 \beta^{-1} \frac{\partial^2}{\partial x^2} + \mathcal{L}_f + \mathcal{L}_y \right) u^\varepsilon \\ &=: \left(\frac{1}{\varepsilon^2} \mathcal{L}_0 + \frac{1}{\varepsilon} \mathcal{L}_1 + \mathcal{L}_2 \right) u^\varepsilon,\end{aligned}\quad (7.67)$$

where \mathcal{L}_0 and \mathcal{L}_1 are defined in (7.63), and

$$\mathcal{L}_2 = \beta^{-1} \frac{\partial^2}{\partial x^2}.$$

We go through the same analysis as in the calculation of the effective drift to obtain the following sequence of equations:

$$\mathcal{L}_0 u_0 = 0, \quad (7.68a)$$

$$\mathcal{L}_0 u_1 = -\mathcal{L}_1 u_0, \quad (7.68b)$$

$$\mathcal{L}_0 u_2 = -\mathcal{L}_1 u_1 - \mathcal{L}_2 u_0, \quad (7.68c)$$

$$\dots = \dots.$$

Equation (7.68a) implies that $u_0 = u(x, t)$. Now (7.68b) becomes

$$\mathcal{L}_0 u_1 = -F(z, y, f) \frac{\partial}{\partial x} u(x, t).$$

Since we have assumed that $U = 0$, the right-hand side of the above equation belongs to the null space of \mathcal{L}_0^* , and this equation is well posed. The solution is

$$u_1(x, z, f, y, t) = \chi(z, y, f) \frac{\partial}{\partial x} u(x, t), \quad (7.69)$$

where the auxiliary field $\chi(z, y, f)$ satisfies the Poisson equation

$$-\mathcal{L}_0 \chi(z, y, f) = F(z, y, f), \quad (7.70)$$

with periodic boundary conditions in z and appropriate boundary conditions in y and f ; see footnote 10.

We proceed now with the analysis of equation (7.68c). The solvability condition for this equation reads (with $dZ = dx dy dz$)

$$\int_{\mathcal{M}} (-\mathcal{L}_1 u_1 - \mathcal{L}_2 u_0) \rho(z, y, f) dZ = 0,$$

from which, on using (7.69), (7.70), and the fact that u_0 depends only on x and t , we obtain the limiting backward Kolmogorov equation for $u(x, t)$:

$$\begin{aligned}\frac{\partial u}{\partial t} &= \beta^{-1} \frac{\partial^2 u}{\partial x^2} + \int_{\mathcal{M}} \left(F(z, f, y) \frac{\partial}{\partial x} + 2\beta^{-1} \frac{\partial^2}{\partial x \partial z} \right) \left(\chi(z, y, f) \frac{\partial}{\partial x} u(x, t) \right) \rho(z, y, f) dZ \\ &= D \frac{\partial^2 u}{\partial x^2}.\end{aligned}$$

The diffusion coefficient is given by

$$D = \beta^{-1} + \int_{\mathcal{M}} \left(F(z, y, f) \chi(z, y, f) + 2\beta^{-1} \frac{\partial \chi(z, y, f)}{\partial z} \right) \rho(z, y, f) dZ.$$

When the effective drift does not vanish, $U \neq 0$ can be studied using the same techniques by studying the rescaled process $X^\varepsilon(t) := \varepsilon \left(X_{t/\varepsilon^2} - \varepsilon^{-2} U t \right)$. The diffusion coefficient is given by (7.71), with χ the solution of (7.70) but with $F(z, y, f)$ replaced by $F(z, y, f) - U$:

$$D = \beta^{-1} + \int_{\mathcal{M}} \left((F(z, y, f) - U) \chi(z, y, f) + 2\beta^{-1} \frac{\partial \chi(z, y, f)}{\partial z} \right) \rho(z, y, f), \quad (7.71)$$

and

$$-\mathcal{L}_0 \chi = F(z, y, f) - U. \quad (7.72)$$

Example 7.2. We consider a one-dimensional reversible diffusion perturbed by a stationary Ornstein–Uhlenbeck process:

$$dX_t = (-V'(X_t) + y_t) dt + \sqrt{2\beta^{-1}} dW_1(t), \quad (7.73a)$$

$$dy_t = -\frac{1}{\tau} y_t dt + \sqrt{\frac{2\sigma}{\tau}} dW_2(t), \quad (7.73b)$$

where $W_i(t)$, $i = 1, 2$, are independent standard 1-dimensional Brownian motions. The potential $V(x)$ is assumed to be periodic with period L . We can use (7.66) and (7.71) to calculate the drift and diffusion coefficients. The stationary Fokker–Planck equation and the Poisson equation (7.71) become

$$-\partial_z ((-\partial_z V(z) + y) \rho(y, z)) + \beta^{-1} \partial_z^2 \rho(y, z) + \frac{1}{\tau} (\partial_y (y \rho(y, z)) + \sigma \partial_y^2 \rho(y, z)) = 0$$

and

$$\begin{aligned}& -(-\partial_z V(z) + y) \partial_z \chi(y, z) - \beta^{-1} \partial_z^2 \chi(y, z) - \frac{1}{\tau} (-y \partial_y \chi(y, z) + \sigma \partial_y^2 \chi(y, z)) \\ &= (-\partial_z V(z) + y) - U.\end{aligned}$$

These equations cannot be solved in closed form. However, we can solve them perturbatively for short correlation time τ of the Ornstein–Uhlenbeck process to obtain approximate analytical expressions for the effective drift and diffusion coefficients. We obtain, see Exercise 7,

$$U = \mathcal{O}(\tau^3) \quad (7.74)$$

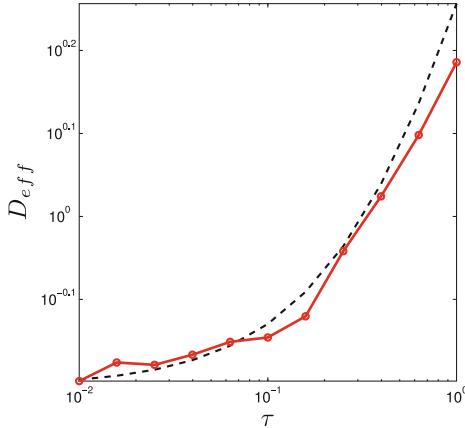


Fig. 7.9 Effective diffusion coefficient for (7.73) with $V(x) = \cos(x)$ as a function of τ , for $\sigma = 1$, $\beta = 1$. *Solid line*: Results from Monte Carlo simulations. *Dashed line*: Results using (7.75)

and

$$D = \frac{L^2}{Z\widehat{Z}} \left(D + \tau\sigma \left(1 + \frac{1}{\gamma D^2} \left(\frac{Z_2}{\widehat{Z}} - \frac{Z_1}{Z} \right) \right) \right) + \mathcal{O}(\tau^2), \quad (7.75)$$

with

$$\begin{aligned} Z &= \int_0^L e^{-\beta V(z)} dz, & \widehat{Z} &= \int_0^L e^{\beta V(z)} dz, \\ Z_1 &= \int_0^L V(z) e^{-\beta V(z)} dz, & Z_2 &= \int_0^L V(z) e^{\beta V(z)} dz. \end{aligned}$$

In the limit as the correlation time of y_t tends to 0, i.e., when y_t is close to a white noise process, X_t is close to a reversible process. Hence, it is not surprising that the effective drift is very weak. Furthermore, as expected, in the limit as $\tau \rightarrow 0$, the diffusion coefficient converges to that for a Brownian particle in a periodic potential; see (6.145).

We can compare the small- τ asymptotics for the diffusion coefficient with Monte Carlo simulations, based on the numerical solution of equations (7.73) using the Euler–Maruyama method (see Sect. 5.2), for the cosine potential $V(x) = \cos(x)$. The results are presented in Figs. 7.9 and 7.10.

7.6 Discussion and Bibliography

Kramers's paper on the modeling of chemical reactions based on diffusion processes and on the calculation of reaction rates is [122]. Our calculations are based on the study of the mean first passage time. See [68, Chaps. 5, 9], [246, Chap. 4], and [82, Sec. VII]. Further information on the calculation of reaction rates using Kramers's approach can be found in [82]. See also [80] and [162].

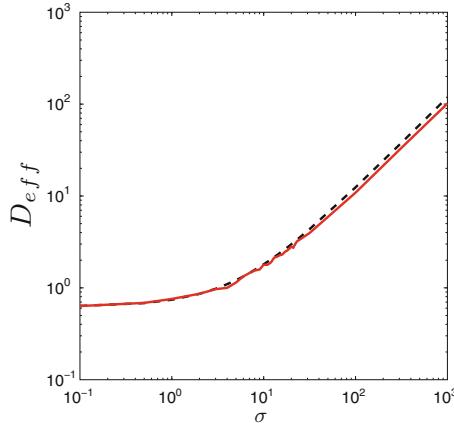


Fig. 7.10 Effective diffusion coefficient for (7.73) with $V(x) = \cos(x)$ as a function of σ , for $\tau = 0.1$, $\beta = 1$. *Solid line*: Results from Monte Carlo simulations. *Dashed line*: Results using (7.75)

As we have seen in Sect. 6.5, the overdamped Langevin dynamics

$$\gamma \dot{q} = -V'(q) + \sqrt{2\gamma\beta^{-1}} \dot{W}, \quad (7.76)$$

is obtained from the Langevin dynamics

$$\ddot{q} = -\partial_q V(q) - \gamma \dot{q} + \sqrt{2\gamma\beta^{-1}} \dot{W}, \quad (7.77)$$

in the limit of large friction. In Sect. 7.3, we obtained an approximate analytical formula for the escape rate in the overdamped regime, valid at low temperatures:

$$\kappa = \frac{\omega_a \omega_b}{2\pi\gamma} \exp(-\beta \Delta V), \quad (7.78)$$

where ΔV denotes the height of the potential barrier. A similar formula is valid when $\beta \gg 1$ for arbitrary values of the friction coefficient:

$$\kappa = v(\gamma) \exp(-\beta \Delta V),$$

where the rate coefficient v depends on the friction coefficient and the temperature. In order to calculate the reaction rate in the underdamped or *energy-diffusion-limited* regime $\gamma \ll 1$, we need to study the diffusion process for the energy; see equations (6.120) and (6.122). The result is (compare with formula (6.156) for the small- γ , large- β asymptotics of the diffusion coefficient of a Brownian particle in a periodic potential)

$$\kappa = \gamma \beta I(\Delta V) \frac{\omega_a}{2\pi} e^{-\beta \Delta V}, \quad (7.79)$$

where $I(\Delta V)$ denotes the action evaluated at the local maximum b . See [82, Sect. IV]. This result was obtained by Kramers in [122].

For intermediate values of the friction, Kramers's formula becomes

$$\kappa = \frac{\sqrt{\frac{\gamma^2}{4} - \omega_b^2 - \frac{\gamma}{2}} \omega_a}{\omega_b} \frac{1}{2\pi} \exp(-\beta \Delta V). \quad (7.80)$$

In the overdamped limit $\gamma \rightarrow +\infty$, (7.80) reduces to (7.78). A formula for the escape rate that is valid for all values of friction coefficient was obtained in [163]. This formula reduces to (7.78) and (7.79) in the overdamped and underdamped limits, respectively.

It is also possible to calculate reaction rates for non-Markovian Langevin equations of the form

$$\ddot{q} = -V'(q) - \int_0^t \gamma(t-s) \dot{q}(s) ds + \xi(t) \quad (7.81a)$$

$$\langle \xi(t) \xi(0) \rangle = \beta^{-1} \gamma(t). \quad (7.81b)$$

We will derive the *generalized Langevin equation* (7.81a), together with the fluctuation–dissipation theorem (7.81b), in Chap. 8. The calculation of reaction rates for the generalized Langevin equation can be found in [80]. A systematic study of the problem of the escape from a potential well, both for Markovian and non-Markovian dynamics (in the Markovian approximation, see Sect. 8.2) based on a careful analysis of the boundary value problem for the mean first passage time using singular perturbation theory, can be found in [157, 158, 216]. In particular, the calculation of the reaction rate that is uniformly valid in the friction coefficient is presented in [158].

Stochastic resonance was first introduced in the context of climate modeling, as a possible mechanism for explaining the phenomenon of periodically recurrent ice ages [18, 19, 173, 175]. In these papers, the “climate” is modeled through a double-well potential, where the minima correspond to the low-temperature state (the “ice age”). The periodic forcing models the modulation of the earth’s eccentricity, whereas all other sources of short-term climate fluctuations are modeled as white noise. A more recent review of the application of stochastic resonance to climate modeling can be found in [174].

Although it is not clear whether stochastic resonance is the correct mechanism for explaining the recurrent ice ages, the concept of stochastic resonance and more generally of noise-induced coherence and signal amplification has found numerous applications in solid-state physics and biology. Further information can be found in the review article [66]. A rigorous study of stochastic resonance can be found in [20, 87]; see also [21, Chap. 4] for a review of the analysis of stochastic resonance based on the Fokker–Planck equation.

The introduction of the auxiliary variable τ_t in (7.24) introduces the problem of choosing an initial condition for this variable. Equivalently, we need to choose an initial distribution (as a function of two variables τ, x) for the Fokker–Planck equation. This issue is discussed in [110]. By introducing this auxiliary variable and studying the ergodic properties of the two-dimensional stochastic differential equation (7.24), we can use the standard definition for the stationary autocorrelation

function (4.109) and for the spectral density of a stationary process (1.7). Alternatively, we could take the starting time to the distant past, so that the system at time $t = 0$ has already forgotten its initial condition [21, Chap. 4] (using the notation \mathbb{E}^{t_0, x_0} to denote the conditional expectation given the initial condition $X_{t_0} = x_0$):

$$\begin{aligned} C_A(t, s) &= \lim_{t_0 \rightarrow -\infty} \mathbb{E}^{t_0, x_0}(X_t X_s) \\ &= \lim_{t_0 \rightarrow -\infty} \int \int x p(x, t|y, s) y p(y, s|x_0, t_0) dy dx \end{aligned}$$

and

$$S_A(\omega) = \lim_{t \rightarrow +\infty} \frac{1}{2t} \left| \int_{-t}^t \lim_{t_0 \rightarrow -\infty} \mathbb{E}^{t_0, x_0}(X_s) e^{i\omega s} ds \right|^2,$$

where X_t denotes the solution of (7.22), and \mathbb{E}^{t_0, x_0} denotes the expectation with the process starting at x_0 at t_0 . The two approaches lead to equivalent results.

The weak forcing regime $A_0 \ll 1$ in (7.22) that we studied using perturbation theory can be cast in the framework of linear response theory. See Chap. 9 for details.

In Sect. 7.4, we used the spectral amplification η to quantify the phenomenon of stochastic resonance. An alternative measure for spectral amplification is provided by the *signal-to-noise ratio* (SNR). The definition of the SNR is based on the observation that the power spectrum (spectral density) of X_t , the solution of (7.22),

$$S_{A_0}(\omega) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{it\omega} C(t) dt, \quad (7.82)$$

can be written as the superposition of a continuous part, e.g., as a sum of Lorentzians, corresponding to the noisy part, and a sequence of delta-like spikes at odd-integer multiples of the input frequency ω_0 , due to the periodic input signal. In the weak external forcing (linear response) regime $A_0 \ll 1$, most of the energy of the delta spikes is concentrated at the first two spikes, and we can write

$$S_{A_0}(\omega) = \frac{\pi}{2} \bar{x}^2(\beta) \left(\delta(\omega - \omega_0) + \delta(\omega + \omega_0) \right) + S_N(\omega),$$

where $S_N(\omega) = S_N^0(\omega)$, the spectral density of the unperturbed problem, up to $\mathcal{O}(A_0^2)$ (the calculation is very similar to the one that we presented in this chapter). The SNR is then defined as¹¹

$$\text{SNR} = \frac{\pi}{2} \frac{\bar{x}^2(\beta)}{S_N^0(\omega_0)}. \quad (7.83)$$

¹¹ More generally, the SNR is essentially the ratio between the strengths of the singular and continuous parts of the power spectrum:

$$\text{SNR} = 2 \frac{\lim_{\Delta\omega \rightarrow 0} \int_{\omega_0 - \Delta\omega}^{\omega_0 + \Delta\omega} S(\omega) d\omega}{S_N(\omega)}.$$

The SNR can be computed approximately in the weak-amplitude, low-temperature regime. See Exercise 4.

Background material on Floquet theory can be found in [128, 194]. Results on the ergodic theory of SDEs with time-dependent periodic coefficients can be found in [17, Chap. 3]; see also [37]. PDE-based arguments for studying the ergodic properties and developing homogenization theory for SDEs with space- and time-periodic coefficients can be found in [73]. Tools from the ergodic theory of hypoelliptic diffusions can also be used [159].

The low-temperature asymptotics of the eigenvalues and eigenfunctions of the generator/Fokker–Planck operator is a problem very closely related to the semiclassical analysis of the Schrödinger operator. See [85] and the references therein. The calculation of the first nonzero eigenvalue and of the corresponding eigenfunction using singular perturbation theory can be found in [156].

Section 7.5 is based on [184]. The example of a Brownian particle moving in a piecewise linear potential and subject to a piecewise constant periodic force is taken from [151]. Brownian motors have been used as simple theoretical models for various intracellular transport processes such as molecular motors [31]. Further information can be found in the review article [196]; there, a classification of Brownian motors depending on whether the additional noise that breaks detailed balance is additive or multiplicative or we have a fluctuating potential can be found, and the terms pulsating, tilting, and temperature ratchets (Brownian motors) can be found. Temperature ratchets can be studied using the same techniques that we used in Sect. 7.5, and in particular, formulas for the drift and the diffusion coefficient can be derived. The model (7.73), in particular the calculation of the effective drift, can be found in [14, 46]. A very careful analysis of the effective velocity for SDEs with space- and time-periodic coefficients, as well as systems of linearly coupled SDEs (switching Markov processes), can be found in [37].

7.7 Exercises

- Calculate the mean exit time from the left well for the piecewise constant potential (7.20). Study its dependence on the temperature and the starting point x .
 - Consider a general quartic bistable potential $V(x) = \sum_{j=1}^4 \alpha_j x^4$. Find optimal piecewise linear and piecewise constant approximations after giving an appropriate definition of optimal approximation in this context. Calculate the mean first passage times from the left and the right wells, with particular emphasis on the asymmetric case. Compare the different approximations for the mean first passage times numerically, in particular in the low-temperature regime.
- Use (7.29) to derive (7.30) and (7.31), together with (7.32).
- Prove the properties of the eigenfunctions and eigenvalues of the generator \mathcal{L} of a reversible diffusion in a double-well potential that were used in Sect. 7.4,

- in particular formulas (7.43) and (7.44) (consult [156] or a book on singular perturbation theory, e.g., [91, 114] if necessary).
4. Calculate the signal-to-noise ratio for the symmetric bistable potential in the weak-amplitude, low-temperature regime. Plot the SNR as a function of the temperature for different values of the signal frequency. Calculate the optimal temperature. Do you obtain the same result as in the optimization of the spectral amplification, equation (7.46)?
 5. Obtain formula (7.79).
 6. Consider the dynamics

$$\dot{x}(t) = -\partial_x V(x(t), t) + y(t) + \sqrt{2T(x(t), t)} \xi(t), \quad (7.84)$$

for the space- and time-periodic potential $V(x, t)$ and temperature $T(x, t) > 0$, and periodic-in-time force $y(t)$ with period L . The equation is interpreted in the Itô sense. Use multiscale analysis to derive formulas for the effective drift and diffusion coefficients.

7. Use perturbation theory to obtain formulas (7.74) and (7.75).

Chapter 8

Derivation of the Langevin Equation

In this chapter, we derive the Langevin equation from a simple mechanical model for a small system (which we will refer to as a Brownian particle) that is in contact with a thermal reservoir that is at thermodynamic equilibrium at time $t = 0$. The full dynamics, Brownian particle plus thermal reservoir, are assumed to be Hamiltonian. The derivation proceeds in three steps. First, we derive a closed stochastic integrodifferential equation for the dynamics of the Brownian particle, the *generalized Langevin equation* (GLE). In the second step, we approximate the GLE by a finite-dimensional Markovian equation in an extended phase space. Finally, we use singular perturbation theory for Markov processes to derive the Langevin equation, under the assumption of rapidly decorrelating noise. This derivation provides a partial justification for the use of stochastic differential equations, in particular the Langevin equation, in the modeling of physical systems.

In Sect. 8.1, we study a simple model of open classical systems, and we derive the generalized Langevin equation. The Markovian approximation of the GLE is studied in Sect. 8.2. The derivation of the Langevin equation from this Markovian approximation is studied in Sect. 8.3. Discussion and bibliographical remarks are included in Sect. 8.4. Exercises can be found in Sect. 8.5.

8.1 Open Classical Systems

We consider a particle in one dimension that is in contact with a thermal reservoir (heat bath), a system with infinite heat capacity at temperature β^{-1} that interacts (exchanges energy) with the particle. We will model the reservoir as a system of infinitely many noninteracting particles that is in thermodynamic equilibrium at time $t = 0$. In particular, we will model the heat bath as a system of infinitely many harmonic oscillators whose initial energy is distributed according to the canonical (Boltzmann–Gibbs) distribution at temperature β^{-1} .

A finite collection of harmonic oscillators is a Hamiltonian system with Hamiltonian

$$H(\mathbf{p}, \mathbf{q}) = \frac{1}{2} \sum_{j=1}^N p_j^2 + \frac{1}{2} \sum_{j=1}^N q_j^2, \quad (8.1)$$

where for simplicity, we have set all the spring constants $\{k_j\}_{j=1}^N$ equal to 1. The corresponding canonical distribution (Gibbs measure) is

$$\mu_\beta(d\mathbf{p}, d\mathbf{q}) = \frac{1}{Z} e^{-\beta H(\mathbf{p}, \mathbf{q})} d\mathbf{p} d\mathbf{q} =: \rho_\beta(q, p) dq dp. \quad (8.2)$$

Since the Hamiltonian (8.1) is quadratic in both positions and momenta, the measure (8.2) is Gaussian. We set $\mathbf{z} = (\mathbf{q}, \mathbf{p}) \in \mathbb{R}^{2N} =: \mathbb{H}$ and denote by $\langle \cdot, \cdot \rangle$ the Euclidean inner product in (the Hilbert space) \mathbb{H} . Then for arbitrary vectors $\mathbf{h}, \mathbf{b} \in \mathbb{H}$, we have

$$\mathbb{E}\langle \mathbf{z}, \mathbf{h} \rangle = 0, \quad \mathbb{E}\left(\langle \mathbf{z}, \mathbf{h} \rangle \langle \mathbf{z}, \mathbf{b} \rangle\right) = \beta^{-1} \langle \mathbf{h}, \mathbf{b} \rangle. \quad (8.3)$$

We want to consider an infinite-dimensional extension of the above model for the heat bath. A natural infinite-dimensional extension of a finite system of harmonic oscillators is the wave equation $\partial_t^2 \varphi = \partial_x^2 \varphi$, which we write as a system of equations

$$\partial_t \varphi = \pi, \quad \partial_t \pi = \partial_x^2 \varphi. \quad (8.4)$$

The wave equation is an infinite-dimensional Hamiltonian system with Hamiltonian

$$\mathcal{H}(\pi, \varphi) = \frac{1}{2} \int_{\mathbb{R}} \left(|\pi|^2 + |\partial_x \varphi|^2 \right) dx. \quad (8.5)$$

It is convenient to introduce the Hilbert space H_E with the (energy) norm

$$\|\phi\|^2 = \int_{\mathbb{R}} \left(|\pi|^2 + |\partial_x \varphi|^2 \right) dx, \quad (8.6)$$

where $\phi = (\varphi, \pi)$. The corresponding inner product is

$$\langle \phi_1, \phi_2 \rangle = \int_{\mathbb{R}} \left(\partial_x \varphi_1(x) \overline{\partial_x \varphi_2(x)} + \pi_1(x) \overline{\pi_2(x)} \right) dx, \quad (8.7)$$

where the overbar denotes the complex conjugate. Using the notation (8.6), we can write the Hamiltonian for the wave equation as

$$\mathcal{H}(\phi) = \frac{1}{2} \|\phi\|^2.$$

We would like to extend the Gibbs distribution (8.2) to this infinite-dimensional system. However, the expression

$$\mu_\beta(d\pi d\varphi) = \frac{1}{Z} e^{-\beta \mathcal{H}(\varphi, \pi)} \prod_{x \in \mathbb{R}} d\pi d\varphi \quad (8.8)$$

is merely formal, since Lebesgue measure does not exist in infinite dimensions. However, this measure is Gaussian (the Hamiltonian \mathcal{H} is a quadratic functional in π and φ), and the theory of Gaussian measures in Hilbert spaces is well developed. This theory goes beyond the scope of this book.¹ For our purposes, it is sufficient to note that if X is a Gaussian random variable in the Hilbert space H_E with inner product (8.7), then $\langle X, f \rangle$ is a scalar Gaussian random variable with mean and variance

$$\mathbb{E}\langle X, f \rangle = 0, \quad \text{and} \quad \mathbb{E}(\langle X, f \rangle \langle X, h \rangle) = \beta^{-1} \langle f, h \rangle, \quad (8.9)$$

respectively. Notice the similarity between the formulas in (8.3) and (8.9).

We assume that the full dynamics of the particle coupled to the heat bath is Hamiltonian, described by a Hamiltonian function

$$\mathcal{H}(p, q, \pi, \varphi) = H(p, q) + \mathcal{H}_{HB}(\pi, \varphi) + H_I(q, \varphi). \quad (8.10)$$

We use $\mathcal{H}_{HB}(\pi, \varphi)$ to denote the Hamiltonian for the wave equation (8.5), while $H(p, q)$ denotes the Hamiltonian of the particle, whereas H_I describes the interaction between the particle and the field $\varphi = (\varphi, \pi)$. We assume that the coupling is only through the position q and φ ; it does not depend on the momentum p and the momentum field π . We assume that the particle is moving in a confining potential $V(q)$. Consequently,

$$H(p, q) = \frac{p^2}{2} + V(q). \quad (8.11)$$

Concerning the coupling, we assume that it is linear in the field φ and that it is translation-invariant:

$$H_I(q, \varphi) = \int_{\mathbb{R}} \varphi(x) \rho(x - q) dx. \quad (8.12)$$

The coupling between the particle and the heat bath depends crucially on the “charge density” $\rho(x)$, which is arbitrary at this point.

Now we make an approximation that will simplify considerably the analysis: Since the particle moves in a confining potential (think of a quadratic potential), we can assume that its position does not change too much. We will refer to this assumption as the *dipole approximation*. Consequently, we can perform a Taylor series expansion in (8.12), which, together with an integration by parts, gives (see Exercise 1)

$$H_I(q, \varphi) \approx q \int_{\mathbb{R}} \partial_x \varphi(x) \rho(x) dx. \quad (8.13)$$

The coupling now is linear in both q and φ . This will enable us to integrate out explicitly the fields φ and π from the equations of motion and to obtain a closed-form equation for the dynamics of the particle.

¹ A few basic facts about Gaussian measures in Hilbert spaces can be found in Sect. B.5 in the appendix.

On putting together (8.11), (8.5), and (8.13), the Hamiltonian (8.10) becomes

$$\mathcal{H}(p, q, \pi, \varphi) = \frac{p^2}{2} + V(q) + \frac{1}{2} \int_{\mathbb{R}} (|\pi|^2 + |\partial_x \varphi|^2) dx + q \int_{\mathbb{R}} \partial_x \varphi(x) \rho(x) dx. \quad (8.14)$$

Now we can derive Hamilton's equations of motion for the coupled particle-field model (8.14):

$$\dot{q} = \frac{\partial \mathcal{H}}{\partial p}, \quad \dot{p} = -\frac{\partial \mathcal{H}}{\partial q}, \quad (8.15a)$$

$$\partial_t \varphi = \frac{\delta \mathcal{H}}{\delta \pi}, \quad \partial_t \pi = -\frac{\delta \mathcal{H}}{\delta \varphi}, \quad (8.15b)$$

where $\frac{\delta \mathcal{H}}{\delta \varphi}$, $\frac{\delta \mathcal{H}}{\delta \pi}$ denote the functional derivatives with respect to φ and π .² Carrying out the differentiations, we obtain

$$\dot{q} = p, \quad \dot{p} = -V'(q) - \int_{\mathbb{R}} \partial_x \varphi(x) \rho(x) dx, \quad (8.16a)$$

$$\partial_t \varphi = \pi, \quad \partial_t \pi = \partial_x^2 \varphi + q \partial_x \rho. \quad (8.16b)$$

Our goal now is to solve Eq. (8.16b), which is a system of linear inhomogeneous differential equations, and then substitute into (8.16a). We will use the variation of constants formula (*Duhamel's principle*). It is more convenient to rewrite (8.16) in a slightly different form. First, we introduce the operator

$$\mathcal{A} = \begin{pmatrix} 0 & 1 \\ \partial_x^2 & 0 \end{pmatrix}, \quad (8.17)$$

acting on functions in H_E with inner product (8.7). It is not hard to show that \mathcal{A} is an antisymmetric operator in this space (see Exercise 2). Furthermore, we introduce the notation $\alpha = (\alpha_1(x), 0) \in H_E$ with $\partial_x \alpha_1(x) = \rho(x)$. Noticing that

$$\mathcal{A} \alpha = (0, \partial_x \rho),$$

we can rewrite (8.16b) in the form

$$\partial_t \phi = \mathcal{A}(\phi + q\alpha) \quad (8.18)$$

with $\phi = (\varphi, \pi)$. Furthermore, the second equation in (8.16a) becomes

$$\dot{p} = -V'(q) - \langle \phi, \alpha \rangle. \quad (8.19)$$

² For a functional of the form $\mathcal{H}(\phi) = \int_{\mathbb{R}} H(\phi, \partial_x \phi) dx$, the functional derivative is given by $\frac{\delta \mathcal{H}}{\delta \phi} = \frac{\partial H}{\partial \phi} - \frac{\partial}{\partial x} \frac{\partial H}{\partial (\partial_x \phi)}$. We apply this definition to the functional (8.14) to obtain $\frac{\delta \mathcal{H}}{\delta \pi} = \pi$, $\frac{\delta \mathcal{H}}{\delta \varphi} = -\partial_x^2 \varphi - q \partial_x \rho$.

Finally, we introduce the function $\psi = \phi + q\alpha$ to rewrite (8.18) as

$$\partial_t \psi = \mathcal{A} \psi + p \alpha. \quad (8.20)$$

Similarly, we introduce ψ in (8.19) to obtain

$$\dot{p} = -V'_{\text{eff}}(q) - \langle \psi, \alpha \rangle, \quad (8.21)$$

where

$$V_{\text{eff}}(q) = V(q) - \frac{1}{2} \|\alpha\|^2 q^2. \quad (8.22)$$

Note that

$$\|\alpha\|^2 = \|\rho\|_{L^2}^2 =: \lambda.$$

The parameter λ measures the strength of the coupling between the particle and the heat bath. The correction term in the potential $V_{\text{eff}}(q)$ is essentially due to the way we have chosen to write the equations of motion for the particle-field system, and it is not fundamental; see Exercise 1.

The solution of (8.20) is

$$\psi(t) = e^{\mathcal{A}t} \psi(0) + \int_0^t e^{\mathcal{A}(t-s)} p(s) \alpha ds.$$

We substitute this into (8.21) to obtain

$$\begin{aligned} \dot{p} &= -V'_{\text{eff}}(q) - \langle \psi, \alpha \rangle \\ &= -V'_{\text{eff}}(q) - \left\langle e^{\mathcal{A}t} \psi(0), \alpha \right\rangle - \int_0^t \left\langle e^{\mathcal{A}(t-s)} \alpha, \alpha \right\rangle p(s) ds \\ &=: -V'_{\text{eff}}(q) - \int_0^t \gamma(t-s) p(s) ds + F(t), \end{aligned}$$

where we have used the antisymmetry of \mathcal{A} in H_E ,

$$F(t) = -\left\langle \psi(0), e^{-\mathcal{A}t} \alpha \right\rangle, \quad (8.23)$$

and

$$\gamma(t) = \left\langle e^{-\mathcal{A}t} \alpha, \alpha \right\rangle. \quad (8.24)$$

Note that $\psi(0) = \phi(0) + q(0)\alpha$ is a Gaussian field with mean and covariance

$$\mathbb{E}\langle \psi(0), f \rangle = q(0) \langle \alpha, f \rangle =: \mu_f$$

and

$$\mathbb{E}((\langle \psi(0), f \rangle - \mu_f)(\langle \psi(0), h \rangle - \mu_h)) = \beta^{-1} \langle f, h \rangle,$$

where we have used (8.9). To simplify things, we will set $q(0) = 0$. Then $F(t)$ is a mean-zero stationary Gaussian process with autocorrelation function

$$\begin{aligned}\mathbb{E}(F(t)F(s)) &= \mathbb{E}\left(\left\langle \psi(0), e^{-\mathcal{A}t} \alpha \right\rangle \left\langle \psi(0), e^{-\mathcal{A}s} \alpha \right\rangle\right) \\ &= \beta^{-1} \left\langle e^{-\mathcal{A}t} \alpha, e^{-\mathcal{A}s} \alpha \right\rangle \\ &= \beta^{-1} \gamma(t-s),\end{aligned}$$

where we have used (8.9). Consequently, the autocorrelation function of the stochastic forcing in (8.23) is precisely the kernel (times the temperature) of the dissipation term in the equation for p . This is an example of the *fluctuation–dissipation theorem*.

To summarize, we have obtained a closed-form equation for the dynamics of the particle, the generalized Langevin equation

$$\ddot{q} = -V'_{\text{eff}}(q) - \int_0^t \gamma(t-s) \dot{q}(s) ds + F(t), \quad (8.25)$$

with $F(t)$ a mean-zero stationary Gaussian process with autocorrelation function given by the fluctuation–dissipation theorem,

$$\mathbb{E}(F(t)F(s)) = \beta^{-1} \gamma(t-s). \quad (8.26)$$

It is clear from formula (8.24) and the definition of α that the autocorrelation function $\gamma(t)$ depends only on the density ρ .³ In fact, we can show that (see Exercise 3)

$$\gamma(t) = \int_{\mathbb{R}} |\hat{\rho}(k)|^2 e^{ikt} dk, \quad (8.27)$$

where $\hat{\rho}(k)$ denotes the Fourier transform of ρ .

Let us now make several remarks on the generalized Langevin equation (8.25). First, notice that the GLE is Newton's equation of motion for the particle, augmented with two additional terms: a linear dissipation term that depends on the history of the particle position and a stochastic forcing term that is related to the dissipation term through the fluctuation–dissipation theorem (8.26). The fact that the fluctuations (noise) and the dissipation in the system satisfy such a relation is not surprising, since they have the same source, namely the interaction between the particle and the field. It is important to note that the noise (and also the fact that it is Gaussian and stationary) in the GLE is due to our assumption that the heat bath is at equilibrium at time $t = 0$, i.e., that the initial equations of the wave equation are distributed according to the (Gaussian) Gibbs measure (8.8). Perhaps surprisingly, the derivation of the GLE and the fluctuation–dissipation theorem are not related to our assumption that the heat bath is described by a field, i.e., it is a dynamical system with infinitely many degrees of freedom. We could have arrived at the GLE

³ Assuming, of course, that the heat bath is described by a wave equation, i.e., assuming that \mathcal{A} is the wave operator.

and the fluctuation–dissipation theorem even if we had only one oscillator in the “heat bath”; see Exercise 6.

Furthermore, the autocorrelation function of the noise depends only on the coupling function $\rho(x)$: different choices of the coupling function lead to different noise processes $F(t)$.⁴

It is also important to emphasize the fact that the GLE (8.25) is *equivalent* to the full Hamiltonian dynamics (8.14) with random initial conditions distributed according to (8.8). So far, no approximation has been made. We have merely used the linearity of the dynamics of the heat bath and the linearity of the coupling in order to integrate out the heat bath variables using the variation of constants formula.

Finally, we remark that an alternative way for writing the GLE is

$$\ddot{q} = -V'(q) - \int_0^t D(t-s)q(s)ds + F(t) \quad (8.28)$$

with

$$D(t) = \langle \mathcal{A}e^{\mathcal{A}t}\alpha, \alpha \rangle. \quad (8.29)$$

The fluctuation–dissipation theorem takes the form

$$\dot{\gamma}(t) = D(t). \quad (8.30)$$

See Exercise 7. In writing the GLE in the form (8.28), there is no need to introduce an effective potential or to assume that $q(0) = 0$.

8.2 The Markovian Approximation

From now on, we will ignore the correction in the potential (8.22); see Exercise 1 for a justification of this. We rewrite the GLE (8.25) as

$$\ddot{q} = -V'(q) - \int_0^t \gamma(t-s)\dot{q}(s)ds + F(t), \quad (8.31)$$

together with the fluctuation–dissipation theorem (8.26). Equation (8.31) is a non-Markovian stochastic equation, since the solution at time t depends on the entire past. In this section, we show that when the autocorrelation function $\gamma(t)$ decays sufficiently fast, the dynamics of the particle can be described by a Markovian system of stochastic differential equations in an extended phase space. The basic observation that was already made in Chap. 2, Exercise 6, that a one-dimensional mean-zero Gaussian stationary process with continuous paths and an exponential autocorrelation function is necessarily an Ornstein–Uhlenbeck process. This is the content of Doob’s theorem. Consequently, if the memory kernel (autocorrelation function) $\gamma(t)$ is decaying exponentially fast, then we expect that we can describe the noise in

⁴ In fact, the autocorrelation function depends also on the operator \mathcal{A} in (8.17).

the GLE by adding a finite number of auxiliary variables. We can formalize this by introducing the concept of a *quasi-Markovian process*:

Definition 8.1. We will say that a stochastic process X_t is quasi-Markovian if it can be represented as a Markovian stochastic process by adding a finite number of additional variables, that is, if there exists a finite-dimensional stochastic process Y_t such that $\{X_t, Y_t\}$ is a Markov process.

Proposition 8.1. Let $\lambda \in \mathbb{R}^m$, let $A \in \mathbb{R}^{m \times m}$ be positive definite, and assume that the autocorrelation function $\gamma(t)$ is given by

$$\gamma(t) = \langle e^{-At} \lambda, \lambda \rangle. \quad (8.32)$$

Then the generalized Langevin equation (8.31) is equivalent to the stochastic differential equation

$$dq(t) = p_t dt, \quad (8.33a)$$

$$dp(t) = (-V'(q(t)) + \langle \lambda, z(t) \rangle) dt, \quad (8.33b)$$

$$dz(t) = (-p(t)\lambda - Az(t)) dt + \Sigma dW(t), \quad z(0) \sim \mathcal{N}(0, \beta^{-1}I), \quad (8.33c)$$

where $z: \mathbb{R}^+ \mapsto \mathbb{R}^m$, $\lambda \in \mathbb{R}^m$, $\Sigma \in \mathbb{R}^{m \times m}$, and the matrix Σ satisfies

$$\Sigma \Sigma^T = \beta^{-1}(A + A^T). \quad (8.34)$$

Remark 8.1.

- (i) The formula for the autocorrelation function (8.32) is similar to (8.24). However, the operator \mathcal{A} in (8.24) is the wave operator (8.17), i.e., the generator of a unitary group, whereas the operator A (or rather, $-A$) that appears in (8.32) is the generator of the contraction semigroup e^{-At} , i.e., a dissipative operator. The sources of the noise in (8.25) and in (8.33) are quite different from each other, even though they have the same effect, when the autocorrelation function is exponentially decaying.
- (ii) Equation (8.34) is the form that the fluctuation–dissipation theorem (8.26) takes when the memory kernel is given by (8.32).

Proof of Proposition 8.1. The solution of (8.33c) is

$$z(t) = e^{-At} z(0) + \int_0^t e^{-A(t-s)} \Sigma dW(s) - \int_0^t e^{-A(t-s)} \lambda p(s) ds. \quad (8.35)$$

We substitute this into (8.33b) to obtain

$$\dot{p} = -V'(q) - \int_0^t \gamma(t-s) p(s) ds + F(t)$$

with

$$\gamma(t) = \langle e^{-At} \lambda, \lambda \rangle$$

and

$$\begin{aligned} F(t) &= \left\langle \lambda, e^{-At} z(0) + \int_0^t e^{-A(t-s)} \Sigma dW(s) \right\rangle \\ &=: \langle \lambda, y(t) \rangle, \end{aligned}$$

where

$$y(t) = S(t)z(0) + \int_0^t S(t-s) \Sigma dW(s)$$

with $S(t) = e^{-At}$. With our assumptions on $Z(0)$ and (8.34), $y(t)$ is a mean-zero stationary Gaussian process with covariance matrix

$$Q(t-s) = \mathbb{E}(y^T(t)y(s)) = \beta^{-1}S(|t-s|). \quad (8.36)$$

To see this, we first note that (using the summation convention)

$$\begin{aligned} \mathbb{E}(y_i(t)y_j(s)) &= S_{i\ell}(t)S_{j\rho}(t)\mathbb{E}(z_\ell(0)z_\rho(0)) \\ &\quad + \int_0^t \int_0^s S_{i\rho}(t-\ell)\Sigma_{\rho k}S_{jn}(s-\tau)\Sigma_{nk}\delta(\ell-\tau)d\ell dm \\ &= \beta^{-1}S_{i\rho}(t)S_{j\rho}(s) + \int_0^{\min(t,s)} S_{i\rho}(t-\tau)\Sigma_{\rho k}\Sigma_{nk}S_{jn}(s-\tau)d\tau. \end{aligned}$$

Consequently, using (8.34),

$$\begin{aligned} \mathbb{E}(y^T(t)y(s)) &= \beta^{-1}S(t)S^T(s) + \int_0^{\min(t,s)} S(t-\tau)\Sigma^T\Sigma S^T(s-\tau)d\tau \\ &= \beta^{-1}S(t) \left(I + \int_0^{\min(t,s)} S(-\tau)(A+A^T)S^T(-\tau)d\tau \right) S^T(s). \end{aligned}$$

Without loss of generality, we may assume that $s \leq t$. Now we claim that

$$\left(I + \int_0^{\min(t,s)} S(-\tau)(A+A^T)S^T(-\tau)d\tau \right) S^T(s) = S(-s).$$

To see this, note that this equation is equivalent to

$$I + \int_0^{\min(t,s)} S(-\tau)(A+A^T)S^T(-\tau)d\tau = S(s)S^T(-s).$$

This equation is clearly valid at $s = 0$. We differentiate to obtain the identity

$$S(-s)(A+A^T)S^T(-s) = \frac{d}{dt}S(s)S^T(-s),$$

which is true for all s . This completes the proof of (8.36). Now we calculate, with $s \leq t$,

$$\begin{aligned}\mathbb{E}(F(t)F(s)) &= \mathbb{E}(\langle \lambda, y(t) \rangle \langle \lambda, y(s) \rangle) \\ &= \langle Q(t-s)\lambda, \lambda \rangle = \beta^{-1} \langle e^{-At}\lambda, \lambda \rangle \\ &= \beta^{-1} \gamma(t-s),\end{aligned}$$

and the proposition is proved. \square

Example 8.1. Consider the case $m = 1$. In this case, the vector λ and the matrix A become scalar quantities. The SDE (8.33) becomes

$$\begin{aligned}dq(t) &= p(t) dt, \\ dp(t) &= (-V'(q(t)) + \lambda z(t)) dt, \\ dz(t) &= (-\lambda p(t) - \alpha z(t)) dt + \sqrt{2\alpha\beta^{-1}} dW(t), \quad z(0) \sim \mathcal{N}(0, \beta^{-1}).\end{aligned}$$

The autocorrelation function is

$$\gamma(t) = \lambda^2 e^{-\alpha t}.$$

Example 8.2. Consider now the case

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

From (8.34), we deduce that

$$A = 2\beta^{-1}\gamma \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Thus, the noise describing the heat bath is degenerate in this example. We also take the coupling vector $\lambda = (\lambda, 0)$. The Markovian GLE takes the form

$$dq = pdt, \tag{8.37a}$$

$$dp = (-V'(q) + \lambda z_1) dt, \tag{8.37b}$$

$$dz_1 = (z_2 - \lambda_1 p) dt, \tag{8.37c}$$

$$dz_2 = (-z_1 - \gamma z_2) dt + \sqrt{2\beta^{-1}\gamma} dW_t. \tag{8.37d}$$

Note that the noise process $F(t)$ in the corresponding GLE is the stationary Gaussian process that is obtained as the solution of the Langevin equation in a harmonic potential with stationary initial conditions:

$$\begin{aligned}dF &= y dt, \\ dy &= (-F - \gamma y) dt + \sqrt{2\beta^{-1}\gamma} dW_t.\end{aligned}$$

This noise process, appropriately rescaled, can be used as a regularized approximation to white noise; it is called the *harmonic noise*. Note that harmonic noise provides us with a smoother approximation to white noise in comparison to the Ornstein–Uhlenbeck approximation that we studied in Sect. 5.1.

The generator of the dynamics (8.33) is

$$\mathcal{L} = p\partial_q - \partial_q V \partial_p + \langle \lambda, z \rangle \partial_p - p\lambda \cdot \nabla_z - Az \cdot \nabla_z + \beta^{-1} A : D_z. \quad (8.38)$$

The Fokker–Planck operator is

$$\mathcal{L}^* = -p\partial_q + \partial_q V \partial_p - \langle \lambda, z \rangle \partial_p + p\lambda \cdot \nabla_z + \nabla_z (Az \cdot) + \beta^{-1} A : D_z. \quad (8.39)$$

When the potential $V(q)$ is confining, then the process $X(t) := (q(t), p(t), z(t))$ has nice ergodic properties. We recall that the Hamiltonian of the system is $H(p, q) = \frac{1}{2}p^2 + V(q)$.

Proposition 8.2. *Assume that the potential V in (8.33) is confining. Then the process $X(t) := (q(t), p(t), z(t))$ is ergodic with invariant distribution*

$$\rho_\beta(q, p, \mathbf{z}) = \frac{1}{Z} e^{-\beta(H(p, q) + \frac{1}{2}\|\mathbf{z}\|^2)}, \quad Z = (2\pi\beta^{-1})^{m+1} \int_{\mathbb{R}} e^{-\beta V(q)} dq. \quad (8.40)$$

Proof. We prove only that (8.40) is an invariant distribution. The uniqueness is discussed in Sect. 8.4. We have to check that ρ_β is a solution of the stationary Fokker–Planck equation

$$\mathcal{L}^* \rho_\beta = 0.$$

We have

$$(-p\partial_q + \partial_q V \partial_p) e^{-\beta H(q, p)} = 0.$$

Furthermore,

$$(-\langle \lambda, z \rangle \partial_p + p\lambda \cdot \nabla_z) e^{-\beta(\frac{1}{2}p^2 + \frac{1}{2}\|\mathbf{z}\|^2)} = 0.$$

Finally,

$$\nabla_z \cdot (Az + \beta^{-1} A \nabla_z) e^{-\frac{1}{2}\beta\|\mathbf{z}\|^2} = 0.$$

The formula for the partition function follows from Gaussian integration. \square

Remark 8.2. Observe that the invariant distribution is independent of the vector λ and the matrix A .

As in the case of the Langevin dynamics, we can work in the weighted L^2 space $L^2(\mathbb{R}^{2+m}; \rho_\beta)$. In this space, the generator (8.38) can be naturally decomposed into its symmetric and antisymmetric parts, similarly to the generator of the Langevin dynamics that was studied in Chap. 6. We denote by A_a and A_s the antisymmetric and symmetric parts of the matrix A , respectively.

⁵ In fact, the last term in (8.38) should read $\beta^{-1} A_s : D_z$, where $A_s = \frac{1}{2}(A + A^T)$ denotes the symmetric part of A . However, since D_z is symmetric, we can write it in the form $\beta^{-1} A : D_z$.

Proposition 8.3. *The generator (8.38) can be written as*

$$\mathcal{L} = \mathcal{A} + \mathcal{S}, \quad (8.41)$$

where

$$\mathcal{A} = p\partial_q - \partial_q V\partial_p + \langle \lambda, z \rangle \partial_p - \langle A_a z, \nabla_z \rangle$$

and

$$\mathcal{S} = \langle -A_s z, \nabla_z \rangle + \beta^{-1} A_s : D_z.$$

Furthermore, \mathcal{A} and \mathcal{S} are antisymmetric and symmetric operators, respectively, with respect to the $L^2(\mathbb{R}^{2+m}; \rho_\beta)$ inner product.

The proof of this proposition is left as an exercise.

8.3 Derivation of the Langevin Equation

Now we are ready to derive the Langevin equation

$$dq(t) = p(t) dt, \quad dp(t) = -V'(q(t)) dt - \gamma p(t) dt + \sqrt{2\gamma\beta^{-1}} dW(t), \quad (8.42)$$

and to obtain a formula for the friction coefficient γ . We will derive (8.42) from the GLE (8.31) in the limit of vanishing correlation time of the noise, $\gamma(t) \rightarrow \delta(t)$. This corresponds to taking the coupling in the full Hamiltonian dynamics (8.14) to be localized, $\rho(x) \rightarrow \delta(x)$.

We focus on the Markovian approximation (8.33) with the family of autocorrelation functions

$$\gamma^\varepsilon(t) = \frac{1}{\varepsilon^2} \langle e^{-\frac{A}{\varepsilon^2}t} \lambda, \lambda \rangle.$$

This corresponds to rescaling λ and A in (8.33) according to $\lambda \mapsto \lambda/\varepsilon$ and $A \mapsto A/\varepsilon^2$. Equation (8.33) becomes

$$dq^\varepsilon(t) = p^\varepsilon(t) dt, \quad (8.43a)$$

$$dp^\varepsilon(t) = \left(-V'(q^\varepsilon(t)) + \frac{1}{\varepsilon} \langle \lambda, z^\varepsilon(t) \rangle \right) dt, \quad (8.43b)$$

$$dz^\varepsilon(t) = \left(-\frac{1}{\varepsilon} p^\varepsilon(t) \lambda - \frac{1}{\varepsilon^2} A z^\varepsilon(t) \right) dt + \frac{1}{\varepsilon} \Sigma dW(t), \quad z^\varepsilon(0) \sim \mathcal{N}(0, \beta^{-1} I),$$

where (8.34) has been used.

Result 8.4. *Let $\{q^\varepsilon(t), p^\varepsilon(t), z^\varepsilon(t)\}$ denote the solution of (8.43), and assume that the matrix A is invertible. Then $\{q^\varepsilon(t), p^\varepsilon(t)\}$ converges weakly to the solution of the Langevin equation (8.42), where the friction coefficient is given by the formula*

$$\gamma(t) = \langle \lambda, A^{-1} \lambda \rangle. \quad (8.44)$$

Remark 8.3. Note that (8.44) is equivalent to the formula

$$\gamma = \int_0^{+\infty} \gamma(t) dt,$$

which is used quite often in statistical mechanics. Formula (8.44) can also be expressed in terms of the solution of an appropriate linear equation:

$$\gamma = \langle \lambda, \phi \rangle, \quad A\phi = \lambda.$$

These formulas are similar to those that we obtained in Chap. 6 for the diffusion coefficient of a Brownian particle in a periodic potential as well as those that we will obtain in Chap. 9 in the context of the Green–Kubo formalism.

Proof of Result 8.4. The backward Kolmogorov equation corresponding to (8.43) is

$$\frac{\partial u^\varepsilon}{\partial t} = \frac{1}{\varepsilon^2} \mathcal{L}_0 + \frac{1}{\varepsilon} \mathcal{L}_1 + \mathcal{L}_2 \quad (8.45)$$

with

$$\begin{aligned} \mathcal{L}_0 &= -\langle Az, \nabla_z \rangle + \beta^{-1} A : D_z, \\ \mathcal{L}_1 &= \langle \lambda, z \rangle \partial_p - p \langle \lambda, \nabla_z \rangle, \\ \mathcal{L}_2 &= p \partial_q - \partial_q V \partial_p. \end{aligned}$$

We look for a solution to (8.45) in the form of a power series expansion in ε :

$$u^\varepsilon = u_0 + \varepsilon u_1 + \varepsilon^2 u_2 + \dots$$

We substitute this into (8.45) and equate powers of ε to obtain the sequence of equations

$$\mathcal{L}_0 u_0 = 0, \quad (8.46a)$$

$$-\mathcal{L}_0 u_1 = \mathcal{L}_1 u_0, \quad (8.46b)$$

$$-\mathcal{L}_0 u_2 = \mathcal{L}_1 u_1 + \mathcal{L}_2 u_0 - \frac{\partial u_0}{\partial t}. \quad (8.46c)$$

$$\dots = \dots$$

From the first equation, we deduce that to leading order, the solution of the Kolmogorov equation is independent of the auxiliary variables \mathbf{z} , $u_0 = u(q, p, t)$. The solvability of the second equation reads

$$\int_{\mathbb{R}^d} \mathcal{L}_1 u_0 e^{-\frac{\beta}{2} \|\mathbf{z}\|^2} dz = 0,$$

which is satisfied, since

$$\mathcal{L}_1 u_0 = \langle \lambda, \mathbf{z} \rangle \frac{\partial u}{\partial p}.$$

The solution to the equation

$$-\mathcal{L}_0 u_1 = \langle \lambda, \mathbf{z} \rangle \frac{\partial u}{\partial p}$$

is

$$u_1(q, p, t) = \langle (A^T)^{-1} \lambda, \mathbf{z} \rangle \frac{\partial u}{\partial p},$$

plus an element in the null space of \mathcal{L}_0 , which, as we expect from similar calculations that we have done in earlier chapters, for example in Sect. 3.2, will not affect the limiting equation.

Now we use the solvability condition for (8.46c) to obtain the backward Kolmogorov equation corresponding to the Langevin equation. The solvability condition gives

$$\frac{\partial u}{\partial t} = \mathcal{L}_2 u + \langle \mathcal{L}_1 u_1 \rangle_\beta,$$

where

$$\langle \cdot \rangle_\beta := (2\pi\beta^{-1})^{-m} \int_{\mathbb{R}^m} \cdot e^{-\frac{\beta}{2} \|\mathbf{z}\|^2} dz.$$

We calculate

$$\langle \mathcal{L}_1 u_1 \rangle_\beta = \beta^{-1} \langle (A^T)^{-1} \lambda, \lambda \rangle \frac{\partial^2 u}{\partial p^2} - \langle (A^T)^{-1} \lambda, \lambda \rangle p \frac{\partial u}{\partial p}.$$

Consequently, u is the solution of the PDE

$$\frac{\partial u}{\partial t} = (p \partial_q - \partial_q V \partial_p - \gamma p \partial_p + \gamma \beta^{-1} \partial_p^2) u,$$

where γ is given by (8.44). This is precisely the backward Kolmogorov equation of the Langevin equation (8.42). \square

8.4 Discussion and Bibliography

Section 8.1 is based on [202]. The generalized Langevin equation was studied extensively in [103–105], where existence and uniqueness of solutions as well as ergodic properties were established. An earlier reference on the construction of heat baths is [139]. The ergodic properties of a chain of anharmonic oscillators, coupled to two Markovian heat baths (i.e., with an exponential autocorrelation function) at

different temperatures, were studied in [48, 50, 51, 203]. See also the review article [201]. The Markovian approximation of the generalized Langevin equation was studied in [129]. See also [179].

A natural question that arises is whether it is possible to approximate the generalized Langevin equation (8.25) with an arbitrary memory kernel by a Markovian system of the form (8.33). This is essentially a problem in approximation theory that was studied in [123, 124, 220]. A systematic methodology for obtaining Markovian approximations to the GLE, which is based on the continued fraction expansion of the Laplace transform of the autocorrelation function of the noise in the GLE, was introduced by Mori in [169].

We can also consider a “Brownian particle” that is coupled to a finite number of harmonic oscillators that are in equilibrium. A calculation similar to what we have done in Sect. 8.1 leads to a GLE in which the noise depends on the number of particles in the heat bath. One then passes to the thermodynamic limit, i.e., the limit in which the number of particles in the heat bath becomes infinite, to obtain the GLE; see Exercise 6. This model is called the Kac–Zwanzig model and was introduced in [60, 245]. See also [59]. Further information on the Kac–Zwanzig model can be found in [4, 38, 71, 141]. Nonlinear coupling between the distinguished particle and the harmonic heat bath is studied in [131]. The Kac–Zwanzig model can be used to compare the results of reaction rate theory that was developed in Chap. 7 with techniques for calculating reaction rates that are appropriate for Hamiltonian systems such as *transition state theory*. See [3, 82, 187, 188].

We emphasize the fact that the GLE obtained in Sect. 8.1 from the coupled particle-field model (8.10) is exact. Of course, all the information about the environment is contained in the noise process and the autocorrelation function. The rather straightforward derivation of the GLE is based on the linearity of the thermal reservoir and on the linear coupling. Similar derivations are also possible for more general Hamiltonian systems of the form (6.7) using projection operator techniques. This approach is usually referred to as the *Mori–Zwanzig formalism*. The Mori–Zwanzig formalism is studied in [127, 161, 246]. It is possible to derive Langevin (or Fokker–Planck) equation in some appropriate asymptotic limit, for example, in the limit as the ratio of the mass of the particles in the bath and the (much heavier) Brownian particle tends to 0. See [161, 223]. This asymptotic limit goes back to Einstein’s original work on Brownian motion. A rigorous study of such a model is presented in [47].

8.5 Exercises

1. Derive (8.13) from (8.12). Show that the next term in the expansion compensates for the correction term in the effective potential (8.22).
2. Show that the operator \mathcal{A} defined in (8.17) is antisymmetric in the Hilbert space H_E with inner product (8.7). Conclude that $(e^{\mathcal{A}t})^* = e^{-\mathcal{A}t}$. Prove that the

one-parameter family of operators $e^{\mathcal{A}t}$ forms a unitary group. (This is usually referred to as *Stone's theorem*. See [193]).

3. Solve the wave equation (8.4) by taking the Fourier transform. In particular, calculate $e^{-\mathcal{A}t}$ in Fourier space. Use this to prove (8.27).
4. Solve the GLE (8.31) for the free particle $V \equiv 0$ and when the potential is quadratic. (*Hint:* Use the Laplace transform; see [129].)
5. (a) Consider a system of N harmonic oscillators governed by the Hamiltonian

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

Assume that the initial conditions are distributed according to the distribution $\frac{1}{Z}e^{-\beta H(\mathbf{p}, \mathbf{q})}$ with $\beta > 0$ and $\mathbf{q} = (q_1, \dots, q_N)$, $\mathbf{p} = (p_1, \dots, p_N)$. Compute the average kinetic energy for this system as a function of time.

- (b) Do the same calculation for the Hamiltonian

$$H(\mathbf{q}, \mathbf{p}) = \frac{1}{2} \langle A\mathbf{p}, \mathbf{p} \rangle + \frac{1}{2} \langle B\mathbf{q}, \mathbf{q} \rangle,$$

where $\mathbf{q}, \mathbf{p} \in \mathbb{R}^N$, $A, B \in \mathbb{R}^{N \times N}$ are symmetric strictly positive definite matrices, and the initial conditions are distributed according to $\frac{1}{Z}e^{-\beta H(\mathbf{p}, \mathbf{q})}$.

6. (The Kac–Zwanzig model). Consider the Hamiltonian

$$H(Q_N, P_N, q, p) = \frac{P_N^2}{2} + V(Q_N) + \sum_{n=1}^N \left[\left(\frac{p_n^2}{2m_n} + \frac{1}{2}m_n\omega_n^2 q_n^2 \right) - \lambda \mu_n q_n Q_N \right], \quad (8.47)$$

where the subscript N in the notation for the position and momentum of the distinguished particle, Q_N and P_N , emphasizes their dependence on the number N of the harmonic oscillators in the heat bath, $V(Q)$ denotes the potential experienced by the Brownian particle, and $\lambda > 0$ is the coupling constant. Assume that the initial conditions of the Brownian particle are deterministic and that those of the particles in the heat bath are Gaussian distributed according to the distribution $\frac{1}{Z}e^{-\beta H(\mathbf{p}, \mathbf{q})}$.

- (a) Obtain the generalized Langevin equation and prove the fluctuation–dissipation theorem.
- (b) Assume that the frequencies $\{\omega_n\}_{n=1}^N$ are random variables. Investigate under what assumptions on their distribution it is possible to pass to the thermodynamic limit (see [71]).
7. Derive Eqs. (8.28), (8.29), and (8.30).
8. Prove Proposition 8.3.
9. Analyze the models studied in this chapter in the multidimensional case, i.e., when the Brownian particle is a d -dimensional Hamiltonian system.

Chapter 9

Linear Response Theory for Diffusion Processes

In this chapter, we study the effect of a weak external forcing on a system at equilibrium. The forcing moves the system away from equilibrium, and we are interested in understanding the response of the system to this forcing. We study this problem for ergodic diffusion processes using perturbation theory. In particular, we develop *linear response theory*. The analysis of weakly perturbed systems leads to fundamental results such as the *fluctuation–dissipation theorem* and to the Green–Kubo formula, which enables us to calculate transport coefficients.

Linear response theory is developed in Sect. 9.1. The fluctuation–dissipation theorem is presented in Sect. 9.2. The Green–Kubo formula is derived in Sect. 9.3. Discussion and bibliographical remarks are included in Sect. 9.4. Exercises can be found in Sect. 9.5.

9.1 Linear Response Theory

The setup that we will consider is the following. Let X_t denote a stationary dynamical system with state space \mathbb{X} and invariant measure $\mu(dx) = f_\infty(x)dx$. We probe the system by adding a time-dependent forcing $\varepsilon F(t)$ with $\varepsilon \ll 1$ at time t_0 .¹ Our goal is to calculate the distribution function $f^\varepsilon(x, t)$ of the perturbed systems X_t^ε , for $\varepsilon \ll 1$, in particular in the long-time limit $t \rightarrow +\infty$. We can then calculate the expectation value of observables as well as correlation functions.

¹ The natural choice is $t_0 = 0$. Sometimes, it is convenient to take $t_0 = -\infty$.

We assume that the distribution function $f^\varepsilon(x, t)$ satisfies a linear kinetic equation e.g., the Liouville or the Fokker–Planck equation:²

$$\frac{\partial f^\varepsilon}{\partial t} = \mathcal{L}^{*\varepsilon} f^\varepsilon, \quad (9.1a)$$

$$f^\varepsilon|_{t=t_0} = f_\infty. \quad (9.1b)$$

The choice of the initial conditions reflects the fact that at $t = t_0$, the system is at equilibrium.

The operator $\mathcal{L}^{*\varepsilon}$ can be written in the form

$$\mathcal{L}^{*\varepsilon} = \mathcal{L}_0^* + \varepsilon \mathcal{L}_1^*, \quad (9.2)$$

where \mathcal{L}_0^* denotes the Liouville or Fokker–Planck operator of the unperturbed system, and \mathcal{L}_1^* is related to the external forcing. Throughout this section, we will assume that \mathcal{L}_1^* is of the form

$$\mathcal{L}_1^* = F(t) \cdot \mathcal{D}, \quad (9.3)$$

where \mathcal{D} is some linear (differential) operator. Since f_∞ is the unique equilibrium distribution, we have that

$$\mathcal{L}_0^* f_\infty = 0. \quad (9.4)$$

Before we proceed with the analysis of (9.1), we present a few examples.

Example 9.1 (A deterministic dynamical system). Let X_t be the solution of the differential equation

$$\frac{dX_t}{dt} = h(X_t), \quad (9.5)$$

on a (possibly compact) state space \mathbb{X} . We add a weak time-dependent forcing to obtain the dynamics

$$\frac{dX_t}{dt} = h(X_t) + \varepsilon F(t). \quad (9.6)$$

We assume that the unperturbed dynamics has a unique invariant distribution f_∞ that is the solution of the stationary Liouville equation

$$\nabla \cdot (h(x) f_\infty) = 0, \quad (9.7)$$

equipped with appropriate boundary conditions. The operator $\mathcal{L}^{*\varepsilon}$ in (9.2) has the form

$$\mathcal{L}^{*\varepsilon} \cdot = -\nabla \cdot (h(x) \cdot) - \varepsilon F(t) \cdot \nabla \cdot.$$

In this example, the operator \mathcal{D} in (9.3) is $\mathcal{D} = -\nabla$.

² Note that to be consistent with the notation that we have used previously in the book, in (9.1a) we use \mathcal{L}^* instead of \mathcal{L} , since the operator that appears in the Liouville or the Fokker–Planck equation is the adjoint of the generator.

A particular case of a deterministic dynamical system of the form (9.5), and the most important in statistical mechanics, is that of an N -body Hamiltonian system.

Example 9.2 (A stochastic dynamical system). Let X_t be the solution of the stochastic differential equation

$$dX_t = h(X_t) dt + \sigma(X_t) dW_t, \quad (9.8)$$

on \mathbb{R}^d , where $\sigma(x)$ is a positive semidefinite matrix and where the Itô interpretation is used. We add a weak time-dependent forcing to obtain the dynamics

$$dX_t = h(X_t) dt + \varepsilon F(t) dt + \sigma(X_t) dW_t. \quad (9.9)$$

We assume that the unperturbed dynamics has a unique invariant distribution f_∞ that is the solution of the stationary Fokker–Planck equation

$$-\nabla \cdot (h(x) f_\infty) + \frac{1}{2} D^2 : (\Sigma(x) f_\infty) = 0, \quad (9.10)$$

where $\Sigma(x) = \sigma(x) \sigma^T(x)$. The operator $\mathcal{L}^{*\varepsilon}$ in (9.2) has the form

$$\mathcal{L}^{*\varepsilon} \cdot = -\nabla \cdot (h(x) \cdot) + \frac{1}{2} D^2 : (\Sigma(x) \cdot) - \varepsilon F(t) \cdot \nabla:$$

As in the previous example, the operator \mathcal{D} in (9.3) is $\mathcal{D} = -\nabla$.

Example 9.3. A particular case of Example 9.2 is the Langevin equation:

$$\ddot{q} = -\nabla V(q) + \varepsilon F(t) - \gamma \dot{q} + \sqrt{2\gamma\beta} \dot{W}. \quad (9.11)$$

Writing (9.11) as a system of SDEs, we have

$$dq_t = p_t dt, \quad dp_t = -\nabla V(q_t) dt + \varepsilon F(t) dt - \gamma p_t dt + \sqrt{2\gamma\beta} dW_t. \quad (9.12)$$

For this example, we have $\mathcal{D} = -\nabla_p$, and assuming that V is a confining potential, $f_\infty = \frac{1}{Z} e^{-\beta H(p,q)}$, $H(p,q) = \frac{1}{2} p^2 + V(q)$. We will study this example in detail later on.

Example 9.4. Consider again the Langevin dynamics with a time-dependent temperature. The perturbed dynamics are

$$dq_t = p_t dt, \quad dp_t = -\nabla V(q_t) dt - \gamma p_t dt + \sqrt{2\gamma\beta^{-1}(1 + \varepsilon T(t))} dW_t, \quad (9.13)$$

with $1 + \varepsilon T(t) > 0$. In this case, the operator \mathcal{D} is

$$\mathcal{D} = \gamma\beta^{-1} \Delta_p.$$

The general case in which both the drift and the diffusion are perturbed is considered in Exercise 1.

Now we proceed with the analysis of (9.1). We look for a solution in the form of a power series expansion in ε :

$$f^\varepsilon = f_0 + \varepsilon f_1 + \dots \quad (9.14)$$

We substitute this into (9.1a) and use the initial condition (9.1b) to obtain the equations

$$\frac{\partial f_0}{\partial t} = \mathcal{L}_0^* f_0, \quad f_0|_{t=0} = f_\infty, \quad (9.15a)$$

$$\frac{\partial f_1}{\partial t} = \mathcal{L}_0^* f_1 + \mathcal{L}_1^* f_0, \quad f_1|_{t=0} = 0. \quad (9.15b)$$

The only solution to (9.15a) is

$$f_0 = f_\infty.$$

We substitute this into (9.15b) and use (9.3) to obtain

$$\frac{\partial f_1}{\partial t} = \mathcal{L}_0^* f_1 + F(t) \cdot \mathcal{D}f_\infty, \quad f_1|_{t=0} = 0.$$

We use the variation of constants formula to solve this equation:

$$f_1(t) = \int_{t_0}^t e^{\mathcal{L}_0^*(t-s)} F(s) \cdot \mathcal{D}f_\infty ds. \quad (9.16)$$

It is possible to calculate higher-order terms in the expansion for f^ε ; see Exercise 2. For our purposes, the calculation of $f_1(t)$ is sufficient.

Now we can calculate the deviation in the expectation value of an observable due to the external forcing. Let $\langle \cdot \rangle_{eq}$ and $\langle \cdot \rangle$ denote the expectation values with respect to f_∞ and f^ε , respectively. Let $A(\cdot)$ be an observable (phase-space function), and denote by $A(t)$ the deviation of its expectation value from equilibrium, to leading order:

$$\begin{aligned} A(t) &:= \langle A(X_t) \rangle - \langle A(X_t) \rangle_{eq} \\ &= \int A(x) (f^\varepsilon(x, t) - f_{eq}(x)) dx \\ &= \varepsilon \int A(x) \left(\int_{t_0}^t e^{\mathcal{L}_0^*(t-s)} F(s) \cdot \mathcal{D}f_\infty ds \right) dx. \end{aligned}$$

Assuming now that we can interchange the order of integration, we can rewrite the above formula as

$$\begin{aligned} A(t) &= \varepsilon \int A(x) \left(\int_{t_0}^t e^{\mathcal{L}_0^*(t-s)} F(s) \cdot \mathcal{D}f_\infty ds \right) dx \\ &= \varepsilon \int_{t_0}^t \left(\int A(x) e^{\mathcal{L}_0^*(t-s)} \cdot \mathcal{D}f_\infty dx \right) F(s) ds \\ &=: \varepsilon \int_{t_0}^t R_{\mathcal{L}_0, A}(t-s) F(s) ds, \end{aligned} \quad (9.17)$$

where we have defined the *response function*

$$R_{\mathcal{L}_0, A}(t) = \int A(x) e^{\mathcal{L}_0^* t} \mathcal{D}f_\infty dx. \quad (9.18)$$

We set now the lower limit of integration in (9.17) to be $t_0 = -\infty$ (we extend the definition of $R_{\mathcal{L}_0, A}(t)$ in (9.18) to be 0 for $t < 0$) and assume that $R_{\mathcal{L}_0, A}(t)$ decays to 0 as $t \rightarrow +\infty$ sufficiently fast, so that we can extend the upper limit of integration to $+\infty$ to write

$$A(t) = \varepsilon \int_{-\infty}^{+\infty} R_{\mathcal{L}_0, A}(t-s) F(s) ds. \quad (9.19)$$

As expected (since we have used linear perturbation theory), the deviation of the expectation value of an observable from its equilibrium value is a linear function of the forcing term. Note also that (9.19) has the form of the solution of a linear differential equation with $R_{\mathcal{L}_0, A}(t)$ playing the role of the Green's function. If we consider a delta-like forcing at $t = 0$, $F(t) = \delta(t)$, then the above formula gives

$$A(t) = \varepsilon R_{\mathcal{L}_0, A}(t).$$

Thus, the response function gives the deviation of the expectation value of an observable from equilibrium for a delta-like force.

Consider now a constant force $F(t) = F\Theta(t)$ that is exerted on the system at time $t = 0$, where $\Theta(t)$ denotes the Heaviside step function. For this forcing, (9.17) becomes

$$A(t) = \varepsilon F \int_0^t R_{\mathcal{L}_0, A}(t-s) ds. \quad (9.20)$$

Example 9.5 (Stochastic resonance; see Sect. 7.4.). Linear response theory provides us with a very elegant method for calculating the noise amplification factor for a particle moving in a double-well potential in the presence of thermal fluctuations under the influence of a weak external forcing. We consider the model [cf. Eq. (7.22)]

$$dX_t = -V'(X_t) dt + A_0 \cos(\omega_0 t) dt + \sqrt{2\beta^{-1}} dW_t. \quad (9.21)$$

Our goal is to calculate the average position $\langle X_t \rangle$ in the regime $A_0 \ll 1$. We can use (9.17) and (9.18). The generator of the unperturbed dynamics is the generator of the reversible dynamics

$$dX_t = -V'(X_t) dt + \sqrt{2\beta^{-1}} dW_t.$$

We have

$$\mathcal{D} = -\frac{\partial}{\partial x}, \quad f_\infty(x) = \frac{1}{Z} e^{-\beta V(x)}, \quad F(t) = \cos(\omega_0 t).$$

The observable in which we are interested is the particle's position. The response function is

$$\begin{aligned} R_{\mathcal{L}_0, x}(t) &= \int x e^{\mathcal{L}_0^* t} \left(-\frac{\partial}{\partial x} f_\infty(x) \right) dx \\ &= \beta \int \left(e^{\mathcal{L}_0 t} x \right) V'(x) f_\infty(x) dx \\ &= \beta \langle X_t V'(X_t) \rangle_{eq}. \end{aligned}$$

Let now $\{\lambda_\ell, \phi_\ell\}_{\ell=0}^\infty$ denote the eigenvalues and eigenfunctions of the unperturbed generator (4.103). We calculate (see Exercise 12)

$$\langle X_t V'(X_t) \rangle_{eq} = \sum_{\ell=1}^{\infty} g_\ell e^{-\lambda_\ell t}$$

with

$$g_\ell = \langle x, \phi_\ell \rangle_{f_\infty} \langle V'(x), \phi_\ell \rangle_{f_\infty},$$

with $\langle g, h \rangle_{f_\infty} = \int g(x) h(x) f_\infty(x) dx$. Consequently (recall that $\langle X_t \rangle_{eq} = 0$; furthermore, to ensure stationarity, we have set $t_0 = -\infty$),

$$\begin{aligned} \langle X_t \rangle &= \beta A_0 \int_{-\infty}^t \sum_{\ell=1}^{\infty} g_\ell e^{-\lambda_\ell(t-s)} \cos(\omega_0 s) ds \\ &= \frac{\beta A_0}{2} \sum_{\ell=1}^{\infty} g_\ell \operatorname{Re} \left(\frac{e^{i\omega_0 t}}{\lambda_\ell + i\omega_0} \right). \end{aligned}$$

We introduce now the *susceptibility*

$$\chi(\omega) = \chi'(\omega) - i\chi''(\omega) = \sum_{\ell=1}^{\infty} \frac{g_\ell}{\lambda_\ell + i\omega},$$

to rewrite

$$\langle X_t \rangle = \bar{x} \cos(\omega_0 t - \bar{\phi}) \quad (9.22)$$

with

$$\bar{x} = \beta A_0 |\chi(\omega_0)| \quad \text{and} \quad \bar{\phi} = \arctan \left(\frac{\chi''(\omega_0)}{\chi'(\omega_0)} \right). \quad (9.23)$$

The spectral amplification factor [see (7.42)] in the linear response approximation is

$$\eta = \beta^2 |\chi(\omega_0)|^2. \quad (9.24)$$

As expected, it is independent of the amplitude of the oscillations. It depends only on the spectrum of the generator of the unperturbed dynamics and the temperature.

9.2 The Fluctuation–Dissipation Theorem

In this section, we establish a connection between the response function (9.18) and stationary autocorrelation functions. Let X_t be a stationary Markov process in \mathbb{R}^d with generator \mathcal{L} and invariant distribution f_∞ , and let $A(\cdot)$ and $B(\cdot)$ be two observables. The stationary autocorrelation function $\langle A(X_t)B(X_0) \rangle_{eq}$ [see Eq. (2.60)] can be calculated as follows:

$$\begin{aligned}\kappa_{A,B}(t) &:= \langle A(X_t)B(X_0) \rangle_{eq} \\ &= \int \int A(x)B(x_0)p(x,t|x_0,0)f_\infty(x_0)dx dx_0 \\ &= \int \int A(x)B(x_0)e^{\mathcal{L}^*t}\delta(x-x_0)f_\infty(x_0)dx dx_0 \\ &= \int \int e^{\mathcal{L}t}A(x)B(x_0)\delta(x-x_0)f_\infty(x_0)dx dx_0 \\ &= \int e^{\mathcal{L}t}A(x)B(x)f_\infty(x)dx,\end{aligned}$$

where \mathcal{L} and \mathcal{L}^* act on functions of x . Thus we have established the formula

$$\kappa_{A,B}(t) = \langle S_t A(x), B(x) \rangle_{f_\infty}, \quad (9.25)$$

where $S_t = e^{\mathcal{L}t}$ denotes the semigroup generated by \mathcal{L} , and $\langle \cdot, \cdot \rangle_{f_\infty}$ denotes the $L^2(\mathbb{R}^d; f_\infty)$ -inner product.

Consider now the particular choice $B(x) = f_\infty^{-1} \mathcal{D} f_\infty$. We combine (9.18) and (9.25) to deduce

$$\kappa_{A, f_\infty^{-1} \mathcal{D} f_\infty}(t) = R_{\mathcal{L}_0, A}(t). \quad (9.26)$$

This is a version of the *fluctuation–dissipation theorem*, and it forms one of the cornerstones of nonequilibrium statistical mechanics. In particular, it enables us to calculate equilibrium correlation functions by measuring the response of the system to a weak external forcing.

Example 9.6. Consider the Langevin equation from Example 9.3 in one dimension with a constant external forcing:

$$dq = p dt, \quad dp = -V'(q) dt + \varepsilon F dt - \gamma p dt + \sqrt{2\gamma\beta^{-1}} dW_t.$$

We have $\mathcal{D} = -\partial_p$ and

$$B = f_\infty^{-1} \mathcal{D} f_\infty = \beta p.$$

We use (9.26) with $A = p$:

$$R_{\mathcal{L}_0, p}(t) = \beta \langle p(t)p(0) \rangle_{eq}.$$

When the potential is harmonic, $V(q) = \frac{1}{2}\omega_0^2 q^2$, we can compute explicitly the response function and, consequently, the velocity autocorrelation function at equilibrium:³

$$R_{\mathcal{L}_0, q}(t) = \frac{1}{\omega_1} e^{-\frac{\gamma t}{2}} \sin(\omega_1 t), \quad \omega_1 = \sqrt{\omega_0^2 - \frac{\gamma^2}{4}}$$

and

$$R_{\mathcal{L}_0, p}(t) = e^{-\frac{\gamma t}{2}} \left(\cos(\omega_1 t) - \frac{\gamma}{2\omega_1} \sin(\omega_1 t) \right).$$

Consequently,

$$\langle p(t)p(0) \rangle_{eq} = \beta^{-1} e^{-\frac{\gamma t}{2}} \left(\cos(\omega_1 t) - \frac{\gamma}{2\omega_1} \sin(\omega_1 t) \right).$$

Similar calculations can be done for more general linear stochastic differential equations. See Exercise 5.

Example 9.7. Consider again the Langevin dynamics with a perturbation in the temperature

$$dq = pdt, \quad dp = -V'(q)dt - \gamma p dt + \sqrt{2\gamma\beta^{-1}(1+\varepsilon F)}dW_t.$$

We have $\mathcal{D} = \gamma\beta^{-1}\partial_p^2$ and

$$B = f_\infty^{-1} \mathcal{D} f_\infty = \gamma\beta(p^2 - \beta^{-1}).$$

Let $H(p, q) = p^2/2 + V(q)$ denote the total energy. We have

$$f_\infty^{-1} \mathcal{L}_0^* H(p, q) f_\infty = \mathcal{L}_0 H(p, q) = \gamma(-p^2 + \beta^{-1}).$$

Consequently (see Exercise 6),

$$\kappa_{A, f_\infty^{-1} \mathcal{D} f_\infty}(t) = -\beta \frac{d}{dt} \kappa_{A, H}(t). \quad (9.27)$$

Setting now $A = H$, we obtain

$$R_{H, \mathcal{L}}(t) = -\beta \frac{d}{dt} \langle H(t)H(0) \rangle_{eq}.$$

9.3 The Green–Kubo Formula

Let us now calculate the long-time limit of $A(t)$ in (9.17) when the external forcing is a step function, Eq. (9.20). The following formal calculations can be justified in particular cases, for example for reversible diffusion processes, in which case

³ Note that this is the Green's function for the damped harmonic oscillator.

the generator of the process is a self-adjoint operator (see Sect. 4.6) and functional calculus can be used. As in previous sections, in this chapter, we assume that the diffusion process X_t with generator \mathcal{L} (to simplify the notation, we will use \mathcal{L} instead of \mathcal{L}_0) and Markov semigroup $S_t = e^{t\mathcal{L}}$ has a unique invariant measure. We will also assume that \mathcal{L} has a spectral gap. This is sufficient to justify the calculations that follow.

Assume that the observable $A(\cdot)$ is mean zero, $\int A(x)f_\infty(x)dx = 0$. We calculate

$$\begin{aligned} \int_0^t R_{\mathcal{L},A}(t-s)ds &= \int_0^t \int A(x)e^{\mathcal{L}^*(t-s)}\mathcal{D}f_\infty dx ds \\ &= \int \int_0^t \left(e^{\mathcal{L}(t-s)}A(x) \right) \mathcal{D}f_\infty ds dx \\ &= \int \left(e^{t\mathcal{L}} \int_0^t e^{(-s)\mathcal{L}} ds A(x) \right) \mathcal{D}f_\infty dx \\ &= \int \left(e^{t\mathcal{L}}(-\mathcal{L})^{-1} \left(e^{(-t)\mathcal{L}} - I \right) A(x) \right) \mathcal{D}f_\infty dx \\ &= \int \left((I - e^{t\mathcal{L}})(-\mathcal{L})^{-1} A(x) \right) \mathcal{D}f_\infty dx. \end{aligned} \quad (9.28)$$

Our assumptions on the generator \mathcal{L} imply that

$$\lim_{t \rightarrow +\infty} S_t h = \int h(x)f_\infty(x)dx.$$

Since the function $h = ((-\mathcal{L}_0)^{-1}A(x))\mathcal{D}f_\infty$ is mean zero, passing to the limit as $t \rightarrow +\infty$ in (9.28) yields

$$\Sigma := \lim_{t \rightarrow +\infty} \int_0^t R_{\mathcal{L},A}(t-s)ds = \int (-\mathcal{L})^{-1}A(x)\mathcal{D}f_\infty dx. \quad (9.29)$$

Using this in (9.20) and relabeling $\varepsilon F \mapsto F$, we deduce that

$$\lim_{F \rightarrow 0} \lim_{t \rightarrow +\infty} \frac{A(t)}{F} = \int (-\mathcal{L})^{-1}A(x)\mathcal{D}f_\infty dx. \quad (9.30)$$

Note that we can interchange the order in which we take the limits in (9.30). We will see later that formulas of the form (9.30) enable us to calculate *transport coefficients* such as the diffusion coefficient. We remark also that we can rewrite the above formula in the form

$$\lim_{F \rightarrow 0} \lim_{t \rightarrow +\infty} \frac{A(t)}{F} = \int \phi \mathcal{D}f_\infty dx,$$

where ϕ is the solution of the Poisson equation

$$-\mathcal{L}\phi = A(x), \quad (9.31)$$

equipped with appropriate boundary conditions. Note that since by assumption, $A(\cdot)$ is of mean zero, the right-hand side of the above equation satisfies the centering

condition of Fredholm's alternative. This equation is precisely the formalism that was used in Chap. 6 in the study of Brownian motion in periodic potentials:

Example 9.8. Consider the Langevin dynamics in a periodic or random potential:

$$dq_t = p_t dt, \quad dp_t = -\nabla V(q_t) dt - \gamma p_t dt + \sqrt{2\gamma\beta^{-1}} dW.$$

From Einstein's formula, we have that the diffusion coefficient is related to the mobility according to

$$D = \beta^{-1} \lim_{F \rightarrow 0} \lim_{t \rightarrow +\infty} \frac{\langle p_t \rangle}{F},$$

where we have used $\langle p_t \rangle_{eq} = 0$. We use now (9.29) with $A(p, q) = p$, $\mathcal{D} = -\nabla_p$, $f_\infty = \frac{1}{Z} e^{-\beta H(q, p)}$ to obtain

$$D = \int \int \phi p f_\infty dp dq = \langle -\mathcal{L}\phi, \phi \rangle_{f_\infty}, \quad (9.32)$$

which is precisely the formula obtained from homogenization theory.

Notice also that on combining (9.26) with (9.29), we obtain

$$\Sigma = \lim_{t \rightarrow +\infty} \int_0^t \kappa_{A, f_\infty^{-1} D f_\infty}(t-s) ds. \quad (9.33)$$

Thus, a transport coefficient such as the diffusion coefficient can be computed in terms of the time integral of an appropriate autocorrelation function. This is an example of the *Green–Kubo formula*.

We can obtain a more general form of the Green–Kubo formalism as follows. First, we define the generalized drift and diffusion coefficients as follows [compare with (2.45) and (2.46)]:

$$V^f(x) = \lim_{h \rightarrow 0} \frac{1}{h} \mathbb{E} \left(f(X_h) - f(X_0) \middle| X_0 = x \right) = \mathcal{L}f \quad (9.34)$$

and

$$\begin{aligned} D^{f,g}(x) &:= \lim_{h \rightarrow 0} \frac{1}{h} \mathbb{E} \left((f(X_{t+h}) - f(X_t))((g(X_{t+h}) - g(X_t)) \middle| X_t = x) \right) \\ &= \mathcal{L}(fg)(x) - (g\mathcal{L}f)(x) - (f\mathcal{L}g)(x), \end{aligned} \quad (9.35)$$

where f, g are C^2 functions.⁴ The equality in (9.34) follows from the definition of the generator of a diffusion process. For the equality in (9.35), see Exercise 3. We remark that $D^{f,g}(x)$ is the *opérateur carré du champ* that was introduced in Chap. 4; see Eq. (4.138). We have the following result.

⁴ In fact, all we need is $f, g \in D(\mathcal{L})$ and $fg \in D(\mathcal{L})$.

Result 9.1 (The Green–Kubo formula). *Let X_t be a stationary diffusion process with state space \mathbb{X} , generator \mathcal{L} , invariant measure $\mu(dx)$, and let $V^f(x)$, $D^{f,g}(x)$ given by (9.34) and (9.35), respectively. Then*

$$\frac{1}{2} \int D^{f,f} \mu(dx) = \int_0^\infty \mathbb{E} \left(V^f(X_t) V^f(X_0) \right) dt. \quad (9.36)$$

Proof. As earlier, we will use the notation $\langle \cdot, \cdot \rangle_\mu$ for the inner product in $L^2(\mathbb{X}, \mu)$. First, we note that

$$\frac{1}{2} \int D^{f,f} \mu(dx) = \langle -\mathcal{L}f, f \rangle_\mu := D_{\mathcal{L}}(f), \quad (9.37)$$

where $D_{\mathcal{L}}(f)$ is the Dirichlet form associated with the generator \mathcal{L} . In view of (9.37), formula (9.36) becomes

$$D_{\mathcal{L}}(f) = \int_0^\infty \mathbb{E} \left(V^f(X_t) V^f(X_0) \right) dt. \quad (9.38)$$

Now we use (9.25), together with the formula

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

which is valid for all mean-zero functions f , to obtain

$$\int_0^\infty \kappa_{A,B}(t) dt = \langle (-\mathcal{L})^{-1} A, B \rangle_\mu.$$

We set now $A = B = V^f = \mathcal{L}f$ in the above formula to obtain (9.38), from which (9.36) follows. \square

We remark that in the reversible case, we have (see Sect. 4.6)

$$\frac{1}{2} \int D^{f,g} \mu(dx) = \int_0^\infty \mathbb{E} \left(V^f(X_t) V^g(X_0) \right) dt. \quad (9.39)$$

Since the generator \mathcal{L} is a self-adjoint operator in $L^2(\mathbb{X}, \mu)$, the formal calculations presented in the above proof can be justified rigorously using functional calculus and spectral theory. See Exercise 7.

Example 9.9. Consider the reversible diffusion process (4.92) from Sect. 4.6:

$$dX_t = \mathbf{b}(X_t) dt + \boldsymbol{\sigma}(X_t) dW_t, \quad (9.40)$$

where the drift satisfies the detailed balance condition (4.98):

$$b(x) = \frac{1}{2} (\nabla \cdot A(x) - A(x) \nabla \Phi(x)),$$

with $A(x) = (\sigma\sigma^T)(x)$ and $\rho = e^{-\Phi}$ the invariant distribution. The generator of (4.92) is, see (4.99),

$$\mathcal{L} = \frac{1}{2} e^{\Phi} \nabla \cdot (A e^{-\Phi} \nabla).$$

Let $f = x_i$, $g = x_j$. We calculate

$$V^{x_i}(x) = \mathcal{L}x_i = \frac{1}{2} \frac{\partial A_{ik}}{\partial x_k} - \frac{1}{2} A_{ik} \frac{\partial \Phi}{\partial x_k}, \quad i = 1, \dots, d, \quad (9.41)$$

where we have used the summation convention. We use (9.35), together with the symmetry of \mathcal{L} in $L^2(\mathbb{R}^d; e^{-\Phi})$ and the previous calculation, to obtain, using again the summation convention,

$$\begin{aligned} \frac{1}{2} \int D^{x_i, x_j} \mu(dx) &= \langle -\mathcal{L}x_i, x_j \rangle_\mu \\ &= -\frac{1}{2} \int \left(\frac{\partial A_{ik}}{\partial x_k} - A_{ik} \frac{\partial \Phi}{\partial x_k} \right) x_j \rho(x) dx \\ &= \frac{1}{2} \int \left(A_{ik} \frac{\partial x_j}{\partial x_k} \rho - A_{ik} x_j \frac{\partial \rho}{\partial x_k} - A_{ik} \frac{\partial \Phi}{\partial x_k} x_j \rho \right) dx \\ &= \frac{1}{2} \int A_{ij}(x) \rho(x) dx. \end{aligned}$$

The Green–Kubo formula (9.36) gives

$$\frac{1}{2} \int A_{ij}(x) \rho(x) dx = \int_0^{+\infty} \mathbb{E} \left(V^{x_i}(X_t) V^{x_j}(X_0) \right) dt, \quad (9.42)$$

where the drift $V^{x_i}(x)$ is given by (9.41).

9.4 Discussion and Bibliography

Linear response theory and the fluctuation–dissipation theorem are in discussed in practically every book on nonequilibrium statistical mechanics, such as [12, 127, 160, 199, 246]. An earlier reference is [41]. An early review article is [126].

In Sect. 9.1, we considered stationary processes whose invariant densities have a smooth density with respect to Lebesgue measure. This excludes several interesting problems such as chaotic dynamical systems and stochastic PDEs. Linear response theory for deterministic dynamical systems is reviewed in [213] and for stochastic PDEs in [78]. Rigorous results on linear response theory and the fluctuation–dissipation theorem for Markov processes are presented in [42]. There is a very large literature on the mathematical justification of linear response theory, the fluctuation–dissipation theorem, and the Green–Kubo formula. Our approach to the Green–Kubo formula in Sect. 9.3, and in particular, Result 9.1, is based on [107, 244]. See also [106].

Formulas of the form (9.32) for the diffusion coefficient can be justified rigorously using either tools from stochastic analysis (the martingale central limit theorem) or the theory of partial differential equations (homogenization theory). The diffusion coefficient for reversible diffusions (together with the functional central limit theorem) is proved in [115]. Einstein's formula for the diffusion coefficient of a Brownian particle in a periodic potential is justified rigorously in [210].

Linear response theory and the fluctuation–dissipation theorem have a found a wide range of applications. Examples include climate modeling [136] and galactic dynamics [22, Chap. 5].

Linear response theory, the fluctuation–dissipation theorem, and Green–Kubo formulas are important topics in quantum nonequilibrium statistical mechanics. See, for example, [101, 102] and the references therein. See also [160].

9.5 Exercises

1. Let X_t be the solution of (9.8), and assume that we add a weak external forcing to both the drift and the noise. Write down the equation for the perturbed dynamics and the formulas for \mathcal{L}_1 and \mathcal{D} .
2. Calculate higher-order terms in the expansion (9.14). Use this to obtain higher-order terms in the calculation of expectation values of observables.
3. Let X_t be a stationary Markov process with state space \mathbb{X} , generator \mathcal{L} , and invariant measure μ , and let $f, g \in D(\mathcal{L})$ and $fg \in D(\mathcal{L})$. Show that

$$\lim_{h \rightarrow 0} \frac{1}{h} \mathbb{E} \left((f(X_h) - f(X_0))(g(X_h) - g(X_0)) \middle| X_0 = x \right) = \mathcal{L}(fg)(x) - (g\mathcal{L}f)(x) - (f\mathcal{L}g)(x). \quad (9.43)$$

4. Let $X_t \in \mathbb{R}^d$ be a dynamical system at equilibrium at $t = -\infty$, which is perturbed away from equilibrium by a weak external force $F(t)$. Let $A(x)$ be an observable, and consider the linear response relation

$$\Delta A(t) = \int_{\mathbb{R}} \gamma(s) F(t-s) ds, \quad (9.44)$$

where $\Delta A(t) = \langle A(X_t) \rangle - \langle A(X_t) \rangle_{eq}$. The *causality principle* implies that

$$\gamma(t) = 0, \quad \text{for } t < 0. \quad (9.45)$$

Assume that $\gamma(t) \in L^1(\mathbb{R})$.

- (a) Show that the linear response relation (9.44) can be written in the form

$$\widehat{\Delta A}(\omega) = \widehat{\gamma}(\omega) \widehat{F}(\omega), \quad (9.46)$$

where $\widehat{f}(\omega)$ denotes the Fourier transform of a function $f(t)$ (we assume that all Fourier transforms in (9.46) exist). The Fourier transform of the response function $\widehat{\gamma}(\omega)$ is called the *susceptibility*.

- (b) Show that the causality principle (9.45) implies that $\hat{\gamma}(\omega)$, $\omega \in \mathbb{C}$ is an analytic function in the complex upper half-plane.
- (c) Assume, furthermore, that $\lim_{|\omega| \rightarrow +\infty} \frac{1}{|\omega|} \gamma(\omega) = 0$. Apply Cauchy's integral theorem to the function

$$f(\omega) = \frac{\hat{\gamma}(\omega)}{\omega - \zeta},$$

where $\zeta \in \mathbb{R}$, and use the residue theorem to prove the *Kramers–Kronig relations*

$$\gamma_R(\zeta) = \frac{1}{\pi} \mathcal{P} \int_{\mathbb{R}} \frac{\gamma(\omega)}{\omega - \zeta} d\omega, \quad (9.47a)$$

$$\gamma_I(\zeta) = -\frac{1}{\pi} \mathcal{P} \int_{\mathbb{R}} \frac{\gamma_R(\omega)}{\omega - \zeta} d\omega, \quad (9.47b)$$

where $\hat{\gamma}(\omega) = \gamma_R(\omega) + i\gamma_I(\omega)$, and \mathcal{P} denotes the Cauchy principal value. (Hint: Integrate the function $f(\omega)$ along \mathbb{R} and a semicircle in the upper half-plane, avoiding the point $\zeta \in \mathbb{R}$ with a small semicircle of radius r in the upper half-plane).

- (d) Use the fact that $\gamma(t)$ is a real-valued function to obtain the alternative formulas

$$\gamma_R(\zeta) = \frac{2}{\pi} \mathcal{P} \int_0^\infty \frac{\omega \gamma_I(\omega)}{\omega^2 - \zeta^2} d\omega, \quad (9.48a)$$

$$\gamma_I(\zeta) = -\frac{1}{\pi} \mathcal{P} \int_{\mathbb{R}} \frac{\zeta \gamma_R(\omega)}{\omega^2 - \zeta^2} d\omega. \quad (9.48b)$$

More information about the Kramers–Kronig relations can be found in [41, Sect. VIII.3], [199, Sect. XI.1.2].

5. Let A and Σ be respectively strictly positive and positive $d \times d$ matrices, and consider the linear stochastic differential equation

$$dX_t = -AX_t dt + \sqrt{2\Sigma} dW_t. \quad (9.49)$$

- (a) Consider a weak external forcing. Calculate the response function. Use this to calculate the equilibrium autocorrelation matrix

$$\langle x(t) \otimes x(0) \rangle_{eg}.$$

- (b) Calculate the susceptibilities corresponding to the response functions $R_{\mathcal{L}_0, x_i}(t)$ (see Exercise 4).

- (c) Consider weak fluctuations in the diffusion matrix Σ . Calculate the response function and the equilibrium autocorrelation function of the (appropriately defined) energy.

6. Use (9.25) to prove (9.27).

7. Let \mathcal{L} be the generator of a reversible diffusion process. Use the spectral theorem for self-adjoint operators to provide a rigorous proof of (9.39).

Appendices

Appendix A

Frequently Used Notation

In the following, all summations are over indices from the set $\{1, 2, \dots, d\}$, d being the dimension of the space. We use \mathbb{R}^d to denote d -dimensional Euclidean space. We denote by $\langle \cdot, \cdot \rangle$ the standard inner product on \mathbb{R}^d . We also use \cdot to denote the inner product of two vectors, so that

$$\langle a, b \rangle = a \cdot b = \sum_i a_i b_i,$$

where $\{\xi_i\}_{i=1}^d$ are the components of a vector $\xi \in \mathbb{R}^d$ with respect to the standard basis $\{e_i\}_{i=1}^d$. The norm induced by this inner product is the Euclidean norm

$$|a| = \sqrt{a \cdot a},$$

and it follows that

$$|a|^2 = \sum_i a_i^2, \quad a \in \mathbb{R}^d.$$

The *inner product between matrices* is denoted by

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

The norm induced by this inner product is the *Frobenius norm*

$$|A|_F = \sqrt{\text{tr}(A^T A)}. \quad (\text{A.1})$$

Let ∇ and $\nabla \cdot$ denote the *gradient* and *divergence* in \mathbb{R}^d . The gradient lifts a scalar (respectively vector) to a vector (respectively matrix), while the divergence contracts a vector (respectively matrix) to a scalar (respectively vector). The gradient acts on scalar-valued functions ϕ , or vector-valued functions v , via

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

The divergence of a vector-valued function $v(x)$ is

$$\nabla \cdot v = \text{Tr}(\nabla v) = \sum_i \frac{\partial v_i}{\partial x_i}.$$

The divergence and gradient operators do not commute:

$$\nabla(\nabla \cdot v) = \nabla \cdot ((\nabla v)^T).$$

The divergence of a matrix-valued function $A(x)$ is the vector field defined as

$$\nabla \cdot A(x) \cdot a = \nabla \cdot (A^T(x)a),$$

for all constant vectors $a \in \mathbb{R}^d$. Componentwise,

$$(\nabla \cdot A)_i = \sum_j \frac{\partial A_{ij}}{\partial x_j}, \quad i = 1, \dots, d.$$

Given a vector-valued function $v(x)$ and a matrix-valued function $A(x)$, we have the following product rule:

$$\nabla \cdot (A^T v) = (\nabla \cdot A) \cdot v + A : \nabla v.$$

For two matrix-valued functions $A(x), B(x)$, we have the following product rule:

$$\nabla \cdot (AB) \cdot a = (\nabla \cdot B) \cdot Aa + B : \nabla(Aa). \quad (\text{A.2})$$

Given vector fields a, v , we use the notation

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

Thus we define this quantity by calculating $a \cdot \nabla v_k$ for each component of the vector v . Likewise, we define

$$a \cdot \nabla \Theta,$$

where Θ is a matrix field, using the above definition componentwise.

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

$$\nabla \nabla \phi$$

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

$$\nabla \nabla v$$

by applying $\nabla \nabla$ to each scalar component of the vector v , or indeed,

$$\nabla \nabla \Theta,$$

again componentwise. We define the Laplacian of a scalar or vector field by

$$\Delta \phi = \nabla \cdot \nabla \phi; \quad \Delta v = \nabla \cdot \nabla v.$$

It follows that $\Delta \phi = I : \nabla \nabla \phi$. Applying this definition componentwise allows for the definition of $\Delta \Theta$. We also use the following notation:

$$A : \nabla \nabla f = A : D^2 f = \text{Tr}(AD^2)f = \sum_{i,j} A_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j}.$$

Appendix B

Elements of Probability Theory

In this appendix we put together some basic definitions and results from probability theory. This is very standard material and can be found in all textbooks on probability theory and stochastic processes. In Sect. B.1, we give some basic definitions from the theory of probability. In Sect. B.2, we present some properties of random variables. In Sect. B.3, we introduce the concept of conditional expectation, and in Sect. B.4, we define the characteristic function. A few calculations with Gaussian measures in finite dimensions and in separable Hilbert spaces are presented in Sect. B.5. Different types of convergence and the basic limit theorems of the theory of probability are discussed in Sect. B.6. Discussion and bibliographical comments are presented in Sect. B.7.

B.1 Basic Definitions from Probability Theory

In order to study stochastic processes, we need to be able to describe the outcome of a random experiment and to calculate functions of this outcome. First, we need to describe the set of all possible experiments.

Definition B.1. The set of all possible outcomes of an experiment is called the *sample space* and is denoted by Ω .

We define events to be subsets of the sample space. Of course, we would like the unions, intersections, and complements of events also to be events. When the sample space Ω is uncountable, then technical difficulties arise. In particular, not all subsets of the sample space need to be events. A definition of the collection of subsets of events that is appropriate for finite additive probability is the following.

Definition B.2. A collection \mathcal{F} of Ω is called a field on Ω if

- (i) $\emptyset \in \mathcal{F}$;
- (ii) if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$;
- (iii) If $A, B \in \mathcal{F}$, then $A \cup B \in \mathcal{F}$.

From the definition of a field, we immediately deduce that \mathcal{F} is closed under finite unions and finite intersections:

$$A_1, \dots, A_n \in \mathcal{F} \Rightarrow \bigcup_{i=1}^n A_i \in \mathcal{F}, \quad \bigcap_{i=1}^n A_i \in \mathcal{F}.$$

When Ω is infinite-dimensional, the above definition is not appropriate, since we need to consider countable unions of events.

Definition B.3 (σ -algebra). A collection \mathcal{F} of Ω is called a σ -field or σ -algebra on Ω if

- (i) $\emptyset \in \mathcal{F}$;
- (ii) if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$;
- (iii) If $A_1, A_2, \dots \in \mathcal{F}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

A σ -algebra is closed under the operation of taking countable intersections. Standard examples of a σ -algebra are $\mathcal{F} = \{\emptyset, \Omega\}$, $\mathcal{F} = \{\emptyset, A, A^c, \Omega\}$, where A is a subset of Ω , and the power set of Ω , denoted by $\{0, 1\}^{\Omega}$, which contains all subsets of Ω .

Let now \mathcal{F} be a collection of subsets of Ω . It can be extended to a σ -algebra (take, for example, the power set of Ω). Consider all the σ -algebras that contain \mathcal{F} and take their intersection, denoted by $\sigma(\mathcal{F})$, i.e., $A \subset \Omega$ if and only if it is in every σ -algebra containing \mathcal{F} . It is a standard exercise to show that $\sigma(\mathcal{F})$ is a σ -algebra. It is the smallest algebra containing \mathcal{F} , and it is called the σ -algebra generated by \mathcal{F} .

Example B.1. Let $\Omega = \mathbb{R}^n$. The σ -algebra generated by the open subsets of \mathbb{R}^n (or equivalently, by the open balls of \mathbb{R}^n) is called the Borel σ -algebra of \mathbb{R}^n and is denoted by $\mathcal{B}(\mathbb{R}^n)$.

Let X be a closed subset of \mathbb{R}^n . Similarly, we can define the Borel σ -algebra of X , denoted by $\mathcal{B}(X)$. A sub- σ -algebra is a collection of subsets of a σ -algebra that satisfies the axioms of a σ -algebra. The σ -field \mathcal{F} of a sample space Ω contains all possible outcomes of the experiment that we want to study. Intuitively, the σ -field contains all the useful information that is available about the random experiment that we are performing.

Now we want to assign probabilities to the possible outcomes of an experiment.

Definition B.4 (Probability measure). A probability measure \mathbb{P} on the measurable space (Ω, \mathcal{F}) is a function $\mathbb{P} : \mathcal{F} \mapsto [0, 1]$ satisfying

- (i) $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(\Omega) = 1$;
- (ii) For A_1, A_2, \dots with $A_i \cap A_j = \emptyset$, $i \neq j$, then

$$\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

Definition B.5. The triple $(\Omega, \mathcal{F}, \mathbb{P})$ comprising a set Ω , a σ -algebra \mathcal{F} of subsets of Ω , and a probability measure \mathbb{P} on (Ω, \mathcal{F}) is called a probability space.

A standard example is $\Omega = [0, 1]$, $\mathcal{F} = \mathcal{B}([0, 1])$, $\mathbb{P} = \text{Leb}([0, 1])$. Then $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space.

B.2 Random Variables

We are usually interested in the consequences of the outcome of an experiment, rather than the experiment itself. The function of the outcome of an experiment is a *random variable*, that is, a map from Ω to \mathbb{R} .

Definition B.6. A sample space Ω equipped with a σ -field of subsets \mathcal{F} is called a measurable space.

Definition B.7. Let (Ω, \mathcal{F}) and (E, \mathcal{G}) be two measurable spaces. A function $X : \Omega \rightarrow E$ such that the *event*

$$\{\omega \in \Omega : X(\omega) \in A\} =: \{X \in A\} \quad (\text{B.1})$$

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

When E is \mathbb{R} equipped with its Borel σ -algebra, then (B.1) can be replaced with

$$\{X \leq x\} \in \mathcal{F} \quad \forall x \in \mathbb{R}.$$

Let X be a random variable (measurable function) from $(\Omega, \mathcal{F}, \mu)$ to (E, \mathcal{G}) . If E is a metric space, then we may define the *expectation* with respect to the measure μ by

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

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

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

Let U be a topological space. We will use the notation $\mathcal{B}(U)$ to denote the Borel σ -algebra of U , the smallest σ -algebra containing all open sets of U . Every random variable from a probability space $(\Omega, \mathcal{F}, \mu)$ to a measurable space $(E, \mathcal{B}(E))$ induces a probability measure on E :

$$\mu_X(B) = \mathbb{P}X^{-1}(B) = \mu(\omega \in \Omega; X(\omega) \in B), \quad B \in \mathcal{B}(E). \quad (\text{B.2})$$

The measure μ_X is called the *distribution* (or sometimes the *law*) of X .

Example B.2. Let \mathcal{I} denote a subset of the positive integers. A vector $\rho_0 = \{\rho_{0,i}, i \in \mathcal{I}\}$ is a distribution on \mathcal{I} if it has nonnegative entries and its total mass equals 1: $\sum_{i \in \mathcal{I}} \rho_{0,i} = 1$.

Consider the case $E = \mathbb{R}$ equipped with the Borel σ -algebra. In this case, a random variable is defined to be a function $X : \Omega \rightarrow \mathbb{R}$ such that

$$\{\omega \in \Omega : X(\omega) \leq x\} \subset \mathcal{F} \quad \forall x \in \mathbb{R}.$$

We can now define the probability distribution function $F_X : \mathbb{R} \rightarrow [0, 1]$ of X as

$$F_X(x) = \mathbb{P}(\{\omega \in \Omega | X(\omega) \leq x\}) =: \mathbb{P}(X \leq x). \quad (\text{B.3})$$

In this case, $(\mathbb{R}, \mathcal{B}(\mathbb{R}), F_X)$ becomes a probability space.

The distribution function $F_X(x)$ of a random variable has the properties $\lim_{x \rightarrow -\infty} F_X(x) = 0$, $\lim_{x \rightarrow +\infty} F(x) = 1$, and that it is right continuous.

Definition B.8. A random variable X with values on \mathbb{R} is called discrete if it takes values in some countable subset $\{x_0, x_1, x_2, \dots\}$ of \mathbb{R} . i.e., $\mathbb{P}(X = x) \neq 0$ only for $x = x_0, x_1, \dots$

With a random variable we can associate the probability mass function $p_k = \mathbb{P}(X = x_k)$. We will consider nonnegative integer-valued discrete random variables. In this case, $p_k = \mathbb{P}(X = k)$, $k = 0, 1, 2, \dots$.

Example B.3. The Poisson random variable is the nonnegative integer-valued random variable with probability mass function

$$p_k = \mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots,$$

where $\lambda > 0$.

Example B.4. The binomial random variable is the nonnegative integer-valued random variable with probability mass function

$$p_k = \mathbb{P}(X = k) = \frac{N!}{n!(N-n)!} p^n q^{N-n} \quad k = 0, 1, 2, \dots, N,$$

where $p \in (0, 1)$, $q = 1 - p$.

Definition B.9. A random variable X with values on \mathbb{R} is called continuous if $\mathbb{P}(X = x) = 0 \forall x \in \mathbb{R}$.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X : \Omega \rightarrow \mathbb{R}$ be a random variable with distribution F_X . This is a probability measure on $\mathcal{B}(\mathbb{R})$. We will assume that it is absolutely continuous with respect to the Lebesgue measure with density ρ_X : $F_X(dx) = \rho(x) dx$. We will call the density $\rho(x)$ the probability density function (PDF) of the random variable X .

Example B.5.

(i) The exponential random variable has PDF

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0, \\ 0 & x \leq 0, \end{cases}$$

with $\lambda > 0$.

(ii) The uniform random variable has PDF

$$f(x) = \begin{cases} \frac{1}{b-a} & a < x < b, \\ 0 & x \notin (a, b), \end{cases}$$

with $a < b$.

Definition B.10. Two random variables X and Y are independent if the events $\{\omega \in \Omega \mid X(\omega) \leq x\}$ and $\{\omega \in \Omega \mid Y(\omega) \leq y\}$ are independent for all $x, y \in \mathbb{R}$.

Let X, Y be two continuous random variables. We can view them as a random vector, i.e., a random variable from Ω to \mathbb{R}^2 . We can then define the joint distribution function

$$F(x, y) = \mathbb{P}(X \leq x, Y \leq y).$$

The mixed derivative of the distribution function $f_{X,Y}(x, y) := \frac{\partial^2 F}{\partial x \partial y}(x, y)$, if it exists, is called the joint PDF of the random vector $\{X, Y\}$:

$$f_{X,Y}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(x, y) dx dy.$$

If the random variables X and Y are independent, then

$$F_{X,Y}(x, y) = F_X(x)F_Y(y)$$

and

$$f_{X,Y}(x, y) = f_X(x)f_Y(y).$$

The joint distribution function has the properties

$$\begin{aligned} F_{X,Y}(x, y) &= F_{Y,X}(y, x), \\ F_{X,Y}(+\infty, y) &= F_Y(y), \quad f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dx. \end{aligned}$$

We can extend the above definition to random vectors of arbitrary finite dimensions. Let X be a random variable from $(\Omega, \mathcal{F}, \mu)$ to $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. The (joint) distribution function $F_X: \mathbb{R}^d \rightarrow [0, 1]$ is defined as

$$F_X(\mathbf{x}) = \mathbb{P}(X \leq \mathbf{x}).$$

Let X be a random variable in \mathbb{R}^d with distribution function $f(x_N)$, where $x_N = \{x_1, \dots, x_N\}$. We define the marginal or reduced distribution function $f^{N-1}(x_{N-1})$ by

$$f^{N-1}(x_{N-1}) = \int_{\mathbb{R}} f^N(x_N) dx_N.$$

We can define other reduced distribution functions:

$$f^{N-2}(x_{N-2}) = \int_{\mathbb{R}} f^{N-1}(x_{N-1}) dx_{N-1} = \int_{\mathbb{R}} \int_{\mathbb{R}} f(x_N) dx_{N-1} dx_N.$$

Expectation of Random Variables

We can use the distribution of a random variable to compute expectations and probabilities:

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}} f(x) dF_X(x) \quad (\text{B.4})$$

and

$$\mathbb{P}[X \in G] = \int_G dF_X(x), \quad G \in \mathcal{B}(E). \quad (\text{B.5})$$

The above formulas apply to both discrete and continuous random variables, provided that we define the integrals in (B.4) and (B.5) appropriately.

When $E = \mathbb{R}^d$ and a PDF exists, $dF_X(x) = f_X(x) dx$, we have

$$F_X(x) := \mathbb{P}(X \leq x) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} f_X(x) dx.$$

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

$$\|X\|_{L^p} = \left(\mathbb{E}|X|^p \right)^{1/p}.$$

Let X be a nonnegative integer-valued random variable with probability mass function p_k . We can compute the expectation of an arbitrary function of X using the formula

$$\mathbb{E}(f(X)) = \sum_{k=0}^{\infty} f(k)p_k.$$

Let X, Y be random variables. We define the covariance of the two random variables as

$$\text{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)] = \mathbb{E}(XY) - \mathbb{E}X\mathbb{E}Y.$$

The correlation coefficient is

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)}\sqrt{\text{var}(Y)}}. \quad (\text{B.6})$$

The Cauchy–Schwarz inequality yields that $\rho(X, Y) \in [-1, 1]$. We will say that two random variables X and Y are uncorrelated if $\rho(X, Y) = 0$. It is not true in general that two uncorrelated random variables are independent. This is true, however, for Gaussian random variables.

Example B.6.

- Consider the random variable $X : \Omega \mapsto \mathbb{R}$ with PDF

$$\gamma_{\sigma,b}(x) := (2\pi\sigma)^{-\frac{1}{2}} \exp\left(-\frac{(x-b)^2}{2\sigma}\right).$$

Such an X is termed a Gaussian or normal random variable. The mean is

$$\mathbb{E}X = \int_{\mathbb{R}} x\gamma_{\sigma,b}(x) dx = b,$$

and the variance is

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

- Let $b \in \mathbb{R}^d$ and $\Sigma \in \mathbb{R}^{d \times d}$ be symmetric and positive definite. The random variable $X : \Omega \mapsto \mathbb{R}^d$ with PDF

$$\gamma_{\Sigma,b}(x) := \left((2\pi)^d \det \Sigma\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \langle \Sigma^{-1}(x-b), (x-b) \rangle\right)$$

is termed a multivariate Gaussian or normal random variable. The mean is

$$\mathbb{E}(X) = b, \tag{B.7}$$

and the covariance matrix is

$$\mathbb{E}((X-b) \otimes (X-b)) = \Sigma. \tag{B.8}$$

Since the mean and variance specify completely a Gaussian random variable on \mathbb{R} , the Gaussian is commonly denoted by $\mathcal{N}(m, \sigma)$. The standard normal random variable is $\mathcal{N}(0, 1)$. Similarly, since the mean and covariance matrix completely specify a Gaussian random variable on \mathbb{R}^d , the Gaussian is commonly denoted by $\mathcal{N}(m, \Sigma)$.

Some analytical calculations for Gaussian random variables will be presented in Sect. B.5.

B.3 Conditional Expectation

One of the most important concepts in probability is that of the dependence between events.

Definition B.11. A family $\{A_i : i \in I\}$ of events is called independent if

$$\mathbb{P}(\cap_{j \in J} A_j) = \prod_{j \in J} \mathbb{P}(A_j)$$

for all finite subsets J of I .

When two events A, B are dependent, it is important to know the probability that the event A will occur, given that B has already occurred. We define this to be *conditional probability*, denoted by $\mathbb{P}(A|B)$. We know from elementary probability that

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

A very useful result is that of the *law of total probability*.

Definition B.12. A family of events $\{B_i : i \in I\}$ is called a partition of Ω if

$$B_i \cap B_j = \emptyset, \quad i \neq j \quad \text{and} \quad \bigcup_{i \in I} B_i = \Omega.$$

Proposition B.1 (Law of total probability). For every event A and partition $\{B_i : i \in I\}$, we have

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A|B_i) \mathbb{P}(B_i).$$

The proof of this result is left as an exercise. In many cases, the calculation of the probability of an event is simplified by choosing an appropriate partition of Ω and using the law of total probability.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and fix $B \in \mathcal{F}$. Then $\mathbb{P}(\cdot|B)$ defines a probability measure on \mathcal{F} . Indeed, we have that

$$\mathbb{P}(\emptyset|B) = 0, \quad \mathbb{P}(\Omega|B) = 1,$$

and (since $A_i \cap A_j = \emptyset$ implies that $(A_i \cap B) \cap (A_j \cap B) = \emptyset$)

$$P(\bigcup_{j=1}^{\infty} A_j|B) = \sum_{j=1}^{\infty} \mathbb{P}(A_j|B),$$

for a countable family of disjoint sets $\{A_j\}_{j=1}^{+\infty}$. Consequently, $(\Omega, \mathcal{F}, \mathbb{P}(\cdot|B))$ is a probability space for every $B \in \mathcal{F}$.

Assume that $X \in L^1(\Omega, \mathcal{F}, \mu)$, and let \mathcal{G} be a sub- σ -algebra of \mathcal{F} . The conditional expectation of X with respect to \mathcal{G} is defined to be the function (random variable) $\mathbb{E}[X|\mathcal{G}] : \Omega \mapsto E$, which is \mathcal{G} -measurable and satisfies

$$\int_G \mathbb{E}[X|\mathcal{G}] d\mu = \int_G X d\mu \quad \forall G \in \mathcal{G}.$$

We can define $\mathbb{E}[f(X)|\mathcal{G}]$ and the conditional probability $\mathbb{P}[X \in F|\mathcal{G}] = \mathbb{E}[I_F(X)|\mathcal{G}]$, where I_F is the indicator function of F , in a similar manner.

We list some of the most important properties of conditional expectation.

Proposition B.2 (Properties of conditional expectation). Let $(\Omega, \mathcal{F}, \mu)$ be a probability space, and let \mathcal{G} be a sub- σ -algebra of \mathcal{F} .

- (a) If X is \mathcal{G} -measurable and integrable, then $\mathbb{E}(X|\mathcal{G}) = X$.
- (b) (Linearity) If X_1, X_2 are integrable and c_1, c_2 constants, then

$$\mathbb{E}(c_1 X_1 + c_2 X_2|\mathcal{G}) = c_1 \mathbb{E}(X_1|\mathcal{G}) + c_2 \mathbb{E}(X_2|\mathcal{G}).$$

- (c) (Order) If X_1, X_2 are integrable and $X_1 \leq X_2$ a.s., then $\mathbb{E}(X_1|\mathcal{G}) \leq \mathbb{E}(X_2|\mathcal{G})$ a.s.
- (d) If Y and XY are integrable, and X is \mathcal{G} -measurable, then $\mathbb{E}(XY|\mathcal{G}) = X\mathbb{E}(Y|\mathcal{G})$.
- (e) (Successive smoothing) If \mathcal{D} is a sub- σ -algebra of \mathcal{F} , $\mathcal{D} \subset \mathcal{G}$ and X is integrable, then $\mathbb{E}(X|\mathcal{D}) = \mathbb{E}[\mathbb{E}(X|\mathcal{G})|\mathcal{D}] = \mathbb{E}[\mathbb{E}(X|\mathcal{D})|\mathcal{G}]$.
- (f) (Convergence) Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random variables such that for all n , $|X_n| \leq Z$, where Z is integrable. If $X_n \rightarrow X$ a.s., then $\mathbb{E}(X_n|\mathcal{G}) \rightarrow \mathbb{E}(X|\mathcal{G})$ a.s. and in L^1 .

Let $(\Omega, \mathcal{F}, \mu)$ be a probability space, X a random variable from $(\Omega, \mathcal{F}, \mu)$ to (E, \mathcal{G}) , and let $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}$. Then

$$\mathbb{E}(\mathbb{E}(X|\mathcal{F}_2)|\mathcal{F}_1) = \mathbb{E}(\mathbb{E}(X|\mathcal{F}_1)|\mathcal{F}_2) = \mathbb{E}(X|\mathcal{F}_1). \quad (\text{B.9})$$

Given $\mathcal{G} \subset \mathcal{F}$, we define the function $P_X(B|\mathcal{G}) = P(X \in B|\mathcal{G})$ for $B \in \mathcal{F}$. Assume that f is such that $\mathbb{E}f(X) < \infty$. Then

$$\mathbb{E}(f(X)|\mathcal{G}) = \int_{\mathbb{R}} f(x) P_X(dx|\mathcal{G}). \quad (\text{B.10})$$

B.4 The Characteristic Function

Many of the properties of (sums of) random variables can be studied using the Fourier transform of the distribution function. Let $F(\lambda)$ be the distribution function of a (discrete or continuous) random variable X . The characteristic function of X is defined to be the Fourier transform of the distribution function

$$\phi(t) = \int_{\mathbb{R}} e^{it\lambda} dF(\lambda) = \mathbb{E}(e^{itX}). \quad (\text{B.11})$$

For a continuous random variable for which the distribution function F has a density $dF(\lambda) = p(\lambda)d\lambda$, (B.11) gives

$$\phi(t) = \int_{\mathbb{R}} e^{it\lambda} p(\lambda) d\lambda.$$

For a discrete random variable for which $\mathbb{P}(X = \lambda_k) = a_k$, (B.11) gives

$$\phi(t) = \sum_{k=0}^{\infty} e^{it\lambda_k} a_k.$$

From the properties of the Fourier transform, we conclude that the characteristic function determines uniquely the distribution function of the random variable, in the sense that there is a one-to-one correspondence between $F(\lambda)$ and $\phi(t)$. Furthermore, we can prove the following two results.

Lemma B.3. Let $\{X_1, X_2, \dots, X_n\}$ be independent random variables with characteristic functions $\phi_j(t)$, $j = 1, \dots, n$, and let $Y = \sum_{j=1}^n X_j$ with characteristic function $\phi_Y(t)$. Then

$$\phi_Y(t) = \prod_{j=1}^n \phi_j(t).$$

Lemma B.4. Let X be a random variable with characteristic function $\phi(t)$, and assume that it has finite moments. Then

$$E(X^k) = \frac{1}{i^k} \phi^{(k)}(0).$$

B.5 Gaussian Random Variables

In this section, we present some useful calculations for Gaussian random variables. In particular, we calculate the normalization constant, the mean and variance, and the characteristic function of multidimensional Gaussian random variables.

Theorem B.1. Let $\mathbf{b} \in \mathbb{R}^d$ and $\Sigma \in \mathbb{R}^{d \times d}$ a symmetric and positive definite matrix. Let \mathbf{X} be the multivariate Gaussian random variable with probability density function

$$\gamma_{\Sigma, \mathbf{b}}(\mathbf{x}) = \frac{1}{Z} \exp \left(-\frac{1}{2} \langle \Sigma^{-1}(\mathbf{x} - \mathbf{b}), \mathbf{x} - \mathbf{b} \rangle \right).$$

Then:

(i) The normalization constant is

$$Z = (2\pi)^{d/2} \sqrt{\det(\Sigma)}.$$

(ii) The mean vector and covariance matrix of \mathbf{X} are given by

$$\mathbb{E}\mathbf{X} = \mathbf{b}$$

and

$$\mathbb{E}((\mathbf{X} - \mathbb{E}\mathbf{X}) \otimes (\mathbf{X} - \mathbb{E}\mathbf{X})) = \Sigma.$$

(iii) The characteristic function of \mathbf{X} is

$$\phi(\mathbf{t}) = e^{i\langle \mathbf{b}, \mathbf{t} \rangle - \frac{1}{2} \langle \mathbf{t}, \Sigma \mathbf{t} \rangle}.$$

Proof.

- (i) From the spectral theorem for symmetric positive definite matrices, we have that there exist a diagonal matrix Λ with positive entries and an orthogonal matrix B such that

$$\Sigma^{-1} = B^T \Lambda^{-1} B.$$

Let $\mathbf{z} = \mathbf{x} - \mathbf{b}$ and $\mathbf{y} = B\mathbf{z}$. We have

$$\begin{aligned}\langle \Sigma^{-1} \mathbf{z}, \mathbf{z} \rangle &= \langle B^T \Lambda^{-1} B \mathbf{z}, \mathbf{z} \rangle \\ &= \langle \Lambda^{-1} B \mathbf{z}, B \mathbf{z} \rangle = \langle \Lambda^{-1} \mathbf{y}, \mathbf{y} \rangle \\ &= \sum_{i=1}^d \lambda_i^{-1} y_i^2.\end{aligned}$$

Furthermore, we have that $\det(\Sigma^{-1}) = \prod_{i=1}^d \lambda_i^{-1}$, that $\det(\Sigma) = \prod_{i=1}^d \lambda_i$, and that the Jacobian of an orthogonal transformation is $J = \det(B) = 1$. Hence,

$$\begin{aligned}\int_{\mathbb{R}^d} \exp\left(-\frac{1}{2} \langle \Sigma^{-1}(\mathbf{x} - \mathbf{b}), \mathbf{x} - \mathbf{b} \rangle\right) d\mathbf{x} &= \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2} \langle \Sigma^{-1} \mathbf{z}, \mathbf{z} \rangle\right) d\mathbf{z} \\ &= \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2} \sum_{i=1}^d \lambda_i^{-1} y_i^2\right) |J| d\mathbf{y} \\ &= \prod_{i=1}^d \int_{\mathbb{R}} \exp\left(-\frac{1}{2} \lambda_i^{-1} y_i^2\right) dy_i \\ &= (2\pi)^{d/2} \prod_{i=1}^d \lambda_i^{1/2} = (2\pi)^{d/2} \sqrt{\det(\Sigma)},\end{aligned}$$

from which we get that

$$Z = (2\pi)^{d/2} \sqrt{\det(\Sigma)}.$$

In the above calculation, we have used the elementary calculus identity

$$\int_{\mathbb{R}} e^{-\alpha \frac{x^2}{2}} dx = \sqrt{\frac{2\pi}{\alpha}}.$$

(ii) From the above calculation, we have that

$$\begin{aligned}\gamma_{\Sigma, b}(\mathbf{x}) d\mathbf{x} &= \gamma_{\Sigma, b}(B^T \mathbf{y} + \mathbf{b}) d\mathbf{y} \\ &= \frac{1}{(2\pi)^{d/2} \sqrt{\det(\Sigma)}} \prod_{i=1}^d \exp\left(-\frac{1}{2} \lambda_i y_i^2\right) dy_i.\end{aligned}$$

Consequently,

$$\begin{aligned}\mathbb{E} \mathbf{X} &= \int_{\mathbb{R}^d} \mathbf{x} \gamma_{\Sigma, b}(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathbb{R}^d} (B^T \mathbf{y} + \mathbf{b}) \gamma_{\Sigma, b}(B^T \mathbf{y} + \mathbf{b}) d\mathbf{y} \\ &= \mathbf{b} \int_{\mathbb{R}^d} \gamma_{\Sigma, b}(B^T \mathbf{y} + \mathbf{b}) d\mathbf{y} = \mathbf{b}.\end{aligned}$$

We note that since $\Sigma^{-1} = B^T \Lambda^{-1} B$, we have that $\Sigma = B^T \Lambda B$. Furthermore, $\mathbf{z} = B^T \mathbf{y}$. We calculate

$$\begin{aligned}
\mathbb{E}((X_i - b_i)(X_j - b_j)) &= \int_{\mathbb{R}^d} z_i z_j \gamma_{\Sigma, b}(\mathbf{z} + \mathbf{b}) d\mathbf{z} \\
&= \frac{1}{(2\pi)^{d/2} \sqrt{\det(\Sigma)}} \int_{\mathbb{R}^d} \sum_k B_{ki} y_k \sum_m B_{mj} y_m \exp\left(-\frac{1}{2} \sum_\ell \lambda_\ell^{-1} y_\ell^2\right) d\mathbf{y} \\
&= \frac{1}{(2\pi)^{d/2} \sqrt{\det(\Sigma)}} \sum_{k,m} B_{ki} B_{mj} \int_{\mathbb{R}^d} y_k y_m \exp\left(-\frac{1}{2} \sum_\ell \lambda_\ell^{-1} y_\ell^2\right) d\mathbf{y} \\
&= \sum_{k,m} B_{ki} B_{mj} \lambda_k \delta_{km} \\
&= \Sigma_{ij}.
\end{aligned}$$

- (iii) Let \mathbf{y} be a multivariate Gaussian random variable with mean $\mathbf{0}$ and covariance I . Let also $C = B\sqrt{\Lambda}$. We have that $\Sigma = C^T C$. We have that

$$\mathbf{X} = C\mathbf{Y} + \mathbf{b}.$$

To see this, we first note that \mathbf{X} is Gaussian, since it is given through a linear transformation of a Gaussian random variable. Furthermore,

$$\mathbb{E}\mathbf{X} = \mathbf{b} \quad \text{and} \quad \mathbb{E}((X_i - b_i)(X_j - b_j)) = \Sigma_{ij}.$$

Now we have

$$\begin{aligned}
\phi(\mathbf{t}) &= \mathbb{E}e^{i\langle \mathbf{X}, \mathbf{t} \rangle} = e^{i\langle \mathbf{b}, \mathbf{t} \rangle} \mathbb{E}e^{i\langle C\mathbf{Y}, \mathbf{t} \rangle} \\
&= e^{i\langle \mathbf{b}, \mathbf{t} \rangle} \mathbb{E}e^{i\langle \mathbf{Y}, C^T \mathbf{t} \rangle} = e^{i\langle \mathbf{b}, \mathbf{t} \rangle} \mathbb{E}e^{i\sum_j (\sum_k C_{jk} t_k) y_j} \\
&= e^{i\langle \mathbf{b}, \mathbf{t} \rangle} e^{-\frac{1}{2} \sum_j |\sum_k C_{jk} t_k|^2} = e^{i\langle \mathbf{b}, \mathbf{t} \rangle} e^{-\frac{1}{2} \langle C\mathbf{t}, C\mathbf{t} \rangle} \\
&= e^{i\langle \mathbf{b}, \mathbf{t} \rangle} e^{-\frac{1}{2} \langle \mathbf{t}, C^T C \mathbf{t} \rangle} = e^{i\langle \mathbf{b}, \mathbf{t} \rangle} e^{-\frac{1}{2} \langle \mathbf{t}, \Sigma \mathbf{t} \rangle}.
\end{aligned}$$

Consequently,

$$\phi(\mathbf{t}) = e^{i\langle \mathbf{b}, \mathbf{t} \rangle - \frac{1}{2} \langle \mathbf{t}, \Sigma \mathbf{t} \rangle}.$$

□

B.5.1 Gaussian Measures in Hilbert Spaces

In the following, we let H be a separable Hilbert space, and we let $\mathcal{B}(H)$ be the Borel σ -algebra on H . We begin with the definition of a Gaussian measure.

Definition B.13. A probability measure μ on $(H, \mathcal{B}(H))$ is called Gaussian if for all $h \in H$, there exists $m \in \mathbb{R}$ such that

$$\mu(x \in H; \langle h, x \rangle \in A) = \mathcal{N}(A), A \in \mathcal{B}(\mathbb{R}). \quad (\text{B.12})$$

Let μ be a Gaussian measure. We define the following continuous functionals:

$$H \rightarrow \mathbb{R} \quad h \rightarrow \int_H \langle h, x \rangle \mu(dx), \quad (\text{B.13a})$$

$$H \times H \rightarrow \mathbb{R} \quad (h_1, h_2) \rightarrow \int_H \langle h_1, x \rangle \langle h_2, x \rangle \mu(dx). \quad (\text{B.13b})$$

The functional in (B.13b) is symmetric. We can use the Riesz representation theorem on H and $H \times H$ to obtain the following result.

Theorem B.2. *There exist $m \in H$ and a symmetric nonnegative continuous operator Q such that*

$$\int_H \langle h, x \rangle \mu(dx) = \langle m, h \rangle \quad \forall h \in H$$

and

$$\int_H \langle h_1, x \rangle \langle h_2, x \rangle \mu(dx) - \langle m, h_1 \rangle \langle m, h_2 \rangle = \langle Qh_1, h_2 \rangle \quad \forall h_1, h_2 \in H.$$

We will call m the mean and Q the covariance operator of the measure μ .

A Gaussian measure μ on H with mean m and covariance Q has the following characteristic function:

$$\mu(\lambda) = \int e^{i\langle \lambda \cdot x \rangle} \mu(dx) = e^{i\langle \lambda \cdot m \rangle - \frac{1}{2}\langle Q\lambda, \lambda \rangle}. \quad (\text{B.14})$$

Consequently, a Gaussian measure is uniquely determined by m and Q . Using the characteristic function of μ , one can prove that Q is a trace class operator.

Let now $\{e_k\}$ and $\{\lambda_k\}$ be the eigenfunctions and eigenvalues of Q , respectively. Since Q is symmetric and bounded, $\{e_k\}$ forms a complete orthonormal basis on H . Further, let $x_k = \langle x, e_k \rangle$, $k \in \mathbb{N}$. In the sequel, we will set $m = 0$.

Lemma B.5. *The random variables (x_1, \dots, x_n) are independent.*

Proof. We compute

$$\begin{aligned} \int_H x_i x_j \mu(dx) &= \int_H \langle x, e_i \rangle \langle x, e_j \rangle \mu(dx) \\ &= \langle Qe_i, e_j \rangle \\ &= \lambda_i \delta_{ij}. \end{aligned} \quad (\text{B.15})$$

□

Now we have the following:

Proposition B.6. *Let $\mu \in \mathcal{N}(0, Q)$ on H . Let*

$$S_Q = \inf_{\lambda \in \sigma(Q)} \frac{1}{2\lambda} = \frac{1}{2\|Q\|}, \quad (\text{B.16})$$

where $\sigma(Q)$ is the spectrum of Q . Then $\forall s \in [0, S_Q]$, we have

$$\begin{aligned} \int_H e^{s|x|^2} \mu(dx) &= \exp \left[-\frac{1}{2} \text{Tr}(\log(I - 2sQ)) \right] \\ &= \exp \left[\frac{1}{2} \sum_{k=1}^{\infty} \frac{(2s)^k}{k} \text{Tr}(Q^k) \right]. \end{aligned} \quad (\text{B.17})$$

Proof.

1. First, we observe that $S_Q > 0$. For $s \in [0, S_Q]$, we now have

$$\log(I - 2sQ) = \sum_{k=1}^{\infty} \frac{(2s)^k}{k}, \quad (\text{B.18})$$

the series being absolutely convergent in $L(H)$. Consequently, the operator $\log(I - 2sQ)$ is also trace class.

2. We fix $s \in [0, S_Q]$, and we consider finite-dimensional truncations of the integral that appear on the left-hand side of Eq. (B.17):

$$\begin{aligned} I_n &= \int_H e^{s \sum_{i=1}^n x_i^2} \mu(dx) \\ &= \prod_{i=1}^n \int_H e^{sx_i^2} \mu(dx) \quad (\{x_i\} \text{ are independent}) \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\lambda_i}} \int_{-\infty}^{\infty} e^{\left(s\xi^2 - \frac{x_i^2}{2\lambda_i}\right)} \mu(dx) \quad (x_i \in \mathcal{N}(0, \lambda_i)) \\ &= \prod_{i=1}^n \frac{1}{\sqrt{1 - 2\lambda_i s}} = e^{\left(-\frac{1}{2} \sum_{i=1}^n \log(1 - 2\lambda_i s)\right)} \\ &= e^{\left(-\frac{1}{2} \text{Tr} \log(I - 2sQ_n)\right)} \end{aligned} \quad (\text{B.19})$$

with

$$Q_n x = \sum_{i=1}^n \lambda_i \langle x, e_i \rangle e_i, \quad x \in H. \quad (\text{B.20})$$

Now we let $n \rightarrow \infty$ and use the fact that $\log(I - 2sQ_n)$ is trace class to obtain (B.17). \square

From the above proposition, we immediately obtain the following corollary:

Corollary B.1. *For arbitrary $p \in \mathbb{N}$, there exists a constant C_p such that*

$$\int_H |x|^{2p} \mu(dx) \leq C_p [\text{Tr}(Q)]^p \quad (\text{B.21})$$

for arbitrary $\mu \in \mathcal{N}(0, Q)$.

Proof. Differentiate Eq. (B.17) p times and set $s = 0$. \square

Now we make a few remarks on the above proposition and corollary. First, C_p is a combinatorial constant and grows in p . Moreover, we have

$$\int_H |x|^2 \mu(dx) = \text{Tr}(Q). \quad (\text{B.22})$$

Let now X be a Gaussian variable on H with distribution $\mu(dx)$. Then we have

$$\mathbb{E}|X|^{2p} = \int_H |x|^{2p} \mu(dx) \leq C_p (\mathbb{E}|X|^2)^p. \quad (\text{B.23})$$

We will use the notation $\mathbb{E}|X|^{2p} := \|X\|_{L^{2p}}^{2p}$. Let X_t be a stationary stochastic process on H with distribution $\mu(dx)$, $\mu \in \mathcal{N}(0, Q)$. Then using the above corollary, we can bound the L^{2p} norm of X_t :

$$\|X_t\|_{L^{2p}} \leq C_p \|X_t\|_{L^2}. \quad (\text{B.24})$$

B.6 Types of Convergence and Limit Theorems

One of the most important aspects of the theory of random variables is the study of limit theorems for sums of random variables. The best-known limit theorems in probability theory are the law of large numbers and the central limit theorem. There are various different types of convergence for sequences of random variables. We list the most important types of convergence below.

Definition B.14. Let $\{Z_n\}_{n=1}^\infty$ be a sequence of random variables. We will say that

(a) Z_n converges to Z with probability 1 if

$$\mathbb{P}\left(\lim_{n \rightarrow +\infty} Z_n = Z\right) = 1.$$

(b) Z_n converges to Z in probability if for every $\varepsilon > 0$,

$$\lim_{n \rightarrow +\infty} \mathbb{P}(|Z_n - Z| > \varepsilon) = 0.$$

(c) Z_n converges to Z in L^p if

$$\lim_{n \rightarrow +\infty} \mathbb{E}[|Z_n - Z|^p] = 0.$$

(d) Let $F_n(\lambda)$, $n = 1, \dots, +\infty$, $F(\lambda)$ be the distribution functions of Z_n , $n = 1, \dots, +\infty$, and Z , respectively. Then Z_n converges to Z in distribution if

$$\lim_{n \rightarrow +\infty} F_n(\lambda) = F(\lambda)$$

for all $\lambda \in \mathbb{R}$ at which F is continuous.

Recall that the distribution function F_X of a random variable from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to \mathbb{R} induces a probability measure on \mathbb{R} and that $(\mathbb{R}, \mathcal{B}(\mathbb{R}), F_X)$ is a probability space. We can show that the convergence in distribution is equivalent to the weak convergence of the probability measures induced by the distribution functions.

Definition B.15. Let (E, d) be a metric space, $\mathcal{B}(E)$ the σ -algebra of its Borel sets, P_n a sequence of probability measures on $(E, \mathcal{B}(E))$, and let $C_b(E)$ denote the space of bounded continuous functions on E . We will say that the sequence P_n converges weakly to the probability measure P if for each $f \in C_b(E)$,

$$\lim_{n \rightarrow +\infty} \int_E f(x) dP_n(x) = \int_E f(x) dP(x).$$

Theorem B.3. Let $F_n(\lambda)$, $n = 1, \dots, +\infty$, $F(\lambda)$ be the distribution functions of Z_n , $n = 1, \dots, +\infty$, and Z , respectively. Then Z_n converges to Z in distribution if and only if for all $g \in C_b(\mathbb{R})$,

$$\lim_{n \rightarrow +\infty} \int_X g(x) dF_n(x) = \int_X g(x) dF(x). \quad (\text{B.25})$$

Note that (B.25) is equivalent to

$$\lim_{n \rightarrow +\infty} \mathbb{E}_n g(X_n) = \mathbb{E} g(X),$$

where E_n and E denote the expectations with respect to F_n and F , respectively.

When the sequence of random variables in whose convergence we are interested takes values in \mathbb{R}^d , or more generally, a metric space (E, d) , then we can use weak convergence of the sequence of probability measures induced by the sequence of random variables to define convergence in distribution.

Definition B.16. A sequence of real-valued random variables X_n defined on a probability spaces $(\Omega_n, \mathcal{F}_n, P_n)$ and taking values on a metric space (E, d) is said to converge in distribution if the induced measures $F_n(B) = P_n(X_n \in B)$ for $B \in \mathcal{B}(E)$ converge weakly to a probability measure P .

Let $\{X_n\}_{n=1}^{\infty}$ be iid random variables with $\mathbb{E}X_n = V$. Then the *strong law of large numbers* states that the average of the sum of the iid random variables converges to V with probability 1:

$$\mathbb{P}\left(\lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{n=1}^N X_n = V\right) = 1. \quad (\text{B.26})$$

The strong law of large numbers provides us with information about the behavior of a sum of random variables (or a large number of repetitions of the same experiment) on average. We can also study fluctuations around the average behavior. Indeed, let $\mathbb{E}(X_n - V)^2 = \sigma^2$. Define the centered iid random variables $Y_n = X_n - V$. Then the sequence of random variables $\frac{1}{\sigma\sqrt{N}} \sum_{n=1}^N Y_n$ converges in distribution to a $\mathcal{N}(0, 1)$ random variable:

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(\frac{1}{\sigma \sqrt{N}} \sum_{n=1}^N Y_n \leq a \right) = \int_{-\infty}^a \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx.$$

This is the *central limit theorem*.

A useful result is Slutsky's theorem.

Theorem B.4 (Slutsky). *Let $\{X_n\}_{n=1}^{+\infty}$, $\{Y_n\}_{n=1}^{+\infty}$ be sequences of random variables such that X_n converges in distribution to a random variable X , and Y_n converges in probability to a constant $c \neq 0$. Then*

$$\lim_{n \rightarrow +\infty} Y_n^{-1} X_n = c^{-1} X,$$

in distribution.

B.7 Discussion and Bibliography

The material of this appendix is very standard and can be found in many books on probability theory and stochastic processes. See, for example, [29, 56, 57, 119, 144, 145, 227].

The connection between conditional expectation and orthogonal projections is discussed in [36].

The reduced distribution functions defined in Sect. B.2 are used extensively in statistical mechanics. A different normalization is usually used in physics textbooks. See, for instance, [13, Sect. 4.2].

The calculations presented in Sect. B.5 are essentially an exercise in linear algebra. See [135, Sect. 10.2]. Section B.5 is based on [190, Sect. 2.3], where additional information on probability measures in infinite-dimensional spaces can be found.

Limit theorems for stochastic processes are studied in detail in [100].

References

- [1] R. J. Adler. *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes*, volume 12 of *Institute of Mathematical Statistics Lecture Notes—Monograph Series*. Institute of Mathematical Statistics, Hayward, CA, 1990.
- [2] Y. Aït-Sahalia. Closed-form likelihood expansions for multivariate diffusions. *Ann. Statist.*, 36(2):906–937, 2008.
- [3] G. Ariel and E. Vanden-Eijnden. Testing transition state theory on Kac-Zwanzig model. *J. Stat. Phys.*, 126(1):43–73, 2007.
- [4] G. Ariel and E. Vanden-Eijnden. A strong limit theorem in the Kac-Zwanzig model. *Nonlinearity*, 22(1):145–162, 2009.
- [5] A. Arnold, J. A. Carrillo, L. Desvillettes, J. Dolbeault, A. Jüngel, C. Lederer, P. A. Markowich, G. Toscani, and C. Villani. Entropies and equilibria of many-particle systems: an essay on recent research. *Monatsh. Math.*, 142(1–2):35–43, 2004.
- [6] A. Arnold, P. Markowich, G. Toscani, and A. Unterreiter. On convex Sobolev inequalities and the rate of convergence to equilibrium for Fokker-Planck type equations. *Comm. Partial Differential Equations*, 26(1–2):43–100, 2001.
- [7] L. Arnold. *Stochastic differential equations: theory and applications*. Wiley-Interscience [John Wiley & Sons], New York, 1974. Translated from the German.
- [8] S. Asmussen and P. W. Glynn. *Stochastic simulation: algorithms and analysis*, volume 57 of *Stochastic Modelling and Applied Probability*. Springer, New York, 2007.
- [9] A. Athreya, T. Kolba, and J. C. Mattingly. Propagating Lyapunov functions to prove noise-induced stabilization. *Electron. J. Probab.*, 17:no. 96, 38, 2012.
- [10] D. Bakry, P. Cattiaux, and A. Guillin. Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré. *J. Funct. Anal.*, 254(3):727–759, 2008.

- [11] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Cham, 2014.
- [12] R. Balescu. *Equilibrium and nonequilibrium statistical mechanics*. Wiley-Interscience [John Wiley & Sons], New York, 1975.
- [13] R. Balescu. *Statistical dynamics. Matter out of equilibrium*. Imperial College Press, London, 1997.
- [14] R. Bartussek, P. Reimann, and P. Hanggi. Precise numerics versus theory for correlation ratchets. *Phys. Rev. Lett.*, 76(7):1166–1169, 1996.
- [15] I.V. Basawa and B.L.S. Prakasa Rao. *Statistical inference for stochastic processes*. Academic Press Inc., London, 1980.
- [16] R. F. Bass, B. M. Hambly, and T. J. Lyons. Extending the Wong-Zakai theorem to reversible Markov processes. *J. Eur. Math. Soc. (JEMS)*, 4(3):237–269, 2002.
- [17] A. Bensoussan, J.-L. Lions, and G. Papanicolaou. *Asymptotic analysis for periodic structures*, volume 5 of *Studies in Mathematics and Its Applications*. North-Holland Publishing Co., Amsterdam, 1978.
- [18] R. Benzi, G. Parisi, A. Sutera, and A. Vulpiani. A theory of stochastic resonance in climatic change. *SIAM J. Appl. Math.*, 43(3):565–478, 1983.
- [19] R. Benzi, A. Sutera, and A. Vulpiani. The mechanism of stochastic resonance. *J. Phys. A*, 14(11):L453–L457, 1981.
- [20] N. Berglund and B. Gentz. A sample-paths approach to noise-induced synchronization: stochastic resonance in a double-well potential. *Ann. Appl. Probab.*, 12(4):1419–1470, 2002.
- [21] N. Berglund and B. Gentz. *Noise-induced phenomena in slow-fast dynamical systems*. Probability and Its Applications (New York). Springer-Verlag London Ltd., London, 2006. A sample-paths approach.
- [22] J. Binney and S. Tremaine. *Galactic Dynamics*. Princeton University Press, Princeton, second edition, 2008.
- [23] J. P. N. Bishwal. *Parameter estimation in stochastic differential equations*, volume 1923 of *Lecture Notes in Mathematics*. Springer, Berlin, 2008.
- [24] G. Blankenship and G.C. Papanicolaou. Stability and control of stochastic systems with wide-band noise disturbances. I. *SIAM J. Appl. Math.*, 34(3):437–476, 1978.
- [25] N. Bleistein and R. A. Handelsman. *Asymptotic expansions of integrals*. Dover Publications Inc., New York, second edition, 1986.
- [26] D. Blömker, M. Hairer, and G. A. Pavliotis. Multiscale analysis for stochastic partial differential equations with quadratic nonlinearities. *Nonlinearity*, 20(7):1721–1744, 2007.
- [27] D. Blömker, S. Maier-Paape, and G. Schneider. The stochastic Landau equation as an amplitude equation. *Discrete Contin. Dyn. Syst. Ser. B*, 1(4):527–541, 2001.

- [28] K. Boďová and C. R. Doering. Noise-induced statistically stable oscillations in a deterministically divergent nonlinear dynamical system. *Commun. Math. Sci.*, 10(1):137–157, 2012.
- [29] L. Breiman. *Probability*, volume 7 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992. Corrected reprint of the 1968 original.
- [30] T. Brettschneider, G. Volpe, L. Helden, J. Wehr, and C. Bechinger. Force measurement in the presence of Brownian noise: Equilibrium-distribution method versus drift method. *Phys. Rev. E*, 83(4, Part 1), APR 15 2011.
- [31] C. Bustamante, D. Keller, and G. Oster. The physics of molecular motors. *Acc. Chem. res.*, 34:412–420, 2001.
- [32] E. A. Carlen. Conservative diffusions. *Comm. Math. Phys.*, 94(3):293–315, 1984.
- [33] S. Cerrai and M. Freidlin. On the Smoluchowski-Kramers approximation for a system with an infinite number of degrees of freedom. *Probab. Theory Related Fields*, 135(3):363–394, 2006.
- [34] S. Cerrai and M. Freidlin. Smoluchowski-Kramers approximation for a general class of SPDEs. *J. Evol. Equ.*, 6(4):657–689, 2006.
- [35] S. Chandrasekhar. Stochastic problems in physics and astronomy. *Rev. Mod. Phys.*, 15(1):1–89, Jan 1943.
- [36] A. J. Chorin and O. H. Hald. *Stochastic tools in mathematics and science*, volume 1 of *Surveys and Tutorials in the Applied Mathematical Sciences*. Springer, New York, 2006.
- [37] P. Collet and S. Martínez. Asymptotic velocity of one dimensional diffusions with periodic drift. *J. Math. Biol.*, 56(6):765–792, 2008.
- [38] E. Cortes, B. J. West, and K. Lindenberg. On the generalized Langevin equation—classical and quantum-mechanical. *Journal of Chemical Physics*, 82(6):2708–2717, 1985.
- [39] M. C. Cross and P. C. Hohenberg. Pattern formation outside of equilibrium. *Rev. Mod. Phys.*, 65:851–1112, Jul 1993.
- [40] R. I. Cukier, K. Lakatos-Lindenberg, and K. E. Shuler. Orthogonal polynomial solutions of the Fokker-Planck equation. *J. Stat. Phys.*, 9(2):137–144, 1973.
- [41] S. R. de Groot and P. Mazur. *Non-equilibrium thermodynamics*. Interscience, New York, 1962.
- [42] A. Dembo and J.-D. Deuschel. Markovian perturbation, response and fluctuation dissipation theorem. *Ann. Inst. Henri Poincaré Probab. Stat.*, 46(3):822–852, 2010.
- [43] L. Desvillettes and C. Villani. On the trend to global equilibrium in spatially inhomogeneous entropy-dissipating systems: the linear Fokker-Planck equation. *Comm. Pure Appl. Math.*, 54(1):1–42, 2001.
- [44] L. Desvillettes and C. Villani. On the trend to global equilibrium for spatially inhomogeneous kinetic systems: the Boltzmann equation. *Invent. Math.*, 159(2):245–316, 2005.

- [45] P. Diaconis. The Markov chain Monte Carlo revolution. *Bull. Amer. Math. Soc. (N.S.)*, 46(2):179–205, 2009.
- [46] C. R. Doering, W. Horsthemke, and J. Riordan. Nonequilibrium fluctuation-induced transport. *Phys. Rev. Lett.*, 72(19):2984–2987, 1994.
- [47] D. Dürr, S. Goldstein, and J. L. Lebowitz. A mechanical model of Brownian motion. *Comm. Math. Phys.*, 78(4):507–530, 1980/81.
- [48] J.-P. Eckmann and M. Hairer. Non-equilibrium statistical mechanics of strongly anharmonic chains of oscillators. *Comm. Math. Phys.*, 212(1):105–164, 2000.
- [49] J.-P. Eckmann and M. Hairer. Spectral properties of hypoelliptic operators. *Comm. Math. Phys.*, 235(2):233–253, 2003.
- [50] J.-P. Eckmann, C.-A. Pillet, and L. Rey-Bellet. Entropy production in nonlinear, thermally driven Hamiltonian systems. *J. Statist. Phys.*, 95(1–2):305–331, 1999.
- [51] J.-P. Eckmann, C.-A. Pillet, and L. Rey-Bellet. Non-equilibrium statistical mechanics of anharmonic chains coupled to two heat baths at different temperatures. *Comm. Math. Phys.*, 201(3):657–697, 1999.
- [52] A. Einstein. *Investigations on the theory of the Brownian movement*. Dover Publications Inc., New York, 1956. Edited with notes by R. Fürth, Translated by A. D. Cowper.
- [53] S. N. Ethier and T. G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986.
- [54] L. C. Evans. *An introduction to stochastic differential equations*. American Mathematical Society, Providence, RI, 2013.
- [55] L. C. Evans. *Partial Differential Equations*. AMS, Providence, Rhode Island, 1998.
- [56] W. Feller. *An introduction to probability theory and its applications. Vol. I*. Third edition. John Wiley & Sons Inc., New York, 1968.
- [57] W. Feller. *An introduction to probability theory and its applications. Vol. II*. Second edition. John Wiley & Sons Inc., New York, 1971.
- [58] J. C. M. Fok, B. Guo, and T. Tang. Combined Hermite spectral-finite difference method for the Fokker-Planck equation. *Math. Comp.*, 71(240):1497–1528 (electronic), 2002.
- [59] G. W. Ford and M. Kac. On the quantum Langevin equation. *J. Statist. Phys.*, 46(5–6):803–810, 1987.
- [60] G. W. Ford, M. Kac, and P. Mazur. Statistical mechanics of assemblies of coupled oscillators. *J. Mathematical Phys.*, 6:504–515, 1965.
- [61] M. I. Freidlin and A. D. Wentzell. Random perturbations of Hamiltonian systems. *Mem. Amer. Math. Soc.*, 109(523):viii+82, 1994.
- [62] M. I. Freidlin and A. D. Wentzell. *Random perturbations of dynamical systems*, volume 260 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Heidelberg, third edition, 2012. Translated from the 1979 Russian original by Joseph Szücs.
- [63] A. Friedman. *Partial differential equations of parabolic type*. Prentice-Hall Inc., Englewood Cliffs, N.J., 1964.

- [64] A. Friedman. *Stochastic differential equations and applications. Vol. 1.* Academic Press, New York, 1975. Probability and Mathematical Statistics, Vol. 28.
- [65] A. Friedman. *Stochastic differential equations and applications. Vol. 2.* Academic Press, New York, 1976. Probability and Mathematical Statistics, Vol. 28.
- [66] L. Gammaitoni, P. Hanggi, P. Jung, and F. Marchesoni. Stochastic resonance. *Reviews of Modern Physics*, 70(1):223–287, 1998.
- [67] T. C. Gard. *Introduction to stochastic differential equations*, volume 114 of *Monographs and Textbooks in Pure and Applied Mathematics*. Marcel Dekker Inc., New York, 1988.
- [68] C. W. Gardiner. *Handbook of stochastic methods*. Springer-Verlag, Berlin, second edition, 1985. For physics, chemistry and the natural sciences.
- [69] R. G. Ghanem and P. D. Spanos. *Stochastic finite elements: a spectral approach*. Springer-Verlag, New York, 1991.
- [70] I. I. Gikhman and A. V. Skorokhod. *Introduction to the theory of random processes*. Dover Publications Inc., Mineola, NY, 1996.
- [71] D. Givon, R. Kupferman, and A. M. Stuart. Extracting macroscopic dynamics: model problems and algorithms. *Nonlinearity*, 17(6):R55–R127, 2004.
- [72] J. Glimm and A. Jaffe. *Quantum physics*. Springer-Verlag, New York, second edition, 1987. A functional integral point of view.
- [73] T. Goudon and F. Poupaud. Homogenization of transport equations: weak mean field approximation. *SIAM J. Math. Anal.*, 36(3):856–881 (electronic), 2004/05.
- [74] H. Grad. Asymptotic theory of the Boltzmann equation. *Phys. Fluids*, 6: 147–181, 1963.
- [75] G. R. Grimmett and D. R. Stirzaker. *Probability and random processes*. Oxford University Press, New York, 2001.
- [76] L. Gross. Logarithmic Sobolev inequalities. *Amer. J. Math.*, 97(4): 1061–1083, 1975.
- [77] A. Guionnet and B. Zegarlinski. Lectures on logarithmic Sobolev inequalities. In *Séminaire de Probabilités, XXXVI*, volume 1801 of *Lecture Notes in Math.*, pages 1–134. Springer, Berlin, 2003.
- [78] M. Hairer and A. J. Majda. A simple framework to justify linear response theory. *Nonlinearity*, 23(4):909–922, 2010.
- [79] M. Hairer and G. A. Pavliotis. From ballistic to diffusive behavior in periodic potentials. *J. Stat. Phys.*, 131(1):175–202, 2008.
- [80] P. Hanggi. Escape from a metastable state. *J. Stat. Phys.*, 42(1/2):105–140, 1986.
- [81] P. Hanggi and P. Jung. Colored noise in dynamical systems. In I. Prigogine and S. A. Rice, editors, *Advances in Chemical Physics*, volume 89, 239–326. 1995.
- [82] P. Hanggi, P. Talkner, and M. Borkovec. Reaction-rate theory: fifty years after Kramers. *Rev. Modern Phys.*, 62(2):251–341, 1990.

- [83] R. Z. Has'minskii. *Stochastic stability of differential equations*, volume 7 of *Monographs and Textbooks on Mechanics of Solids and Fluids: Mechanics and Analysis*. Sijthoff & Noordhoff, Alphen aan den Rijn, 1980.
- [84] U. G. Haussmann and É. Pardoux. Time reversal of diffusions. *Ann. Probab.*, 14(4):1188–1205, 1986.
- [85] B. Helffer and F. Nier. *Hypoelliptic estimates and spectral theory for Fokker-Planck operators and Witten Laplacians*, volume 1862 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2005.
- [86] F. Hérau and F. Nier. Isotropic hypoellipticity and trend to equilibrium for the Fokker-Planck equation with a high-degree potential. *Arch. Ration. Mech. Anal.*, 171(2):151–218, 2004.
- [87] S. Herrmann, P. Imkeller, and I. Pavlyukevich. Two mathematical approaches to stochastic resonance. In *Interacting stochastic systems*, pages 327–351. Springer, Berlin, 2005.
- [88] D. J. Higham. Mean-square and asymptotic stability of the stochastic theta method. *SIAM J. Numer. Anal.*, 38(3):753–769 (electronic), 2000.
- [89] D. J. Higham. An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Rev.*, 43(3):525–546 (electronic), 2001.
- [90] R. A. Holley, S. Kusuoka, and D. W. Stroock. Asymptotics of the spectral gap with applications to the theory of simulated annealing. *J. Funct. Anal.*, 83(2):333–347, 1989.
- [91] M. H. Holmes. *Introduction to perturbation methods*. Springer-Verlag, New York, 1995.
- [92] L. Hörmander. Hypoelliptic second order differential equations. *Acta Math.*, 119:147–171, 1967.
- [93] W. Horsthemke and R. Lefever. *Noise-induced transitions*, volume 15 of *Springer Series in Synergetics*. Springer-Verlag, Berlin, 1984. Theory and applications in physics, chemistry, and biology.
- [94] S. Hottovy, G. Volpe, and J. Wehr. Noise-Induced Drift in Stochastic Differential Equations with Arbitrary Friction and Diffusion in the Smoluchowski-Kramers Limit. *J. Stat. Phys.*, 146(4):762–773, 2012.
- [95] C.-R. Hwang, S.-Y. Hwang-Ma, and S.-J. Sheu. Accelerating Gaussian diffusions. *Ann. Appl. Probab.*, 3(3):897–913, 1993.
- [96] C.-R. Hwang, S.-Y. Hwang-Ma, and S.-J. Sheu. Accelerating diffusions. *Ann. Appl. Probab.*, 15(2):1433–1444, 2005.
- [97] S. M. Iacus. *Simulation and inference for stochastic differential equations*. Springer Series in Statistics. Springer, New York, 2008.
- [98] N. Ikeda and S. Watanabe. *Stochastic differential equations and diffusion processes*. North-Holland Publishing Co., Amsterdam, second edition, 1989.
- [99] H. M. Ito. Ergodicity of randomly perturbed Lorenz model. *J. Statist. Phys.*, 35(1–2):151–158, 1984.
- [100] J. Jacod and A. N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2003.

- [101] V. Jakšić, Y. Ogata, and C.-A. Pillet. The Green-Kubo formula and the Onsager reciprocity relations in quantum statistical mechanics. *Comm. Math. Phys.*, 265(3):721–738, 2006.
- [102] V. Jakšić, Y. Ogata, and C.-A. Pillet. Linear response theory for thermally driven quantum open systems. *J. Stat. Phys.*, 123(3):547–569, 2006.
- [103] V. Jakšić and C.-A. Pillet. Ergodic properties of the non-Markovian Langevin equation. *Lett. Math. Phys.*, 41(1):49–57, 1997.
- [104] V. Jakšić and C.-A. Pillet. Spectral theory of thermal relaxation. *J. Math. Phys.*, 38(4):1757–1780, 1997. Quantum problems in condensed matter physics.
- [105] V. Jakšić and C.-A. Pillet. Ergodic properties of classical dissipative systems. I. *Acta Math.*, 181(2):245–282, 1998.
- [106] D.-Q. Jiang, M. Qian, and M.-P. Qian. *Mathematical theory of nonequilibrium steady states*, volume 1833 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2004.
- [107] D.-Q. Jiang and F.-X. Zhang. The Green-Kubo formula and power spectrum of reversible Markov processes. *J. Math. Phys.*, 44(10):4681–4689, 2003.
- [108] V. V. Jikov, S. M. Kozlov, and O. A. Oleinik. *Homogenization of differential operators and integral functionals*. Springer-Verlag, Berlin, 1994.
- [109] F. John. *Partial differential equations*, volume 1 of *Applied Mathematical Sciences*. Springer-Verlag, New York, fourth edition, 1991.
- [110] P. Jung and P. Hanggi. Stochastic nonlinear dynamics modulated by external periodic forces. *Europhysics Letters*, 8(6):505–510, 1989.
- [111] I. Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1991.
- [112] S. Karlin and H. M. Taylor. *A first course in stochastic processes*. Academic Press, New York-London, 1975.
- [113] S. Karlin and H. M. Taylor. *A second course in stochastic processes*. Academic Press Inc., New York, 1981.
- [114] J. Kevorkian and J. D. Cole. *Multiple scale and singular perturbation methods*, volume 114 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1996.
- [115] C. Kipnis and S. R. S. Varadhan. Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Comm. Math. Phys.*, 104(1):1–19, 1986.
- [116] Y. L. Klimontovich. Ito, Stratonovich and kinetic forms of stochastic-equations. *Physica A*, 163(2):515–532, FEB 15 1990.
- [117] P. E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*, volume 23 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1992.
- [118] T. Komorowski, C. Landim, and S. Olla. *Fluctuations in Markov processes*, volume 345 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Heidelberg, 2012. Time symmetry and martingale approximation.

- [119] L. B. Koralov and Y. G. Sinai. *Theory of probability and random processes*. Universitext. Springer, Berlin, second edition, 2007.
- [120] S. M. Kozlov. Effective diffusion for the Fokker-Planck equation. *Mat. Zametki*, 45(5):19–31, 124, 1989.
- [121] S. M. Kozlov. Geometric aspects of averaging. *Uspekhi Mat. Nauk*, 44(2(266)):79–120, 1989.
- [122] H. A. Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7:284–304, 1940.
- [123] P. Krée. Realization of Gaussian random fields. *J. Math. Phys.*, 24(11):2573–2580, 1983.
- [124] P. Krée and C. Soize. Markovianization of nonlinear oscillators with colored input. *Rend. Sem. Mat. Univ. Politec. Torino*, pages 135–150, 1982.
- [125] N. V. Krylov. *Introduction to the theory of random processes*, volume 43 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2002.
- [126] R. Kubo. The fluctuation-dissipation theorem. *Rep. Prog. Phys.*, 29:255–284, 1966.
- [127] R. Kubo, M. Toda, and N. Hashitsume. *Statistical physics. II*, volume 31 of *Springer Series in Solid-State Sciences*. Springer-Verlag, Berlin, second edition, 1991. Nonequilibrium statistical mechanics.
- [128] P. Kuchment. *Floquet theory for partial differential equations*, volume 60 of *Operator Theory: Advances and Applications*. Birkhäuser Verlag, Basel, 1993.
- [129] R. Kupferman. Fractional kinetics in Kac-Zwanzig heat bath models. *J. Statist. Phys.*, 114(1–2):291–326, 2004.
- [130] R. Kupferman, G. A. Pavliotis, and A. M. Stuart. Itô versus Stratonovich white-noise limits for systems with inertia and colored multiplicative noise. *Phys. Rev. E (3)*, 70(3):036120, 9, 2004.
- [131] R. Kupferman and A. M. Stuart. Fitting SDE models to nonlinear Kac-Zwanzig heat bath models. *Phys. D*, 199(3–4):279–316, 2004.
- [132] Y. Kuramoto. *Chemical oscillations, waves, and turbulence*, volume 19 of *Springer Series in Synergetics*. Springer-Verlag, Berlin, 1984.
- [133] Y. A. Kutoyants. *Statistical inference for ergodic diffusion processes*. Springer Series in Statistics. Springer-Verlag London Ltd., London, 2004.
- [134] A. Lasota and M. C. Mackey. *Chaos, fractals, and noise*, volume 97 of *Applied Mathematical Sciences*. Springer-Verlag, New York, second edition, 1994.
- [135] P. D. Lax. *Linear algebra and its applications*. Pure and Applied Mathematics (Hoboken). Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second edition, 2007.
- [136] C. E. Leith. Climate response and fluctuation dissipation. *Journal of the Atmospheric Sciences*, 32(10):2022–2026, 1975.
- [137] T. Lelievre, F. Nier, and G. A. Pavliotis. Optimal non-reversible linear drift for the convergence to equilibrium of a diffusion. *J. Stat. Phys.*, 152(2):237–274, 2013.

- [138] T. Lelièvre, M. Rousset, and G. Stoltz. *Free energy computations*. Imperial College Press, London, 2010. A mathematical perspective.
- [139] J. T. Lewis and L. C. Thomas. How to make a heat bath. In *Functional integration and its applications (Proc. Internat. Conf., London, 1974)*, pages 97–123. Clarendon Press, Oxford, 1975.
- [140] S. Lifson and J. L. Jackson. On the self-diffusion of ions in polyelectrolytic solution. *J. Chem. Phys.*, 36:2410, 1962.
- [141] K. Lindenberg and B. J. West. *The nonequilibrium statistical mechanics of open and closed systems*. VCH Publishers Inc., New York, 1990.
- [142] R. S. Liptser and A. N. Shiryaev. *Statistics of random processes. I*, volume 5 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, expanded edition, 2001.
- [143] J. S. Liu. *Monte Carlo strategies in scientific computing*. Springer Series in Statistics. Springer, New York, 2008.
- [144] M. Loèvè. *Probability theory. I*. Springer-Verlag, New York, fourth edition, 1977. Graduate Texts in Mathematics, Vol. 45.
- [145] M. Loèvè. *Probability theory. II*. Springer-Verlag, New York, fourth edition, 1978. Graduate Texts in Mathematics, Vol. 46.
- [146] L. Lorenzi and M. Bertoldi. *Analytical Methods for Markov Semigroups*. CRC Press, New York, 2006.
- [147] T. J. Lyons. Differential equations driven by rough signals. *Rev. Mat. Iberoamericana*, 14(2):215–310, 1998.
- [148] Z. M. Ma and M. Röckner. *Introduction to the theory of (nonsymmetric) Dirichlet forms*. Universitext. Springer-Verlag, Berlin, 1992.
- [149] M. C. Mackey. *Time's arrow*. Dover Publications Inc., Mineola, NY, 2003. The origins of thermodynamic behavior, Reprint of the 1992 original [Springer, New York; MR1140408].
- [150] M. C. Mackey, A. Longtin, and A. Lasota. Noise-induced global asymptotic stability. *J. Statist. Phys.*, 60(5–6):735–751, 1990.
- [151] M. O. Magnasco. Forced thermal ratchets. *Phys. Rev. Lett.*, 71(10):1477–1481, 1993.
- [152] B. B. Mandelbrot and J. W. Van Ness. Fractional Brownian motions, fractional noises and applications. *SIAM Rev.*, 10:422–437, 1968.
- [153] P. Mandl. *Analytical treatment of one-dimensional Markov processes*. Die Grundlehren der mathematischen Wissenschaften, Band 151. Academia Publishing House of the Czechoslovak Academy of Sciences, Prague, 1968.
- [154] X. Mao. *Stochastic differential equations and their applications*. Horwood Publishing Series in Mathematics & Applications. Horwood Publishing Limited, Chichester, 1997.
- [155] P. A. Markowich and C. Villani. On the trend to equilibrium for the Fokker-Planck equation: an interplay between physics and functional analysis. *Mat. Contemp.*, 19:1–29, 2000.
- [156] B. J. Matkowsky and Z. Schuss. Eigenvalues of the Fokker-Planck operator and the approach to equilibrium for diffusions in potential fields. *SIAM J. Appl. Math.*, 40(2):242–254, 1981.

- [157] B. J. Matkowsky, Z. Schuss, and E. Ben-Jacob. A singular perturbation approach to Kramers' diffusion problem. *SIAM J. Appl. Math.*, 42(4):835–849, 1982.
- [158] B. J. Matkowsky, Z. Schuss, and C. Tier. Uniform expansion of the transition rate in Kramers' problem. *J. Statist. Phys.*, 35(3–4):443–456, 1984.
- [159] J. C. Mattingly and A. M. Stuart. Geometric ergodicity of some hypoelliptic diffusions for particle motions. *Markov Processes and Related Fields*, 8(2):199–214, 2002.
- [160] G. F. Mazenko. *Nonequilibrium statistical mechanics*. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 2006.
- [161] R.M. Mazo. *Brownian motion*, volume 112 of *International Series of Monographs on Physics*. Oxford University Press, New York, 2002.
- [162] V. I. Melnikov. The Kramers problem—50 years of development. *Physics Reports*, 209(1–2):1–71, 1991.
- [163] V. I. Melnikov and S. V. Meshkov. Theory of activated rate-processes—exact solution of the Kramers problem. *J. Chem. Phys.*, 85(2):1018–1027, JUL 15 1986.
- [164] G. Metafune, D. Pallara, and E. Priola. Spectrum of Ornstein-Uhlenbeck operators in L^p spaces with respect to invariant measures. *J. Funct. Anal.*, 196(1):40–60, 2002.
- [165] J. Meyer and J. Schröter. Proper and normal solutions of the Fokker-Planck equation. *Arch. Rational Mech. Anal.*, 76(3):193–246, 1981.
- [166] J. Meyer and J. Schröter. Comments on the Grad procedure for the Fokker-Planck equation. *J. Statist. Phys.*, 32(1):53–69, 1983.
- [167] S. P. Meyn and R. L. Tweedie. *Markov chains and stochastic stability*. Communications and Control Engineering Series. Springer-Verlag London Ltd., London, 1993.
- [168] A. Mira and C. J. Geyer. On non-reversible Markov chains. In *Monte Carlo methods (Toronto, ON, 1998)*, volume 26 of *Fields Inst. Commun.*, pages 95–110. Amer. Math. Soc., Providence, RI, 2000.
- [169] H. Mori. A continued-fraction representation of the time-correlation functions. *Progress of Theoretical Physics*, 34(3):399–416, 1965.
- [170] E. Nelson. The adjoint Markoff process. *Duke Math. J.*, 25:671–690, 1958.
- [171] E. Nelson. *Dynamical theories of Brownian motion*. Princeton University Press, Princeton, N.J., 1967.
- [172] E. Nelson. *Quantum fluctuations*. Princeton Series in Physics. Princeton University Press, Princeton, NJ, 1985.
- [173] C. Nicolis. Stochastic aspects of climatic transitions-response to a periodic forcing. *Tellus*, 34(1):1–9, 1982.
- [174] C. Nicolis. Long-term climatic transitions and stochastic resonance. *Journal of Statistical Physics*, 70(1–2):3–14, jan 1993.
- [175] C. Nicolis and G. Nicolis. Stochastic aspects of climatic transitions—additive fluctuations. *Tellus*, 33(3):225–234, 1981.
- [176] J. R. Norris. *Markov chains*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 1998.

- [177] D. Nualart. *The Malliavin calculus and related topics*. Probability and Its Applications (New York). Springer-Verlag, Berlin, second edition, 2006.
- [178] B. Øksendal. *Stochastic differential equations*. Universitext. Springer-Verlag, Berlin, 2003.
- [179] M. Ottobre and G. A. Pavliotis. Asymptotic analysis for the generalized Langevin equation. *Nonlinearity*, 24(5):1629–1653, 2011.
- [180] M. Ottobre, G. A. Pavliotis, and K. Pravda-Starov. Exponential return to equilibrium for hypoelliptic quadratic systems. *J. Funct. Anal.*, 262(9):4000–4039, 2012.
- [181] G. A. Pavliotis and A. M. Stuart. White noise limits for inertial particles in a random field. *Multiscale Model. Simul.*, 1(4):527–533 (electronic), 2003.
- [182] G. A. Pavliotis and A. M. Stuart. Analysis of white noise limits for stochastic systems with two fast relaxation times. *Multiscale Model. Simul.*, 4(1):1–35 (electronic), 2005.
- [183] G. A. Pavliotis and A. Vogiannou. Diffusive transport in periodic potentials: Underdamped dynamics. *Fluct. Noise Lett.*, 8(2):L155–173, 2008.
- [184] G. A. Pavliotis. A multiscale approach to Brownian motors. *Phys. Lett. A*, 344:331–345, 2005.
- [185] G. A. Pavliotis and A. M. Stuart. *Multiscale methods*, volume 53 of *Texts in Applied Mathematics*. Springer, New York, 2008. Averaging and homogenization.
- [186] R. F. Pawula. Approximation of the linear Boltzmann equation by the Fokker–Planck equation. *Phys. Rev.*, 162(1):186–188, 1967.
- [187] E. Pollak. Theory of Activated Rate-processes—a new derivation of Kramers expression. *Journal of Chemical Physics*, 85(2):865–867, 1986.
- [188] E. Pollak, H. Grabert, and P. Hanggi. Theory of activated rate processes for arbitrary frequency dependent friction: Solution of the turnover problem. *The Journal of Chemical Physics*, 91(7):4073–4087, 1989.
- [189] M. Pradas, D. Tseluiko, S. Kalliadasis, D. T. Papageorgiou, and G. A. Pavliotis. Noise induced state transitions, intermittency, and universality in the noisy Kuramoto–Sivashinsky equation. *Phys. Rev. Lett.*, 106(6):060602, Feb 2011.
- [190] G. Da Prato and J. Zabczyk. *Stochastic Equations in Infinite Dimensions*, volume 44 of *Encyclopedia of Mathematics and Its Applications*. Cambridge University Press, 1992.
- [191] C. Prévôt and M. Röckner. *A concise course on stochastic partial differential equations*, volume 1905 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007.
- [192] H. Qian, M. Qian, and X. Tang. Thermodynamics of the general diffusion process: time–reversibility and entropy production. *J. Stat. Phys.*, 107(5/6):1129–1141, 2002.
- [193] M. Reed and B. Simon. *Methods of modern mathematical physics. II. Fourier analysis, self-adjointness*. Academic Press, New York, 1975.
- [194] M. Reed and B. Simon. *Methods of modern mathematical physics. IV. Analysis of operators*. Academic Press, New York, 1978.

- [195] M. Reed and B. Simon. *Methods of modern mathematical physics. I.* Academic Press Inc., New York, second edition, 1980. Functional analysis.
- [196] P. Reimann. Brownian motors: noisy transport far from equilibrium. *Phys. Rep.*, 361(2–4):57–265, 2002.
- [197] P. Reimann, C. Van den Broeck, H. Linke, P. Hänggi, J. M. Rubi, and A. Perez-Madrid. Diffusion in tilted periodic potentials: enhancement, universality and scaling. *Phys. Rev. E*, 65(3):031104, 2002.
- [198] P. Reimann, C. Van den Broeck, H. Linke, J. M. Rubi, and A. Perez-Madrid. Giant acceleration of free diffusion by use of tilted periodic potentials. *Phys. Rev. Lett.*, 87(1):010602, 2001.
- [199] P. Resibois and M. De Leener. *Classical Kinetic Theory of Fluids*. Wiley, New York, 1977.
- [200] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999.
- [201] L. Rey-Bellet. Ergodic properties of markov processes. In S. Attal, A. Joye, and C.-A. Pillet, editors, *Open Quantum Systems II*, volume 1881 of *Lecture Notes in Mathematics*, pages 1–39. Springer, 2006.
- [202] L. Rey-Bellet. Open classical systems. In *Open quantum systems. II*, volume 1881 of *Lecture Notes in Math.*, pages 41–78. Springer, Berlin, 2006.
- [203] L. Rey-Bellet and L. E. Thomas. Exponential convergence to non-equilibrium stationary states in classical statistical mechanics. *Comm. Math. Phys.*, 225(2):305–329, 2002.
- [204] F. Riesz and B. Sz.-Nagy. *Functional analysis*. Dover Publications Inc., New York, 1990. Translated from the second French edition by Leo F. Boron, Reprint of the 1955 original.
- [205] H. Risken. Solutions of F-Planck equation in detailed balance. *Z. Physik*, 251(3):231, 1972.
- [206] H. Risken. *The Fokker-Planck equation*, volume 18 of *Springer Series in Synergetics*. Springer-Verlag, Berlin, 1989.
- [207] G. O. Roberts and J. S. Rosenthal. Optimal scaling for various Metropolis-Hastings algorithms. *Statist. Sci.*, 16(4):351–367, 2001.
- [208] G. O. Roberts and O. Stramer. Langevin diffusions and Metropolis-Hastings algorithms. *Methodol. Comput. Appl. Probab.*, 4(4):337–357 (2003), 2002. International Workshop in Applied Probability (Caracas, 2002).
- [209] G. O. Roberts and R. L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, 1996.
- [210] H. Rodenhausen. Einstein’s relation between diffusion constant and mobility for a diffusion model. *J. Statist. Phys.*, 55(5–6):1065–1088, 1989.
- [211] L. C. G. Rogers and D. Williams. *Diffusions, Markov processes, and martingales. Vol. 1*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2000.

- [212] L. C. G. Rogers and D. Williams. *Diffusions, Markov processes, and martingales. Vol. 2*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2000.
- [213] D. Ruelle. A review of linear response theory for general differentiable dynamical systems. *Nonlinearity*, 22(4):855–870, 2009.
- [214] Y. Saito and T. Mitsui. Stability analysis of numerical schemes for stochastic differential equations. *SIAM J. Numer. Anal.*, 33(6):2254–2267, 1996.
- [215] J. Schröter. The complete Chapman-Enskog procedure for the Fokker-Planck equation. *Arch. Rational. Mech. Anal.*, 66(2):183–199, 1977.
- [216] Z. Schuss. Singular perturbation methods in stochastic differential equations of mathematical physics. *SIAM Review*, 22(2):119–155, 1980.
- [217] C. Schwab and R.A. Todor. Karhunen-Loëve approximation of random fields by generalized fast multipole methods. *J. Comput. Phys.*, 217(1):100–122, 2006.
- [218] B. Simon. Schrödinger semigroups. *Bull. Amer. Math. Soc. (N.S.)*, 7(3):447–526, 1982.
- [219] B. Simon. *Functional integration and quantum physics*. AMS Chelsea Publishing, Providence, RI, second edition, 2005.
- [220] C. Soize. *The Fokker-Planck equation for stochastic dynamical systems and its explicit steady state solutions*, volume 17 of *Series on Advances in Mathematics for Applied Sciences*. World Scientific Publishing Co. Inc., River Edge, NJ, 1994.
- [221] M. Sørensen. Estimating functions for diffusion-type processes. In *Statistical methods for stochastic differential equations*, volume 124 of *Monogr. Statist. Appl. Probab.*, pages 1–107. CRC Press, Boca Raton, FL, 2012.
- [222] R.B. Sowers. A boundary layer theory for diffusively perturbed transport around a heteroclinic cycle. *Comm. Pure Appl. Math.*, 58(1):30–84, 2005.
- [223] H. Spohn. *Large Scale Dynamics of Interacting Particle Systems*. Springer, Heidelberg, 1991.
- [224] R. L. Stratonovich. A new representation for stochastic integrals and equations. *SIAM J. Control*, 4:362–371, 1966.
- [225] R. L. Stratonovich. *Topics in the theory of random noise. Vol. II*. Revised English edition. Translated from the Russian by Richard A. Silverman. Gordon and Breach Science Publishers, New York, 1967.
- [226] D. W. Stroock. *Partial differential equations for probabilists*, volume 112 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2008.
- [227] D. W. Stroock. *Probability theory, an analytic view*. Cambridge University Press, Cambridge, 1993.
- [228] D. W. Stroock. *An introduction to Markov processes*, volume 230 of *Graduate Texts in Mathematics*. Springer-Verlag, Berlin, 2005.
- [229] H. J. Sussmann. On the gap between deterministic and stochastic ordinary differential equations. *Ann. Probability*, 6(1):19–41, 1978.
- [230] H. J. Sussmann. Limits of the Wong-Zakai type with a modified drift term. In *Stochastic analysis*, pages 475–493. Academic Press, Boston, MA, 1991.

- [231] G. I. Taylor. Diffusion by continuous movements. *London Math. Soc.*, 20:196, 1921.
- [232] G. Teschl. *Mathematical methods in quantum mechanics*, volume 99 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2009. With applications to Schrödinger operators.
- [233] U. M. Titulaer. A systematic solution procedure for the Fokker-Planck equation of a Brownian particle in the high-friction case. *Phys. A*, 91(3–4):321–344, 1978.
- [234] G. E. Uhlenbeck and L. S. Ornstein. On the theory of the Brownian motion. *Phys. Rev.*, 36(5):823–841, Sep 1930.
- [235] N. G. van Kampen. Itô versus Stratonovich. *J. Statist. Phys.*, 24(1):175–187, 1981.
- [236] N. G. van Kampen. *Stochastic processes in physics and chemistry*. North-Holland Publishing Co., Amsterdam, 1981. Lecture Notes in Mathematics, 888.
- [237] C. Villani. Hypocoercivity. *Mem. Amer. Math. Soc.*, 202(950):iv+141, 2009.
- [238] G. Volpe, L. Helden, T. Brettschneider, J. Wehr, and C. Bechinger. Influence of Noise on Force Measurements. *Physical Review Letters*, 104(17), APR 30 2010.
- [239] N. Wax (editor). *Selected Papers on Noise and Stochastic Processes*. Dover, New York, 1954.
- [240] E. Wong and M. Zakai. On the convergence of ordinary integrals to stochastic integrals. *Ann. Math. Statist.*, 36:1560–1564, 1965.
- [241] E. Wong and M. Zakai. Riemann-Stieltjes approximations of stochastic integrals. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 12:87–97, 1969.
- [242] D. Wycoff and N. L. Balazs. Multiple time scales analysis for the Kramers-Chandrasekhar equation. *Phys. A*, 146(1–2):175–200, 1987.
- [243] D. Wycoff and N. L. Balazs. Multiple time scales analysis for the Kramers-Chandrasekhar equation with a weak magnetic field. *Phys. A*, 146(1–2):201–218, 1987.
- [244] F. Yang, Y. Chen, and Y. Liu. The Green-Kubo formula for general Markov processes with a continuous time parameter. *J. Phys. A*, 43(24):245002, 17, 2010.
- [245] R. Zwanzig. Nonlinear generalized Langevin equations. *J. Stat. Phys.*, 9(3):215–220, 1973.
- [246] R. Zwanzig. *Nonequilibrium statistical mechanics*. Oxford University Press, New York, 2001.

Index

- additive noise, 56
- adiabatic approximation, 253
- asymptotic normality
 - of the maximum likelihood estimator, 164
- autocorrelation function, 5
- backward Kolmogorov equation, 39
- Bakry–Emery criterion, 113
- Birkhoff’s ergodic theorem, 4
- Bochner’s theorem, 6, 7
- Brinkman’s hierarchy, 229
- Brownian motion, 11
 - scaling and symmetry properties, 13
- Brownian motor, 252, 254
- canonical commutation relations, 197
- canonical distribution, 183
- Cauchy distribution, 8
- Cauchy–Lorentz function, 122
- causality principle, 295
- centering condition, 217
- central limit theorem, 319
- Chapman–Kolmogorov equation, 31, 35
- characteristic function
 - of a random variable, 311
- collision operator, 229
- colored noise, 139, 140, 172
- conditional expectation, 310
- conditional probability, 310
- conditional probability density, 30
- confinement time, 238
- confining potential, 110
- contraction semigroup, 51
- correlation coefficient, 308
- covariance function, 5
- Cox–Ingersoll–Ross model, 78
- creation–annihilation operators
 - Ornstein–Uhlenbeck process, 107
- Csiszar–Kullback inequality, 133
- detailed balance, 100, 119
- diffusion coefficient, 10
- diffusion process, 33
 - confinement time, 238
 - joint probability density
 - stationary, 96
 - mean exit time, 237
 - mean first passage time, 237
 - reversible, 103, 116
- dipole approximation, 269
- Dirichlet form, 112, 227, 293

- distribution
 - canonical, 183
 - Gibbs, 183
 - Maxwell–Boltzmann, 183
- Doob's theorem, 43, 273
- Duhamel's principle, 270
- Dynkin's formula, 83, 238
- equation
 - Chapman–Kolmogorov, 31, 35
 - Fokker–Planck, 48, 69, 70, 87, 88
 - stationary, 41
 - forward Kolmogorov, 40, 48
 - generalized Langevin, 263
 - Klein–Kramers–Chandrasekhar, 228
 - Landau–Stuart, 56, 157, 166
 - Langevin, 181, 228
 - overdamped Langevin, 183, 252
 - Poisson, 259
 - Smoluchowski, 87, 109, 252
 - Tanaka, 83
- ergodic Markov process, 40
- ergodic stochastic process, 4
- estimator
 - asymptotically unbiased, 160
 - maximum likelihood, 160
- Euler–Maruyama method, 146
- Feynman–Kac formula, 69, 84
- filtration, 33
- first exit time, 58, 238
- first passage time, 238
- first passage time distribution, 239
- Floquet theory, 247
- fluctuation–dissipation theorem, 173, 181, 263, 272, 283, 289
- Fokker–Planck equation, 48, 69, 70, 87, 88
 - classical solution of, 89
- forward Kolmogorov equation, 40, 48
- fractional Brownian motion, 16
- Fredholm alternative, 141
- free energy functional, 115
- friction coefficient, 181, 278
- Gaussian stochastic process, 2
- generalized Langevin equation, 263, 267, 272
- generator
 - diffusion process, 50, 66
 - Markov process, 38
 - Markov semigroup, 38
- geometric Brownian motion, 74, 98
- Gibbs distribution, 110, 183
- Gibbs measure, 112, 268
- Girsanov's theorem, 77, 79, 163
- gluing condition, 230
- Green–Kubo formula, 10, 231, 283, 292
- Hörmander's theorem, 188
- harmonic noise, 276
- heat bath, 267
- Hermite polynomials, 106
- Hessian, 300
- Heston model, 157
- high-frequency limit, 157
- Holley–Stroock lemma, 133
- hopping rate, 237
- hypocoercivity, 190
- hypoelliptic operator, 187
- hypoellipticity, 185
- implicit numerical method for SDEs, 155
- inequality
 - Csiszar–Kullback, 133
 - logarithmic Sobolev, 132
 - Pinsker, 133
 - Poincaré, 112
- inner product
 - matrix, 299
- invariant measure, 41
 - Markov process, 51
- Itô isometry, 60
- Itô stochastic calculus, 57
- Itô stochastic differential equation, 61
- Itô stochastic integral, 59
- Itô versus Stratonovich problem, 172
- Itô's formula, 57, 67
- Itô-to-Stratonovich correction, 62

- Kac–Zwanzig model, 281, 282
Karhunen–Loève Expansion, 17
for Brownian motion, 21
Klimontovich stochastic integral, 61, 173
Kramers rate, 243
Kramers time, 237, 243, 246
Kramers–Kronig relations, 296

Lamperti transformation, 77, 78, 85
Landau–Stuart equation, 56, 157, 166
Langevin equation, 181, 228
energy-diffusion-limited regime, 214
overdamped limit, 204
underdamped limit, 204
underdamped regime, 200
large sample limit, 161, 164
law
of a random variable, 305
law of large numbers
strong, 318
law of the iterated logarithm, 175
law of total probability, 310
lemma
Holley–Stroock, 133
Weyl, 188
Lepingle–Ribemont method, 154
Lie algebra, 187
likelihood function, 79, 159, 160
linear response theory, 264, 283
linear stochastic differential equation
autocorrelation matrix, 80
Liouville operator, 67, 182
logarithmic Sobolev inequality (LSI), 115, 132
Lorentz distribution, 8
Lyapunov equation, 81
Lyapunov function, 82, 114, 131

Markov chain
continuous-time, 30
discrete-time, 30
generator, 32
transition matrix, 31
Markov chain Monte Carlo, 123

Markov process, 29
ergodic, 40
generator, 38
stationary, 42
time-homogeneous, 31, 36
transition function, 35
transition probability, 31
Markov semigroup, 38
martingale, 60
maximum likelihood estimator, 159, 160
Maxwell–Boltzmann distribution, 183
mean exit time, 236–238
mean first passage time, 236–238
mean reverting Ornstein–Uhlenbeck
process, 74
mean-square stability, 155
Mercer’s theorem, 19
metastable state, 236
Milstein’s method, 146
Mori–Zwanzig formalism, 281
multiplicative noise, 56

noise amplification factor, 287
noise induced stabilization, 170
noise-induced transition, 169
norm
Euclidean, 299
Frobenius, 299
numerical method
A-stable, 155
Euler–Maruyama, 146
implicit, 155
Lepingle–Ribemont, 154
Milstein, 146
stochastic theta method, 156
strong convergence, 147
weak convergence, 150

opérateur carré du champ, 131, 292
open classical system, 182, 267
operator
coercive, 186
Hilbert–Schmidt, 26
hypoelliptic, 182, 187
Liouville, 182

- Ornstein–Uhlenbeck process, 8, 42
 - Fokker–Planck equation, 95
 - mean reverting, 74
- overdamped Langevin equation, 183, 252
- partition function, 110, 183
- Pawula's theorem, 52
- Pinsker inequality, 133
- Poincaré inequality, 112
 - for Gaussian measures, 105
- Poisson equation, 259
- Poisson process, 25
- probability density function
 - of a random variable, 306
- probability flux, 89
- quadratic variation, 157
- quasi-Markovian process, 274
- random variable, 305
 - Gaussian, 308
 - uncorrelated, 308
- rare event, 236
- rate coefficient, 237, 262
- reaction rate, 237
- reaction rate theory, 281
- reduced distribution function, 307
- relative entropy, 115, 132, 135
- response function, 287
- reversible diffusion, 103, 116
- sample path, 2
- sample space, 303
- semigroup of operators, 38
- signal-to-noise ratio, 264
- Slutsky's theorem, 166, 319
- Smoluchowski equation, 87, 109, 110, 252
- spectral amplification, 248, 264
- spectral density, 7
- spectral measure, 7
- stationary Fokker–Planck equation, 41
- stationary Markov process, 42
- stationary process, 3
 - second-order stationary, 4
 - strictly stationary, 3
 - wide-sense stationary, 4
- stochastic differential equation, 14
 - additive noise, 56
 - Cox–Ingersoll–Ross, 78
 - Itô, 61
 - multiplicative noise, 56
 - stable in mean square, 155
 - Stratonovich, 61
 - strong solution, 64, 82
 - Verhulst, 85
 - weak solution, 82
- stochastic integral
 - Itô, 59
 - Klimontovich, 61, 173
 - Stratonovich, 59
- stochastic matrix, 31
- stochastic mechanics, 135
- stochastic process
 - definition, 1
 - equivalent, 2
 - ergodic, 4
 - finite-dimensional distributions, 2
 - Gaussian, 2
 - sample path, 2
 - second-order stationary, 5
 - stationary, 3
 - strictly stationary, 3
- stochastic resonance, 245, 246
- stochastic theta method, 156
- Stone's theorem, 282
- stopping time, 58, 83, 238
- Stratonovich stochastic differential equation, 61
- Stratonovich stochastic integral, 59
- strong solution of a stochastic differential equation, 64
- susceptibility, 250, 288, 295
- Tanaka equation, 83
- temperature, 181
- theorem
 - Birkhoff, 4

- Bochner, 6, 7
Doob, 43, 273
fluctuation–dissipation, 173, 181, 263, 272, 283, 289
Girsanov, 77, 79, 163
Hörmander, 188
Mercer, 19
Pawula, 52
Slutsky, 166, 319
Stone, 282
Wong–Zakai, 139, 172
theory of rough paths, 172
thermal reservoir, 267
time-homogeneous Markov process, 36
transformation
 Lamperti, 78, 85
transition function, 35
transition matrix
 Markov chain, 31
transition probability
 of a Markov process, 31
transition probability density, 36
transition state theory, 281
transport coefficient, 10, 231, 291
uniform ellipticity
 Fokker–Planck equation, 88
Verhulst stochastic differential equation, 85
wave equation, 268
Weyl’s lemma, 188
Wiener process, 11
Wong–Zakai theorem, 139, 172