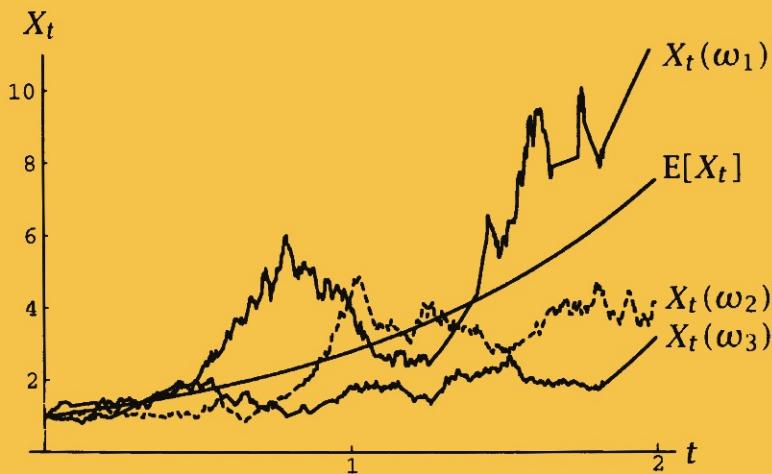


Bernt Øksendal

Stochastic Differential Equations

An Introduction with Applications

Fourth Edition



Springer

Universität

Universitext

Springer-Verlag Berlin Heidelberg GmbH

Bernt Øksendal

Stochastic Differential Equations

An Introduction with Applications

Fourth Edition



Springer

Bernt Øksendal

Department of Mathematics
University of Oslo
Box 1053, Blindern
N-0316 Oslo
Norway

The front cover shows three sample paths $X_t(\omega_1)$, $X_t(\omega_2)$ and $X_t(\omega_3)$ of a geometric Brownian motion $X_t(\omega)$, i.e. of the solution of a (1-dimensional) stochastic differential equation of the form

$$\frac{dX_t}{dt} = (r + \alpha \cdot W_t)X_t \quad t \geq 0; X_0 = x$$

where x , r and α are constants and $W_t = W_t(\omega)$ is white noise. This process is often used to model “exponential growth under uncertainty”. See Chapters V, X and XI.

The figure is a computer simulation for the case $x = r = 1$, $\alpha = 0.6$. The mean value of X_t , $E[X_t] = \exp(t)$, is also drawn. Courtesy of Jan Ubøe, Stord/Haugesund College.

Mathematics Subject Classification (1991): 60H10, 60G35, 60G40, 60G44, 93E20, 60J45, 60J25

ISBN 978-3-540-60243-9

Cataloging-in-Publication Data applied for

Die Deutsche Bibliothek – CIP-Einheitsaufnahme
Øksendal, Bernt:

Stochastic differential equations: an introduction with applications / Bernt Øksendal. – 4. ed.

(Universitext)

ISBN 978-3-540-60243-9 ISBN 978-3-662-03185-8 (eBook)

DOI 10.1007/978-3-662-03185-8

This work is subject to copyright. All rights are reserved, 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 way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer-Verlag Berlin Heidelberg GmbH.

Violations are liable for prosecution under the German Copyright Law.

© Springer-Verlag Berlin Heidelberg 1995

Originally published by Springer-Verlag Berlin Heidelberg New York Tokyo in 1995

Typesetting: Camera-ready copy produced by author using a Springer TeX macro package

SPIN: 10500248

41/3143-543210 – Printed on acid-free paper

*To My Family
Eva, Elise, Anders and Karina*

We have not succeeded in answering all our problems. The answers we have found only serve to raise a whole set of new questions. In some ways we feel we are as confused as ever, but we believe we are confused on a higher level and about more important things.

Posted outside the mathematics reading room,
Tromsø University

Preface to the Fourth Edition

In this edition I have added some material which is particularly useful for the applications, namely the martingale representation theorem (Chapter IV), the variational inequalities associated to optimal stopping problems (Chapter X) and stochastic control with terminal conditions (Chapter XI). In addition solutions and extra hints to some of the exercises are now included. Moreover, the proof and the discussion of the Girsanov theorem have been changed in order to make it more easy to apply, e.g. in economics. And the presentation in general has been corrected and revised throughout the text, in order to make the book better and more useful.

During this work I have benefitted from valuable comments from several persons, including Knut Aase, Sigmund Berntsen, Mark H. A. Davis, Helge Holden, Yaozhong Hu, Tom Lindstrøm, Trygve Nilsen, Paulo Ruffino, Isaac Saais, Clint Scovel, Jan Ubøe, Suleyman Ustunel, Qinghua Zhang, Tusheng Zhang and Victor Daniel Zerkowski. I am grateful to them all for their help.

My special thanks go to Håkon Nyhus, who carefully read large portions of the manuscript and gave me a long list of improvements, as well as many other useful suggestions.

Finally I wish to express my gratitude to Tove Møller and Dina Haraldsson, who typed the manuscript with impressive proficiency.

Oslo, June 1995

Bernt Øksendal

Preface to the Third Edition

The main new feature of the third edition is that exercises have been included to each of the chapters II–XI. The purpose of these exercises is to help the reader to get a better understanding of the text. Some of the exercises are quite routine, intended to illustrate the results, while other exercises are harder and more challenging and some serve to extend the theory.

I have also continued the effort to correct misprints and errors and to improve the presentation. I have benefitted from valuable comments and suggestions from Mark H. A. Davis, Håkon Gjessing, Torgny Lindvall and Håkon Nyhus, My best thanks to them all.

A quite noticeable non-mathematical improvement is that the book is now typed in *T_EX*. Tove Lieberg did a great typing job (as usual) and I am very grateful to her for her effort and infinite patience.

Oslo, June 1991

Bernt Øksendal

Preface to the Second Edition

In the second edition I have split the chapter on diffusion processes in two, the new Chapters VII and VIII: Chapter VII treats only those basic properties of diffusions that are needed for the applications in the last 3 chapters. The readers that are anxious to get to the applications as soon as possible can therefore jump directly from Chapter VII to Chapters IX, X and XI.

In Chapter VIII other important properties of diffusions are discussed. While not strictly necessary for the rest of the book, these properties are central in today's theory of stochastic analysis and crucial for many other applications.

Hopefully this change will make the book more flexible for the different purposes. I have also made an effort to improve the presentation at some points and I have corrected the misprints and errors that I knew about, hopefully without introducing new ones. I am grateful for the responses that I have received on the book and in particular I wish to thank Henrik Martens for his helpful comments.

Tove Lieberg has impressed me with her unique combination of typing accuracy and speed. I wish to thank her for her help and patience, together with Dina Haraldsson and Tone Rasmussen who sometimes assisted on the typing.

Oslo, August 1989

Bernt Øksendal

Preface to the First Edition

These notes are based on a postgraduate course I gave on stochastic differential equations at Edinburgh University in the spring 1982. No previous knowledge about the subject was assumed, but the presentation is based on some background in measure theory.

There are several reasons why one should learn more about stochastic differential equations: They have a wide range of applications outside mathematics, there are many fruitful connections to other mathematical disciplines and the subject has a rapidly developing life of its own as a fascinating research field with many interesting unanswered questions.

Unfortunately most of the literature about stochastic differential equations seems to place so much emphasis on rigor and completeness that it scares many nonexperts away. These notes are an attempt to approach the subject from the nonexpert point of view: Not knowing anything (except rumours, maybe) about a subject to start with, what would I like to know first of all? My answer would be:

- 1) In what situations does the subject arise?
- 2) What are its essential features?
- 3) What are the applications and the connections to other fields?

I would not be so interested in the proof of the most general case, but rather in an easier proof of a special case, which may give just as much of the basic idea in the argument. And I would be willing to believe some basic results without proof (at first stage, anyway) in order to have time for some more basic applications.

These notes reflect this point of view. Such an approach enables us to reach the highlights of the theory quicker and easier. Thus it is hoped that these notes may contribute to fill a gap in the existing literature. The course is meant to be an appetizer. If it succeeds in awaking further interest, the reader will have a large selection of excellent literature available for the study of the whole story. Some of this literature is listed at the back.

In the introduction we state 6 problems where stochastic differential equations play an essential role in the solution. In Chapter II we introduce the basic mathematical notions needed for the mathematical model of some of these problems,

Preface to the First Edition

leading to the concept of Ito integrals in Chapter III. In Chapter IV we develop the stochastic calculus (the Ito formula) and in Chapter V we use this to solve some stochastic differential equations, including the first two problems in the introduction. In Chapter VI we present a solution of *the linear filtering problem* (of which problem 3 is an example), using the stochastic calculus. Problem 4 is *the Dirichlet problem*. Although this is purely deterministic we outline in Chapters VII and VIII how the introduction of an associated Ito diffusion (i.e. solution of a stochastic differential equation) leads to a simple, intuitive and useful stochastic solution, which is the cornerstone of stochastic potential theory. Problem 5 is an *optimal stopping problem*. In Chapter IX we represent the state of a game at time t by an Ito diffusion and solve the corresponding optimal stopping problem. The solution involves potential theoretic notions, such as the generalized harmonic extension provided by the solution of the Dirichlet problem in Chapter VIII. Problem 6 is a stochastic version of F.P. Ramsey's classical control problem from 1928. In Chapter X we formulate the general *stochastic control problem* in terms of stochastic differential equations, and we apply the results of Chapters VII and VIII to show that the problem can be reduced to solving the (deterministic) Hamilton-Jacobi-Bellman equation. As an illustration we solve a problem about optimal portfolio selection.

After the course was first given in Edinburgh in 1982, revised and expanded versions were presented at Agder College, Kristiansand and University of Oslo. Every time about half of the audience have come from the applied section, the others being so-called "pure" mathematicians. This fruitful combination has created a broad variety of valuable comments, for which I am very grateful. I particularly wish to express my gratitude to K.K. Aase, L. Csink and A.M. Davie for many useful discussions.

I wish to thank the Science and Engineering Research Council, U.K. and Norges Almenvitenskapelige Forskningsråd (NAVF), Norway for their financial support. And I am greatly indebted to Ingrid Skram, Agder College and Inger Prestbakken, University of Oslo for their excellent typing - and their patience with the innumerable changes in the manuscript during these two years.

Oslo, June 1985

Bernt Øksendal

Note: Chapters VIII, IX, X of the First Edition have become Chapters IX, X, XI of the Second Edition.

Contents

I. Introduction	1
Some problems (1-6) where stochastic differential equations play an essential role in the solution	
II. Some Mathematical Preliminaries	5
Random variables, independence, stochastic processes.....	5
Kolmogorov's extension theorem.....	9
Brownian motion.....	9
Basic properties of Brownian motion	10
Versions of processes and Kolmogorov's continuity theorem	12
Exercises.....	13
III. Ito Integrals.....	18
Mathematical interpretation of equations involving "noise".....	18
The Ito integral.....	22
Some properties of the Ito integral.....	27
Martingales.....	27
Extensions of the Ito integral.....	31
Comparison between Ito and Stratonovich integrals.....	32
Exercises.....	34
IV. Ito Processes and the Ito Formula.....	40
Ito processes (stochastic integrals).....	41
The 1-dimensional Ito formula.....	41
The multi-dimensional Ito formula.....	45
The martingale representation theorem.....	46
Exercises.....	51
V. Stochastic Differential Equations	59
The population growth model and other examples	59

Brownian motion on the unit circle	63
Existence and uniqueness theorem for stochastic differential equations	64
Weak and strong solutions	68
Exercises	69
VI. The Filtering Problem	75
Statement of the general problem	76
The linear filtering problem	77
Step 1: Z-linear and Z-measurable estimates	81
Step 2: The innovation process	82
Step 3: The innovation process and Brownian motion	86
Step 4: An explicit formula for X_t	87
Step 5: The stochastic differential equation for \hat{X}_t	88
The 1-dimensional Kalman-Bucy filter	91
Examples	91
The multi-dimensional Kalman-Bucy filter	97
Exercises	98
VII. Diffusions: Basic Properties	103
Definition of an Ito diffusion	104
(A) The Markov property	104
(B) The strong Markov property	106
Hitting distribution, harmonic measure and the mean value property	110
(C) The generator of a diffusion	111
(D) The Dynkin formula	114
(E) The characteristic operator	116
Exercises	118
VIII. Other Topics in Diffusion Theory	124
(A) Kolmogorov's backward equation The resolvent	124
(B) The Feynman-Kac formula. Killing	128
(C) The martingale problem	131
(D) When is an Ito process a diffusion? How to recognize a Brownian motion	133
(E) Random time change	138
Time change formula for Ito integrals	141
Examples: Brownian motion on the unit sphere	142
Harmonic and analytic functions	143

(F) The Girsanov theorem	145
Exercises	152
IX. Applications to Boundary Value Problems	160
(A) The Dirichlet problem.....	160
Regular points	162
The stochastic Dirichlet problem	166
Existence and uniqueness of solution	166
When is the solution of the stochastic Dirichlet problem also a solution of the original Dirichlet problem?.....	168
(B) The Poisson problem.....	172
A stochastic version.....	172
Existence of solution	172
Uniqueness of solution	174
The combined Dirichlet-Poisson problem	175
The Green measure	175
Exercises	177
X. Application to Optimal Stopping	183
Statement of the problem	183
Least superharmonic majorants	186
Existence theorem for optimal stopping	189
Uniqueness theorem for optimal stopping.....	193
Examples: Some stopping problems for Brownian motion	194
Reward functions assuming negative values.....	195
The time-inhomogeneous case	196
Example: When is the right time to sell the stocks?.....	197
Optimal stopping problems involving an integral.....	200
Connection with variational inequalities	202
Exercises	206
XI. Application to Stochastic Control	212
Statement of the problem	212
The Hamilton-Jacobi-Bellman (HJB) equation.....	214
A converse of the HJB equation	217
Markov controls versus general adaptive controls	219
The linear stochastic regulator problem	220
An optimal portfolio selection problem	223
A simple problem where the optimal process is discontinuous	225
Stochastic control problems with terminal conditions	228
Exercises	229

XVI Contents

Appendix A: Normal Random Variables	236
Appendix B: Conditional Expectations	239
Appendix C: Uniform Integrability and Martingale Convergence	241
Solutions and additional hints to some of the exercises ...	244
Bibliography	252
List of Frequently Used Notation and Symbols	261
Index	265

I. Introduction

To convince the reader that stochastic differential equations is an important subject let us mention some situations where such equations appear and can be used:

(A) Stochastic analogs of classical differential equations

If we allow for some randomness in some of the coefficients of a differential equation we often obtain a more realistic mathematical model of the situation.

PROBLEM 1. Consider the simple population growth model

$$\frac{dN}{dt} = a(t)N(t), \quad N(0) = A \quad (1.1)$$

where $N(t)$ is the size of the population at time t , and $a(t)$ is the relative rate of growth at time t . It might happen that $a(t)$ is not completely known, but subject to some random environmental effects, so that we have

$$a(t) = r(t) + \text{"noise"},$$

where we do not know the exact behaviour of the noise term, only its probability distribution. The function $r(t)$ is assumed to be nonrandom. How do we solve (1.1) in this case?

PROBLEM 2. The charge $Q(t)$ at time t at a fixed point in an electric circuit satisfies the differential equation

$$L \cdot Q''(t) + R \cdot Q'(t) + \frac{1}{C} \cdot Q(t) = F(t), \quad Q(0) = Q_0, \quad Q'(0) = I_0 \quad (1.2)$$

where L is inductance, R is resistance, C is capacitance and $F(t)$ the potential source at time t .

Again we may have a situation where some of the coefficients, say $F(t)$, are not deterministic but of the form

$$F(t) = G(t) + \text{"noise"}. \quad (1.3)$$

How do we solve (1.2) in this case?

More generally, the equation we obtain by allowing randomness in the coefficients of a differential equation is called a *stochastic differential equation*. This will be made more precise later. It is clear that any solution of a stochastic differential equation must involve some randomness, i.e. we can only hope to be able to say something about the probability distributions of the solutions.

(B) Filtering problems

PROBLEM 3. Suppose that we, in order to improve our knowledge about the solution, say of Problem 2, perform observations $Z(s)$ of $Q(s)$ at times $s \leq t$. However, due to inaccuracies in our measurements we do not really measure $Q(s)$ but a disturbed version of it:

$$Z(s) = Q(s) + \text{“noise”}. \quad (1.4)$$

So in this case there are two sources of noise, the second coming from the error of measurement.

The *filtering problem* is: What is the best estimate of $Q(t)$ satisfying (1.2), based on the observations Z_s in (1.4), where $s \leq t$? Intuitively, the problem is to “filter” the noise away from the observations in an optimal way.

In 1960 Kalman and in 1961 Kalman and Bucy proved what is now known as the Kalman-Bucy filter. Basically the filter gives a procedure for estimating the state of a system which satisfies a “noise” linear differential equation, based on a series of “noisy” observations.

Almost immediately the discovery found applications in aerospace engineering (Ranger, Mariner, Apollo etc.) and it now has a broad range of applications.

Thus the Kalman-Bucy filter is an example of a recent mathematical discovery which has already proved to be useful - it is not just “potentially” useful.

It is also a counterexample to the assertion that “applied mathematics is bad mathematics” and to the assertion that “the only really useful mathematics is the elementary mathematics”. For the Kalman-Bucy filter - as the whole subject of stochastic differential equations - involves advanced, interesting and first class mathematics.

(C) Stochastic approach to deterministic boundary value problems

PROBLEM 4. The most celebrated example is the stochastic solution of the *Dirichlet problem*:

Given a (reasonable) domain U in \mathbf{R}^n and a continuous function f on the boundary of $U, \partial U$. Find a function \tilde{f} continuous on the closure \overline{U} of U such that

- (i) $\tilde{f} = f$ on ∂U
- (ii) \tilde{f} is harmonic in U , i.e.

$$\Delta \tilde{f} := \sum_{i=1}^n \frac{\partial^2 \tilde{f}}{\partial x_i^2} = 0 \text{ in } U.$$

In 1944 Kakutani proved that the solution could be expressed in terms of *Brownian motion* (which will be constructed in Chapter II): $\tilde{f}(x)$ is the expected value of f at the first exit point from U of the Brownian motion starting at $x \in U$.

It turned out that this was just the tip of an iceberg: For a large class of semielliptic second order partial differential equations the corresponding Dirichlet boundary value problem can be solved using a stochastic process which is a solution of an associated stochastic differential equation.

(D) Optimal stopping

PROBLEM 5. Suppose a person has an asset or resource (e.g. a house, stocks, oil...) that she is planning to sell. The price X_t at time t of her asset on the open market varies according to a stochastic differential equation of the same type as in Problem 1:

$$\frac{dX_t}{dt} = rX_t + \alpha X_t \cdot \text{“noise”}$$

where r, α are known constants. The discount rate is a known constant ρ . At what time should she decide to sell?

We assume that she knows the behaviour of X_s up to the present time t , but because of the noise in the system she can of course never be sure at the time of the sale if her choice of time will turn out to be the best. So what we are searching for is a stopping strategy that gives the best result in the long run, i.e. maximizes the *expected profit* when the inflation is taken into account.

This is an optimal stopping problem. It turns out that the solution can be expressed in terms of the solution of a corresponding boundary value problem (Problem 4), except that the boundary is unknown (free) as well and this is compensated by a double set of boundary conditions.

(E) Stochastic control

PROBLEM 6. An optimal portfolio problem.

Suppose that a person has two possible investments:

- i) A *risky* investment, where the price/value $p_1(t)$ at time t satisfies a stochastic differential equation of the type discussed in Problem 1:

$$\frac{dp_1}{dt} = (a + \alpha \cdot \text{"noise"})p_1 \quad (1.5)$$

where $a > 0$ and $\alpha \in \mathbf{R}$ are constants

- ii) A *safe* investment, where the price $p_2(t)$ at time t grows exponentially:

$$\frac{dp_2}{dt} = bp_2 \quad (1.6)$$

where b is a constant, $0 < b < a$.

At each instant t the person can choose how large portion (fraction) $u_t \in [0, 1]$ of his fortune X_t he wants to place in the risky investment, thereby placing $(1 - u_t)X_t$ in the safe investment. Given a utility function K and a terminal time T the problem is to find the *optimal portfolio* $u_t \in [0, 1]$ i.e. find the investment distribution $u_t; 0 \leq t \leq T$ which maximizes the expected utility of the corresponding terminal fortune $X_T^{(u)}$:

$$\max_{0 \leq u_t \leq 1} \{E[K(X_T^{(u)})]\}. \quad (1.7)$$

We will return to these problems in later chapters, after having developed the necessary mathematical machinery. We solve Problem 1 and Problem 2 in Chapter V. Problems involving filtering (Problem 3) are treated in Chapter VI, the generalized Dirichlet problem (Problem 4) in Chapter IX. Problem 5 is solved in Chapter X while stochastic control problems (Problem 6) are discussed in Chapter XI.

II. Some Mathematical Preliminaries

Having stated the problems we would like to solve, we now proceed to find reasonable mathematical notions corresponding to the quantities mentioned and mathematical models for the problems. In short, here is a first list of the notions that need a mathematical interpretation:

- (1) A random quantity
- (2) Independence
- (3) Parametrized (discrete or continuous) families of random quantities
- (4) What is meant by a “best” estimate in the filtering problem (Problem 3)
- (5) What is meant by an estimate “based on” some observations (Problem 3)?
- (6) What is the mathematical interpretation of the “noise” terms?
- (7) What is the mathematical interpretation of the stochastic differential equations?

In this chapter we will discuss (1) - (3) briefly. In the next chapter (III) we will consider (6), which leads to the notion of an Ito stochastic integral (7). In Chapter VI we will consider (4)-(5).

The mathematical model for a random quantity is a *random variable*. Before we define this, we recall some concepts from general probability theory. The reader is referred to e.g. Williams (1991) for more information.

DEFINITION 2.1. If Ω is a given set, then a σ -algebra \mathcal{F} on Ω is a family \mathcal{F} of subsets of Ω with the following properties:

- (i) $\emptyset \in \mathcal{F}$
- (ii) $F \in \mathcal{F} \Rightarrow F^C \in \mathcal{F}$, where $F^C = \Omega \setminus F$ is the complement of F in Ω
- (iii) $A_1, A_2, \dots \in \mathcal{F} \Rightarrow A := \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

The pair (Ω, \mathcal{F}) is called a *measurable space*. A *probability measure* P on a measurable space (Ω, \mathcal{F}) is a function $P : \mathcal{F} \rightarrow [0, 1]$ such that

- (a) $P(\emptyset) = 0, P(\Omega) = 1$
- (b) if $A_1, A_2, \dots \in \mathcal{F}$ and $\{A_i\}_{i=1}^{\infty}$ is disjoint (i.e. $A_i \cap A_j = \emptyset$ if $i \neq j$) then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

The triple (Ω, \mathcal{F}, P) is called a *probability space*. It is called a *complete probability space* if \mathcal{F} contains all subsets G of Ω with P -outer measure zero, i.e. with

$$P^*(G) := \inf\{P(F); F \in \mathcal{F}, G \subset F\} = 0$$

Any probability space can be made complete simply by adding to \mathcal{F} all sets of outer measure 0 and by extending P accordingly.

The subsets F of Ω which belong to \mathcal{F} are called \mathcal{F} -measurable sets. In a probability context these sets are called *events* and we use the interpretation

$$P(F) = \text{"the probability that the event } F \text{ occurs"}$$

In particular, if $P(F) = 1$ we say that " F occurs with probability 1, or almost surely (a.s.)"

Given any family \mathcal{U} of subsets of Ω there is a smallest σ -algebra $\mathcal{H}_{\mathcal{U}}$ containing \mathcal{U} , namely

$$\mathcal{H}_{\mathcal{U}} = \bigcap\{\mathcal{H}; \mathcal{H} \text{ } \sigma\text{-algebra of } \Omega, \mathcal{U} \subset \mathcal{H}\}$$

(See Exercise 2.2)

We call $\mathcal{H}_{\mathcal{U}}$ the σ -algebra generated by \mathcal{U} .

For example, if \mathcal{U} is the collection of all open subsets of a topological space Ω (e.g. $\Omega = \mathbf{R}^n$), the $\mathcal{B} = \mathcal{H}_{\mathcal{U}}$ is called the *Borel σ -algebra* on Ω and the elements $B \in \mathcal{B}$ are called *Borel sets*. \mathcal{B} contains all open sets, all closed sets, all countable unions of closed sets, all countable intersections of such countable unions etc.

If (Ω, \mathcal{F}, P) is a given probability space, then a function $Y : \Omega \rightarrow \mathbf{R}^n$ is called \mathcal{F} -measurable if

$$Y^{-1}(U) := \{\omega \in \Omega; Y(\omega) \in U\} \in \mathcal{F}$$

for all open sets $U \in \mathbf{R}^n$ (or, equivalently, for all Borel sets $U \subset \mathbf{R}^n$).

If $X : \Omega \rightarrow \mathbf{R}^n$ is any function, then the σ -algebra \mathcal{H}_X generated by X is the smallest σ -algebra on Ω containing all the sets

$$X^{-1}(U); \quad U \subset \mathbf{R}^n \text{ open}$$

It is not hard to show that

$$\mathcal{H}_X = \{X^{-1}(B); B \in \mathcal{B}\},$$

where \mathcal{B} is the Borel σ -algebra on \mathbf{R}^n . Clearly, X will then be \mathcal{H}_X -measurable and \mathcal{H}_X is the smallest σ -algebra with this property.

The following result is useful. It is a special case of a result sometimes called the *Doob-Dynkin lemma*. See e.g. M. M. Rao (1984), Prop. 3, p. 7.

LEMMA 2.1. If $X, Y : \Omega \rightarrow \mathbf{R}^n$ are two given functions, then Y is \mathcal{H}_X -measurable if and only if there exists a Borel measurable function $g : \mathbf{R}^n \rightarrow \mathbf{R}^n$ such that

$$Y = g(X).$$

In the following we let (Ω, \mathcal{F}, P) denote a given complete probability space. A random variable X is an \mathcal{F} -measurable function $X : \Omega \rightarrow \mathbf{R}^n$. Every random variable induces a probability measure μ_X on \mathbf{R}^n , defined by

$$\mu_X(B) = P(X^{-1}(B)).$$

μ_X is called the *distribution* of X .

If $\int_{\Omega} |X(\omega)| dP(\omega) < \infty$ then the number

$$E[X] := \int_{\Omega} X(\omega) dP(\omega) = \int_{\mathbf{R}^n} x d\mu_X(x)$$

is called the *expectation* of X (w.r.t. P).

More generally, if $f : \mathbf{R}^n \rightarrow \mathbf{R}$ is Borel measurable and $\int_{\Omega} |f(X(\omega))| dP(\omega) < \infty$ then we have

$$E[f(X)] := \int_{\Omega} f(x(\omega)) dP(\omega) = \int_{\mathbf{R}^n} f(x) d\mu_X(x).$$

The mathematical model for independence is the following:

DEFINITION 2.2. Two subsets $A, B \in \mathcal{F}$ are called *independent* if

$$P(A \cap B) = P(A) \cdot P(B).$$

A collection $\mathcal{A} = \{\mathcal{H}_i; i \in I\}$ of families \mathcal{H}_i of measurable sets is *independent* if

$$P(H_{i_1} \cap \cdots \cap H_{i_k}) = P(H_{i_1}) \cdots P(H_{i_k})$$

for all choices of $H_{i_1} \in \mathcal{H}_{i_1}, \dots, H_{i_k} \in \mathcal{H}_{i_k}$ with different indices i_1, \dots, i_k .

A collection of random variables $\{X_i; i \in I\}$ is *independent* if the collection of induced σ -algebras \mathcal{H}_{X_i} is independent.

If two random variables $X, Y : \Omega \rightarrow \mathbf{R}$ are independent then

$$E[XY] = E[X]E[Y],$$

provided that $E[|XY|] < \infty$, $E[|X|] < \infty$ and $E[|Y|] < \infty$. (See Exercise 2.4.)

DEFINITION 2.3. A *stochastic process* is a parametrized collection of random variables

$$\{X_t\}_{t \in T}$$

defined on a probability space (Ω, \mathcal{F}, P) and assuming values in \mathbf{R}^n . The parameter space T is usually (as in this book) the halfline $[0, \infty)$, but it may also be an interval $[a, b]$, the non-negative integers and even subsets of \mathbf{R}^n for $n \geq 1$. Note that for each $t \in T$ fixed we have a random variable

$$\omega \rightarrow X_t(\omega); \omega \in \Omega.$$

On the other hand, fixing $\omega \in \Omega$ we can consider the function

$$t \rightarrow X_t(\omega); t \in T$$

which is called a *path* of X_t .

It may be useful for the intuition to think of t as “time” and each ω as an individual “particle” or “experiment”. With this picture $X_t(\omega)$ would represent the position (or result) at time t of the particle (experiment) ω . Sometimes it is convenient to write $X(t, \omega)$ instead of $X_t(\omega)$. Thus we may also regard the process as a function of two variables

$$(t, \omega) \rightarrow X(t, \omega)$$

from $T \times \Omega$ into \mathbf{R}^n . This is often a natural point of view in stochastic analysis, because (as we shall see) there it is crucial to have $X(t, \omega)$ jointly measurable in (t, ω) .

Finally we note that we may identify each ω with the function $t \rightarrow X_t(\omega)$ from T into \mathbf{R}^n . Thus we may regard Ω as a subset of the space $\tilde{\Omega} = (\mathbf{R}^n)^T$ of all functions from T into \mathbf{R}^n . Then the σ -algebra \mathcal{F} will contain the σ -algebra \mathcal{B} generated by sets of the form

$$\{\omega; \omega(t_1) \in F_1, \dots, \omega(t_k) \in F_k\}, F_i \subset \mathbf{R}^n \text{ Borel sets}$$

(\mathcal{B} is the same as the Borel σ -algebra on $\tilde{\Omega}$ if $T = [0, \infty)$ and $\tilde{\Omega}$ is given the product topology). Therefore one may also adopt the point of view that a stochastic process is a *probability measure* P on the measurable space $((\mathbf{R}^n)^T, \mathcal{B})$.

The (*finite-dimensional*) *distributions* of the process $X = \{X_t\}_{t \in T}$ are the measures μ_{t_1, \dots, t_k} defined on \mathbf{R}^{nk} , $k = 1, 2, \dots$, by

$$\mu_{t_1, \dots, t_k}(F_1 \times F_2 \times \dots \times F_k) = P[X_{t_1} \in F_1, \dots, X_{t_k} \in F_k]; t_i \in T.$$

Here F_1, \dots, F_k denote Borel sets in \mathbf{R}^n .

The family of all finite-dimensional distributions determine many (but not all) important properties of the process X .

Conversely, given a family $\{\nu_{t_1, \dots, t_k}; k \in \mathbf{N}, t_i \in T\}$ of probability measures on \mathbf{R}^{nk} it is important to be able to construct a stochastic process $Y = \{Y_t\}_{t \in T}$ having ν_{t_1, \dots, t_k} as its finite-dimensional distributions. One of Kolmogorov's famous theorems states that this can be done provided $\{\nu_{t_1, \dots, t_k}\}$ satisfies two natural consistency conditions: (See Lamperti (1977))

THEOREM 2.4 (Kolmogorov's extension theorem).

For all $t_1, \dots, t_k \in T, k \in \mathbf{N}$ let ν_{t_1, \dots, t_k} be probability measures on \mathbf{R}^{nk} s.t.

$$\nu_{t_{\sigma(1)}, \dots, t_{\sigma(k)}}(F_1 \times \dots \times F_k) = \nu_{t_1, \dots, t_k}(F_{\sigma^{-1}(1)} \times \dots \times F_{\sigma^{-1}(k)}) \quad (K1)$$

for all permutations σ on $\{1, 2, \dots, k\}$ and

$$\nu_{t_1, \dots, t_k}(F_1 \times \dots \times F_k) = \nu_{t_1, \dots, t_k, t_{k+1}, \dots, t_{k+m}}(F_1 \times \dots \times F_k \times \mathbf{R}^n \times \dots \times \mathbf{R}^n) \quad (K2)$$

for all $m \in \mathbf{N}$, where (of course) the set on the right hand side has a total of $k + m$ factors.

Then there exists a probability space (Ω, \mathcal{F}, P) and a stochastic process $\{X_t\}$ on $\Omega, X_t : \Omega \rightarrow \mathbf{R}^n$, s.t.

$$\nu_{t_1, \dots, t_k}(F_1 \times \dots \times F_k) = P[X_{t_1} \in F_1, \dots, X_{t_k} \in F_k],$$

for all $t_i \in T, k \in \mathbf{N}$ and all Borel sets F_i .

An important example: Brownian motion

In 1828 the Scottish botanist Robert Brown observed that pollen grains suspended in liquid performed an irregular motion. The motion was later explained by the random collisions with the molecules of the liquid. To describe the motion mathematically it is natural to use the concept of a stochastic process $B_t(\omega)$, interpreted as the position at time t of the pollen grain ω . We will generalize slightly and consider an n-dimensional analog.

To construct $\{B_t\}_{t \geq 0}$ it suffices, by the Kolmogorov extension theorem, to specify a family $\{\nu_{t_1, \dots, t_k}\}$ of probability measures satisfying (K1) and (K2). These measures will be chosen so that they agree with our observations of the pollen grain behaviour:

Fix $x \in \mathbf{R}^n$ and define

$$p(t, x, y) = (2\pi t)^{-n/2} \cdot \exp\left(-\frac{|x - y|^2}{2t}\right) \quad \text{for } y \in \mathbf{R}^n, t > 0$$

If $0 \leq t_1 \leq t_2 \leq \cdots \leq t_k$ define a measure ν_{t_1, \dots, t_k} on \mathbf{R}^{nk} by

$$\nu_{t_1, \dots, t_k}(F_1 \times \cdots \times F_k) = \int_{F_1 \times \cdots \times F_k} p(t_1, x, x_1)p(t_2 - t_1, x_1, x_2) \cdots p(t_k - t_{k-1}, x_{k-1}, x_k) dx_1 \cdots dx_k \quad (2.1)$$

where we use the notation $dy = dy_1 \cdots dy_k$ for Lebesgue measure and the convention that $p(0, x, y)dy = \delta_x(y)$, the unit point mass at x .

Extend this definition to all finite sequences of t_i 's by using (K1). Since $\int_{\mathbf{R}^n} p(t, x, y)dy = 1$ for all $t \geq 0$, (K2) holds, so by Kolmogorov's theorem

there exists a probability space $(\Omega, \mathcal{F}, P^x)$ and a stochastic process $\{B_t\}_{t \geq 0}$ on Ω such that the finite-dimensional distributions of B_t are given by (2.1), i.e.

$$P^x(B_{t_1} \in F_1, \dots, B_{t_k} \in F_k) = \int_{F_1 \times \cdots \times F_k} p(t_1, x, x_1) \cdots p(t_k - t_{k-1}, x_{k-1}, x_k) dx_1 \cdots dx_k. \quad (2.2)$$

Such a process is called (a version of) *Brownian motion starting at x* (observe that $P^x(B_0 = x) = 1$).

The Brownian motion thus defined is not unique, i.e. there exist several quadruples $(B_t, \Omega, \mathcal{F}, P^x)$ such that (2.2) holds. However, for our purposes this is not important, we may simply choose any version to work with. As we shall soon see, the paths of a Brownian motion are (or, more correctly, can be chosen to be) continuous, a.s. Therefore we may identify (a.a.) $\omega \in \Omega$ with a continuous function $t \rightarrow B_t(\omega)$ from $[0, \infty)$ into \mathbf{R}^n . Thus we may adopt the point of view that Brownian motion is just the space $C([0, \infty), \mathbf{R}^n)$ equipped with certain probability measures P^x (given by (2.1) and (2.2) above). This version is called the *canonical* Brownian motion. Besides having the advantage of being intuitive, this point of view is useful for the further analysis of measures on $C([0, \infty), \mathbf{R}^n)$, since this space is Polish (i.e. a complete separable metric space). See Stroock and Varadhan (1979).

We state some basic properties of Brownian motion:

- (i) B_t is a *Gaussian process*, i.e. for all $0 \leq t_1 \leq \cdots \leq t_k$ the random variable $Z = (B_{t_1}, \dots, B_{t_k}) \in \mathbf{R}^{nk}$ has a *(multi)normal distribution*. This means that there exists a vector $M \in \mathbf{R}^{nk}$ and a non-negative definite matrix $C = [c_{jm}] \in \mathbf{R}^{nk \times nk}$ (the set of all $nk \times nk$ -matrices with real entries) such that

$$E^x[\exp(i \sum_{j=1}^{nk} u_j Z_j)] = \exp(-\frac{1}{2} \sum_{j,m} u_j c_{jm} u_m + i \sum_j u_j M_j) \quad (2.3)$$

for all $u = (u_1, \dots, u_{nk}) \in \mathbf{R}^{nk}$, where $i = \sqrt{-1}$ is the imaginary unit and E^x denotes expectation with respect to P^x . Moreover, if (2.3) holds then

$$M = E^x[Z] \text{ is the mean value of } Z \quad (2.4)$$

and

$$c_{jm} = E^x[(Z_j - M_j)(Z_m - M_m)] \text{ is the covariance matrix of } Z. \quad (2.5)$$

(See Appendix A).

To see that (2.3) holds for $Z = (B_{t_1}, \dots, B_{t_k})$ we calculate its left hand side explicitly by using (2.2) (see Appendix A) and obtain (2.3) with

$$M = E^x[Z] = (x, x, \dots, x) \in \mathbf{R}^{nk} \quad (2.6)$$

and

$$C = \begin{pmatrix} t_1 I_n & t_1 I_n & \cdots & t_1 I_n \\ t_1 I_n & t_2 I_n & \cdots & t_2 I_n \\ \vdots & \vdots & & \vdots \\ t_1 I_n & t_2 I_n & \cdots & t_k I_n \end{pmatrix} \quad (2.7)$$

Hence

$$E^x[B_t] = x \text{ for all } t \geq 0 \quad (2.8)$$

and

$$E^x[(B_t - x)^2] = nt, E^x[(B_t - x)(B_s - x)] = n \min(s, t). \quad (2.9)$$

Moreover,

$$E^x[(B_t - B_s)^2] = n(t - s) \text{ if } t \geq s, \quad (2.10)$$

since

$$\begin{aligned} E^x[(B_t - B_s)^2] &= E^x[(B_t - x)^2 - 2(B_t - x)(B_s - x) + (B_s - x)^2] \\ &= n(t - 2s + s) = n(t - s), \text{ when } t \geq s. \end{aligned}$$

(ii) B_t has *independent increments*, i.e.

$$\begin{aligned} B_{t_1}, B_{t_2} - B_{t_1}, \dots, B_{t_k} - B_{t_{k-1}} \text{ are independent} \\ \text{for all } 0 \leq t_1 < t_2 \dots < t_k. \end{aligned} \quad (2.11)$$

To prove this we use the fact that normal random variables are independent iff they are uncorrelated. (See Appendix A). So it is enough to prove that

$$E^x[(B_{t_i} - B_{t_{i-1}})(B_{t_j} - B_{t_{j-1}})] = 0 \text{ when } t_i < t_j, \quad (2.12)$$

which follows from the form of C :

$$\begin{aligned} E^x[B_{t_i} B_{t_j} - B_{t_{i-1}} B_{t_j} - B_{t_i} B_{t_{j-1}} + B_{t_{i-1}} B_{t_{j-1}}] \\ = n(t_i - t_{i-1} - t_j + t_{j-1}) = 0. \end{aligned}$$

(iii) Finally we ask: Is $t \rightarrow B_t(\omega)$ continuous for almost all ω ? Stated like this the question does not make sense, because the set $H = \{\omega; t \rightarrow B_t(\omega)$ is continuous} is not measurable with respect to the Borel σ -algebra \mathcal{B} on $(\mathbf{R}^n)^{[0,\infty)}$ mentioned above (H involves an uncountable number of t's). However, if modified slightly the question can be given a positive answer. To explain this we need the following important concept:

DEFINITION 2.5. Suppose that $\{X_t\}$ and $\{Y_t\}$ are stochastic processes on (Ω, \mathcal{F}, P) . Then we say that $\{X_t\}$ is a version of $\{Y_t\}$ if

$$P(\{\omega; X_t(\omega) = Y_t(\omega)\}) = 1 \text{ for all } t.$$

Note that if X_t is a version of Y_t , then X_t and Y_t have the same finite-dimensional distributions. Thus from the point of view that a stochastic process is a probability law on $(\mathbf{R}^n)^{[0,\infty)}$ two such processes are indistinguishable, but nevertheless their path properties may be different.

The continuity question of Brownian motion can be answered by using another famous theorem of Kolmogorov:

THEOREM 2.6. (Kolmogorov's continuity theorem). Suppose that the process $X = \{X_t\}_{t \geq 0}$ satisfies the following condition: For all $T > 0$ there exist positive constants α, β, D such that

$$E[|X_t - X_s|^\alpha] \leq D \cdot |t - s|^{1+\beta}; 0 \leq s, t \leq T. \quad (2.13)$$

Then there exists a continuous version of X .

(See for example Stroock and Varadhan (1979, p. 51)).

For Brownian motion B_t it is not hard to prove that

$$E^x[|B_t - B_s|^4] = n(n+2)|t-s|^2. \quad (2.14)$$

So Brownian motion satisfies Kolmogorov's condition (2.13) with $\alpha = 4$, $D = n(n+2)$ and $\beta = 1$, and therefore it has a continuous version. From now on we will assume that B_t is such a continuous version.

Finally we note that

If $B_t = (B_t^{(1)}, \dots, B_t^{(n)})$ is n-dimensional Brownian motion, then the 1-dimensional processes $\{B_t^{(j)}\}_{t \geq 0}$, $1 \leq j \leq n$ are independent, 1-dimensional Brownian motions. (2.15)

EXERCISES

2.1.

Let $X : \Omega \rightarrow \mathbf{R}$ be a random variable. The *distribution function* F of X is defined by

$$F(x) = P[X \leq x].$$

a) Prove that F has the following properties:

- (i) $0 \leq F \leq 1$, $\lim_{r \rightarrow -\infty} F(x) = 0$, $\lim_{x \rightarrow \infty} F(x) = 1$.
- (ii) F is increasing (= non-decreasing).
- (iii) F is right-continuous, i.e. $F(x) = \lim_{h \rightarrow 0^+} F(x + h)$.

b) Let $g : \mathbf{R} \rightarrow \mathbf{R}$ be measurable such that $E[|g(X)|] < \infty$. Prove that

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)dF(x),$$

where the integral on the right is interpreted in the Lebesgue-Stieltjes sense.

c) Let $p(x) \geq 0$ be a measurable function on \mathbf{R} . We say that X has the density p if

$$F(x) = \int_{-\infty}^x p(y)dy \quad \text{for all } x.$$

Thus from (2.1)–(2.2) we know that 1-dimensional Brownian motion B_t at time t with $B_0 = 0$ has the density

$$p(x) = \frac{1}{\sqrt{2\pi t}} \exp(-\frac{x^2}{2t}); \quad x \in \mathbf{R}.$$

Find the density of B_t^2 .

2.2.

Let $\{\mathcal{H}_i\}_{i \in I}$ be a family of σ -algebras on Ω . Prove that

$$\mathcal{H} = \bigcap \{\mathcal{H}_i; i \in I\}$$

is again a σ -algebra.

2.3.

a) Let $X : \Omega \rightarrow \mathbf{R}'$ be a random variable such that

$$E[|X|^p] < \infty \quad \text{for some } p, 0 < p < \infty.$$

Prove *Chebychev's inequality*:

$$P[|X| \geq \lambda] \leq \frac{1}{\lambda^p} E[|X|^p] \quad \text{for all } \lambda \geq 0.$$

Hint: $\int_{\Omega} |X|^p dP \geq \int_A |X|^p dP$, where $A = \{\omega : |X| \geq \lambda\}$.

- b) Suppose there exists $k > 0$ such that

$$M = E[\exp(k|X|)] < \infty.$$

Prove that $P[|X| \geq \lambda] \leq M e^{-k\lambda}$ for all $\lambda \geq 0$.

2.4.

Let $X, Y : \Omega \rightarrow \mathbf{R}$ be two independent random variables and assume for simplicity that X and Y are bounded. Prove that

$$E[XY] = E[X]E[Y].$$

(Hint: Assume $|X| \leq M$, $|Y| \leq N$. Approximate X and Y by simple functions $\varphi(\omega) = \sum_{i=1}^m a_i \chi_{F_i}(\omega)$, $\psi(\omega) = \sum_{j=1}^n b_j \chi_{G_j}(\omega)$, respectively, where $F_i = X^{-1}([a_i, a_{i+1}))$, $G_j = Y^{-1}([b_j, b_{j+1}))$, $-M = a_0 < a_1 < \dots < a_m = M$, $-N = b_0 < b_1 < \dots < b_n = N$. Then

$$E[X] \approx E[\varphi] = \sum_i a_i P(F_i), \quad E[Y] \approx E[\psi] = \sum_j b_j P(G_j)$$

and

$$E[XY] \approx E[\varphi\psi] = \sum_{i,j} a_i b_j P(F_i \cap G_j) \dots \Big)$$

2.5.

Let (Ω, \mathcal{F}, P) be a probability space and let A_1, A_2, \dots be sets in \mathcal{F} such that

$$\sum_{k=1}^{\infty} P(A_k) < \infty.$$

Prove the *Borel-Cantelli lemma*:

$$P\left(\bigcap_{m=1}^{\infty} \bigcup_{k=m}^{\infty} A_k\right) = 0,$$

i.e. the probability that ω belongs to infinitely many A'_k 's is zero.

2.6.

- a) Suppose G_1, G_2, \dots, G_n are disjoint subsets of Ω such that

$$\Omega = \bigcup_{i=1}^n G_i.$$

Prove that the family \mathcal{G} consisting of \emptyset and all unions of some (or all) of G_1, \dots, G_n constitutes a σ -algebra on Ω .

- b) Prove that any *finite* σ -algebra \mathcal{F} on Ω is of the type described in a).
- c) Let \mathcal{F} be a *finite* σ -algebra on Ω and let $X : \Omega \rightarrow \mathbf{R}$ be \mathcal{F} -measurable. Prove that X assumes only finitely many possible values. More precisely, there exists a disjoint family of subsets $F_1, \dots, F_m \in \mathcal{F}$ and real numbers c_1, \dots, c_m such that

$$X(\omega) = \sum_{i=1}^m c_i \chi_{F_i}(\omega).$$

2.7.

Let B_t be Brownian motion on \mathbf{R} , $B_0 = 0$. Put $E = E^0$.

- a) Use (2.3) to prove that

$$E[e^{iuB_t}] = \exp\left(-\frac{1}{2}u^2t\right) \quad \text{for all } u \in \mathbf{R}$$

- b) Use the power series expansion of the exponential function on both sides, compare the terms with the same power of u and deduce that

$$E[B_t^4] = 3t^2$$

and more generally that

$$E[B_t^{2k}] = \frac{(2k)!}{2^k \cdot k!} t^k; \quad k \in \mathbf{N}.$$

- c) If you feel uneasy about the lack of rigour in the method in b), you can proceed as follows: Prove that (2.1) implies that

$$E[f(B_t)] = \frac{1}{\sqrt{2\pi t}} \int_{\mathbf{R}} f(x) e^{-\frac{x^2}{2t}} dx$$

for all functions f such that the integral on the right converges. Then apply this to $f(x) = x^{2k}$ and use integration by parts and induction on k .

d) Prove (2.14), for example by using b) and induction on n .

2.8.

To illustrate that the (finite-dimensional) distributions alone do not give all the information regarding the continuity properties of a process, consider the following example:

Let $(\Omega, \mathcal{F}, P) = ([0, \infty), \mathcal{B}, d\mu)$ where \mathcal{B} denotes the Borel σ -algebra on $[0, \infty)$ and $d\mu$ is a probability measure on $[0, \infty)$ with no mass on single points. Define

$$X_t(\omega) = \begin{cases} 1 & \text{if } t = \omega \\ 0 & \text{otherwise} \end{cases}$$

and

$$Y_t(\omega) = 0 \quad \text{for all } (t, \omega) \in [0, \infty) \times [0, \infty).$$

Prove that $\{X_t\}$ and $\{Y_t\}$ have the same distributions and that X_t is a version of Y_t . And yet we have that $t \rightarrow Y_t(\omega)$ is continuous for all ω , while $t \rightarrow X_t(\omega)$ is discontinuous for all ω .

2.9.

A stochastic process X_t is called *stationary* if $\{X_t\}$ has the same distribution as $\{X_{t+h}\}$ for any $h > 0$. Prove that Brownian motion B_t has stationary increments, i.e. that the process $\{B_{t+h} - B_t\}_{h \geq 0}$ has the same distribution for all t .

2.10.

Prove (2.15).

2.11.

Let B_t be Brownian motion and fix $t_0 \geq 0$. Prove that

$$\tilde{B}_t := B_{t_0+t} - B_{t_0}; t \geq 0$$

is a Brownian motion.

2.12.

Let B_t be 2-dimensional Brownian motion and put

$$D_\rho = \{x \in \mathbf{R}^2; |x| < \rho\} \text{ for } \rho > 0.$$

Compute

$$P^0[B_t \in D_\rho].$$

2.13.

Let B_t be n-dimensional Brownian motion and let $K \subset \mathbf{R}^n$ have zero n-dimensional Lebesgue measure. Prove that the expected total length of

time that B_t spends in K is zero. (This implies that the Green measure associated with B_t is absolutely continuous with respect to Lebesgue measure. See Ch. IX).

2.14.

Let B_t be n-dimensional Brownian motion starting at 0 and let $U \in \mathbf{R}^{n \times n}$ be a (constant) orthogonal matrix, i.e. $UU^T = I$. Prove that

$$\tilde{B}_t := UB_t$$

is also a Brownian motion.

2.15.

If $X_t(\cdot) : \Omega \rightarrow \mathbf{R}$ is a continuous stochastic process, then for $p > 0$ the p -th variation process of X_t , $\langle X, X \rangle_t^{(p)}$ is defined by

$$\langle X, X \rangle_t^{(p)}(\omega) = \limsup_{\Delta t_k \rightarrow 0} \sum_{t_k \leq t} |X_{t_{k+1}}(\omega) - X_{t_k}(\omega)|^p$$

where $0 = t_1 < t_2 < \dots < t_n = t$ and $\Delta t_k = t_{k+1} - t_k$. In particular, if $p = 1$ this process is called the *total variation process* and if $p = 2$ this is called the *quadratic variation process*. For Brownian motion $B_t \in \mathbf{R}$ one can show that the quadratic variation process is simply

$$\langle B, B \rangle_t(\omega) = \langle B, B \rangle_t^{(2)}(\omega) = t \text{ a.s.}$$

(See Exercise 4.7). Here we will settle with a weaker result:

a) Define

$$t_k^{(n)} = k \cdot 2^{-n}t; \quad n = 1, 2, \dots, \quad k = 0, 1, \dots, 2^n$$

and put

$$Y_n(t, \omega) = \sum_{k=0}^{2^n} (\Delta B_k(\omega))^2,$$

where

$$\Delta B_k = B_{t_{k+1}^{(n)}} - B_{t_k^{(n)}}; \quad k = 0, \dots, 2^n - 1.$$

Show that $Y_n(t, \cdot) \rightarrow t$ in $L^2(P)$ as $n \rightarrow \infty$. Hence deduce that there is a subsequence n_j ; $j = 1, 2, \dots$ such that

$$Y_{n_j}(t, \omega) \rightarrow t \text{ as } j \rightarrow \infty, \text{ for a.a. } \omega.$$

b) Use a) to prove that a.a. paths of Brownian motion do not have a bounded variation on $[0, t]$, i.e. the total variation of Brownian motion is infinite, a.s.

III. Ito Integrals

We now turn to the question of finding a reasonable mathematical interpretation of the “noise” term in the equation of Problem 1 in the Introduction:

$$\frac{dN}{dt} = (r(t) + \text{“noise”})N(t)$$

or more generally in equations of the form

$$\frac{dX}{dt} = b(t, X_t) + \sigma(t, X_t) \cdot \text{“noise”}, \quad (3.1)$$

where b and σ are some given functions. Let us first concentrate on the case when the noise is 1-dimensional. It is reasonable to look for some stochastic process W_t to represent the noise term, so that

$$\frac{dX}{dt} = b(t, X_t) + \sigma(t, X_t) \cdot W_t. \quad (3.2)$$

Based on many situations, for example in engineering, one is led to assume that W_t has, at least approximately, these properties:

- (i) $t_1 \neq t_2 \Rightarrow W_{t_1}$ and W_{t_2} are independent
- (ii) $\{W_t\}$ is stationary, i.e. the (joint) distribution of $\{W_{t_1+t}, \dots, W_{t_k+t}\}$ does not depend on t .
- (iii) $E[W_t] = 0$ for all t .

However, it turns out there does *not* exist any “reasonable” stochastic process satisfying (i) and (ii): Such a W_t cannot have continuous paths. (See Exercise 3.9.) If we require $E[W_t^2] = 1$ then the function $(t, \omega) \rightarrow W_t(\omega)$ cannot even be measurable, with respect to the σ -algebra $\mathcal{B} \times \mathcal{F}$, where \mathcal{B} is the Borel σ -algebra on $[0, \infty]$. (See Kallianpur (1980, p. 10).)

Nevertheless it is possible to represent W_t as a generalized stochastic process called the *white noise process*.

That the process is *generalized* means that it can be constructed as a probability measure on the space \mathcal{S}' of tempered distributions on $[0, \infty)$, and not as a probability measure on the much smaller space $\mathbf{R}^{[0, \infty)}$, like an ordinary process

can. See e.g. Hida (1980), Adler (1981), Rozanov (1982) or Hida, Kuo, Potthoff and Streit (1993).

We will avoid this kind of construction and rather try to rewrite equation (3.2) in a form that suggests a replacement of W_t by a proper stochastic process: Let $0 = t_0 < t_1 < \dots < t_m = t$ and consider a discrete version of (3.2):

$$X_{k+1} - X_k = b(t_k, X_k) \Delta t_k + \sigma(t_k, X_k) W_k \Delta t_k, \quad (3.3)$$

where

$$X_j = X(t_j), W_k = W_{t_k}, \Delta t_k = t_{k+1} - t_k.$$

We abandon the W_k -notation and replace $W_k \Delta t_k$ by $\Delta V_k = V_{k+1} - V_k$, where $\{V_t\}_{t \geq 0}$ is some suitable stochastic process. The assumptions (i), (ii) and (iii) on W_t suggest that V_t should have *stationary independent increments with mean 0*. It turns out that the only such process with continuous paths is the Brownian motion B_t . (See Knight (1981)). Thus we put $V_t = B_t$ and obtain from (3.3):

$$X_k = X_0 + \sum_{j=0}^{k-1} b(t_j, X_j) \Delta t_j + \sum_{j=0}^{k-1} \sigma(t_j, X_j) \Delta B_j. \quad (3.4)$$

Is it possible to prove that the limit of the right hand side of (3.4) exists, in some sense, when $\Delta t_j \rightarrow 0$? If so, then by applying the usual integration notation we should obtain

$$X_t = X_0 + \int_0^t b(s, X_s) ds + \left(\int_0^t \sigma(s, X_s) dB_s \right) \quad (3.5)$$

and we would adopt as a convention that (3.2) really means that $X_t = X_t(\omega)$ is a stochastic process satisfying (3.5).

Thus, in the remainder of this chapter we will prove the existence, in a certain sense, of

$$\left(\int_0^t f(s, \omega) dB_s(\omega) \right)$$

where $B_t(\omega)$ is 1-dimensional Brownian motion starting at the origin, for a wide class of functions $f : [0, \infty] \times \Omega \rightarrow \mathbf{R}$. Then, in Chapter V, we will return to the solution of (3.5).

Suppose $0 \leq S < T$ and $f(t, \omega)$ is given. We want to define

$$\int_S^T f(t, \omega) dB_t(\omega). \quad (3.6)$$

It is reasonable to start with a definition for a simple class of functions f and then extend by some approximation procedure. Thus, let us first assume that f has the form

$$\phi(t, \omega) = \sum_{j \geq 0} e_j(\omega) \cdot \mathcal{X}_{[j \cdot 2^{-n}, (j+1)2^{-n})}(t), \quad (3.7)$$

where \mathcal{X} denotes the characteristic (indicator) function and n is a natural number. For such functions it is reasonable to define

$$\int_S^T \phi(t, \omega) dB_t(\omega) = \sum_{j \geq 0} e_j(\omega) [B_{t_{j+1}} - B_{t_j}](\omega), \quad (3.8)$$

where

$$t_k = t_k^{(n)} = \begin{cases} k \cdot 2^{-n} & \text{if } S \leq k \cdot 2^{-n} \leq T \\ S & \text{if } k \cdot 2^{-n} < S \\ T & \text{if } k \cdot 2^{-n} > T \end{cases}$$

However, without any further assumptions on the functions $e_j(\omega)$ this leads to difficulties, as the next example shows.

EXAMPLE 3.1. Choose

$$\begin{aligned} \phi_1(t, \omega) &= \sum_{j \geq 0} B_{j \cdot 2^{-n}}(\omega) \cdot \mathcal{X}_{[j \cdot 2^{-n}, (j+1)2^{-n})}(t) \\ \phi_2(t, \omega) &= \sum_{j \geq 0} B_{(j+1)2^{-n}}(\omega) \cdot \mathcal{X}_{[j \cdot 2^{-n}, (j+1)2^{-n})}(t) \end{aligned}$$

Then

$$E \left[\int_0^T \phi_1(t, \omega) dB_t(\omega) \right] = \sum_{j \geq 0} E[B_{t_j} (B_{t_{j+1}} - B_{t_j})] = 0,$$

since $\{B_t\}$ has independent increments. (Here-and in the following - E means the same as E^0 , the expectation w.r.t. the law P^0 for Brownian motion starting at 0.) But

$$\begin{aligned} E \left[\int_0^T \phi_2(t, \omega) dB_t(\omega) \right] &= \sum_{j \geq 0} E[B_{t_{j+1}} \cdot (B_{t_{j+1}} - B_{t_j})] \\ &= \sum_{j \geq 0} E[(B_{t_{j+1}} - B_{t_j})^2] = T, \text{ by (2.10).} \end{aligned}$$

So, in spite of the fact that both ϕ_1 and ϕ_2 appear to be very reasonable approximations to

$$f(t, \omega) = B_t(\omega),$$

their integrals according to (3.8) are not close to each other at all, no matter how large n is chosen.

This only reflects the fact that the variations of the paths of B_t are too big to enable us to define the integral (3.6) in the Riemann-Stieltjes sense. In fact one can show that the paths $t \rightarrow B_t$ of Brownian motion are nowhere differentiable, almost surely (a.s.). (See Breiman (1968)). In particular, the total variation of the path is infinite, a.s.

In general it is natural to approximate a given function $f(t, \omega)$ by

$$\sum_j f(t_j^*, \omega) \cdot \mathcal{X}_{[t_j, t_{j+1})}(t)$$

where the points t_j^* belong to the intervals $[t_j, t_{j+1}]$, and then define

$\int_S^T f(t, \omega) dB_t(\omega)$ as the limit (in a sense that we will explain) of $\sum_j f(t_j^*, \omega)[B_{t_{j+1}} - B_{t_j}](\omega)$ as $n \rightarrow \infty$. However, the example above shows that – unlike the Riemann-Stieltjes integral – it does make a difference here what points t_j^* we choose. The following two choices have turned out to be the most useful ones:

- 1) $t_j^* = t_j$ (the left end point), which leads to the *Ito integral*, from now on denoted by

$$\int_S^T f(t, \omega) dB_t(\omega),$$

and

- 2) $t_j^* = (t_j + t_{j+1})/2$ (the mid point), which leads to the *Stratonovich integral*, denoted by

$$\int_S^T f(t, \omega) \circ dB_t(\omega).$$

(See Protter (1990, Th. V. 5.30)).

In the end of this chapter we will explain why these choices are the best and discuss the relations and distinctions between the corresponding integrals.

In any case one must restrict oneself to a special class of functions $f(t, \omega)$ in (3.6), also if they have the particular form (3.7), in order to obtain a reasonable definition of the integral. We will here present Ito's choice $t_j^* = t_j$. The approximation procedure indicated above will work out successfully provided that f has the property that each of the functions $\omega \rightarrow f(t_j, \omega)$ only depends on the behaviour of $B_s(\omega)$ up to time t_j . This leads to the following important concepts:

DEFINITION 3.2. Let $B_t(\omega)$ be n-dimensional Brownian motion. Then we define $\mathcal{F}_t = \mathcal{F}_t^{(n)}$ to be the σ -algebra generated by the random variables $B_s(\cdot); s \leq t$. In other words, \mathcal{F}_t is the smallest σ -algebra containing all sets of the form

$$\{\omega; B_{t_1}(\omega) \in F_1, \dots, B_{t_k}(\omega) \in F_k\},$$

where $k = 1, 2, \dots, t_j \leq t$ and $F_j \subset \mathbf{R}^n$ are Borel sets. (We assume that all sets of measure zero are included in \mathcal{F}_t).

One often thinks of \mathcal{F}_t as “the history of B_s up to time t ”. A function $h(\omega)$ will be \mathcal{F}_t -measurable if and only if h can be written as the pointwise a.e. limit of sums of functions of the form

$$g_1(B_{t_1})g_2(B_{t_2}) \cdots g_k(B_{t_k}),$$

where g_1, \dots, g_k are bounded continuous functions and $t_j \leq t$ for $j \leq k$, $k = 1, 2, \dots$ (See Exercise 3.14.) Intuitively, that h is \mathcal{F}_t -measurable means that the value of $h(\omega)$ can be decided from the values of $B_s(\omega)$ for $s \leq t$. For example, $h_1(\omega) = B_{t/2}(\omega)$ is \mathcal{F}_t -measurable, while $h_2(\omega) = B_{2t}(\omega)$ is not.

Note that $\mathcal{F}_s \subset \mathcal{F}_t$ for $s < t$ (i.e. $\{\mathcal{F}_t\}$ is *increasing*) and that $\mathcal{F}_t \subset \mathcal{F}$ for all t .

DEFINITION 3.3. Let $\{\mathcal{N}_t\}_{t \geq 0}$ be an increasing family of σ -algebras of subsets of Ω . A process $g(t, \omega) : [0, \infty) \times \Omega \rightarrow \mathbf{R}^n$ is called \mathcal{N}_t -adapted if for each $t \geq 0$ the function

$$\omega \rightarrow g(t, \omega)$$

is \mathcal{N}_t -measurable.

Thus the process $h_1(t, \omega) = B_{t/2}(\omega)$ is \mathcal{F}_t -adapted, while $h_2(t, \omega) = B_{2t}(\omega)$ is not.

We now describe our class of functions for which the Ito integral will be defined:

DEFINITION 3.4. Let $\mathcal{V} = \mathcal{V}(S, T)$ be the class of functions

$$f(t, \omega) : [0, \infty) \times \Omega \rightarrow \mathbf{R}$$

such that

- (i) $(t, \omega) \rightarrow f(t, \omega)$ is $\mathcal{B} \times \mathcal{F}$ -measurable, where \mathcal{B} denotes the Borel σ -algebra on $[0, \infty)$.
- (ii) $f(t, \omega)$ is \mathcal{F}_t -adapted
- (iii) $E[\int_S^T f(t, \omega)^2 dt] < \infty$

The Ito integral

For functions $f \in \mathcal{V}$ we will now show how to define the *Ito integral*

$$\mathcal{I}[f](\omega) = \int_S^T f(t, \omega) dB_t(\omega),$$

where B_t is 1-dimensional Brownian motion.

The idea is natural: First we define $\mathcal{I}[\phi]$ for a simple class of functions ϕ . Then we show that each $f \in \mathcal{V}$ can be approximated (in an appropriate sense) by such ϕ 's and we use this to define $\int f dB$ as the limit of $\int \phi dB$ as $\phi \rightarrow f$.

We now give the details of this construction: A function $\phi \in \mathcal{V}$ is called elementary if it has the form

$$\phi(t, \omega) = \sum_j e_j(\omega) \cdot \chi_{[t_j, t_{j+1})}(t). \quad (3.9)$$

Note that since $\phi \in \mathcal{V}$ each function e_j must be \mathcal{F}_{t_j} -measurable. Thus in Example 1 above the function ϕ_1 is elementary while ϕ_2 is not.

For elementary functions $\phi(t, \omega)$ we define the integral according to (3.8), i.e.

$$\int_S^T \phi(t, \omega) dB_t(\omega) = \sum_{j \geq 0} e_j(\omega) [B_{t_{j+1}} - B_{t_j}](\omega). \quad (3.10)$$

Now we make the following important observation:

LEMMA 3.5. (The Ito isometry). If $\phi(t, \omega)$ is bounded and elementary then

$$E\left[\left(\int_S^T \phi(t, \omega) dB_t(\omega)\right)^2\right] = E\left[\int_S^T \phi(t, \omega)^2 dt\right]. \quad (3.11)$$

Proof: Put $\Delta B_j = B_{t_{j+1}} - B_{t_j}$. Then

$$E[e_i e_j \Delta B_i \Delta B_j] = \begin{cases} 0 & \text{if } i \neq j \\ E[e_j^2] \cdot (t_{j+1} - t_j) & \text{if } i = j \end{cases}$$

using that $e_i e_j \Delta B_i$ and ΔB_j are independent if $i < j$. Thus

$$\begin{aligned} E\left[\left(\int_S^T \phi dB\right)^2\right] &= \sum_{i,j} E[e_i e_j \Delta B_i \Delta B_j] = \sum_j E[e_j^2] \cdot (t_{j+1} - t_j) \\ &= E\left[\int_S^T \phi^2 dt\right]. \end{aligned}$$

The idea is now to use the isometry (3.11) to extend the definition from elementary functions to functions in \mathcal{V} . We do this in several steps:

STEP 1. Let $g \in \mathcal{V}$ be bounded and $g(\cdot, \omega)$ continuous for each ω . Then there exist elementary functions $\phi_n \in \mathcal{V}$ such that

$$E\left[\int_S^T (g - \phi_n)^2 dt\right] \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Proof: Define $\phi_n(t, \omega) = \sum_j g(t_j, \omega) \cdot \chi_{[t_j, t_{j+1})}(t)$. Then ϕ_n is elementary since $g \in \mathcal{V}$, and

$$\int_S^T (g - \phi_n)^2 dt \rightarrow 0 \text{ as } n \rightarrow \infty, \text{ for each } \omega,$$

since $g(\cdot, \omega)$ is continuous for each ω . Hence $E\left[\int_S^T (g - \phi_n)^2 dt\right] \rightarrow 0$ as $n \rightarrow \infty$, by bounded convergence.

STEP 2. Let $h \in \mathcal{V}$ be bounded. Then there exist bounded functions $g_n \in \mathcal{V}$ such that $g_n(\cdot, \omega)$ is continuous for all ω and n , and

$$E\left[\int_S^T (h - g_n)^2 dt\right] \rightarrow 0.$$

Proof: Suppose $|h(t, \omega)| \leq M$ for all (t, ω) . For each n let ψ_n be a non-negative, continuous function on \mathbf{R} such that

$$(i) \quad \psi_n(x) = 0 \text{ for } x \leq -\frac{1}{n} \text{ and } x \geq 0$$

and

$$(ii) \quad \int_{-\infty}^{\infty} \psi_n(x) dx = 1$$

Define

$$g_n(t, \omega) = \int_0^t \psi_n(s-t) h(s, \omega) ds.$$

Then $g_n(\cdot, \omega)$ is continuous for each ω and $|g_n(t, \omega)| \leq M$. Since $h \in \mathcal{V}$ we see that $g_n(t, \cdot)$ is \mathcal{F}_t -measurable for all t . (Use sums to approximate the integral defining g_n). Moreover,

$$\int_S^T (g_n(s, \omega) - h(s, \omega))^2 ds \rightarrow 0 \text{ as } n \rightarrow \infty, \text{ for each } \omega,$$

since $\{\psi_n\}_n$ constitute an approximate identity. (See e.g. Hoffman (1962, p.22).) So by bounded convergence

$$E\left[\int_S^T (h(t, \omega) - g_n(t, \omega))^2 dt\right] \rightarrow 0 \text{ as } n \rightarrow \infty,$$

as asserted.

STEP 3. Let $f \in \mathcal{V}$. Then there exists a sequence $\{h_n\} \subset \mathcal{V}$ such that h_n is bounded for each n and

$$E\left[\int_S^T (f - h_n)^2 dt\right] \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Proof. Put

$$h_n(t, \omega) = \begin{cases} -n & \text{if } f(t, \omega) < -n \\ f(t, \omega) & \text{if } -n \leq f(t, \omega) \leq n \\ n & \text{if } f(t, \omega) > n \end{cases}$$

Then the conclusion follows by dominated convergence.

That completes the approximation procedure.

We are now ready to complete the definition of the Ito integral

$$\int_S^T f(t, \omega) dB_t(\omega) \quad \text{for } f \in \mathcal{V} :$$

If $f \in \mathcal{V}$ we choose, by Steps 1-3, elementary functions $\phi_n \in \mathcal{V}$ such that

$$E\left[\int_S^T |f - \phi_n|^2 dt\right] \rightarrow 0.$$

Then define

$$\mathcal{I}[f](\omega) := \int_S^T f(t, \omega) dB_t(\omega) := \lim_{n \rightarrow \infty} \int_S^T \phi_n(t, \omega) dB_t(\omega).$$

The limit exists as an element of $L^2(P)$, since $\{\int_S^T \phi_n(t, \omega) dB_t(\omega)\}$ forms a Cauchy sequence in $L^2(P)$, by (3.11).

We summarize this as follows:

DEFINITION 3.6. (The Ito integral) Let $f \in \mathcal{V}(S, T)$. Then the *Ito integral* of f (from S to T) is defined by

$$\int_S^T f(t, \omega) dB_t(\omega) = \lim_{n \rightarrow \infty} \int_S^T \phi_n(t, \omega) dB_t(\omega) \quad (\text{limit in } L^2(P)) \quad (3.12)$$

where $\{\phi_n\}$ is a sequence of elementary functions such that

$$E\left[\int_S^T (f(t, \omega) - \phi_n(t, \omega))^2 dt\right] \rightarrow 0 \quad \text{as } n \rightarrow \infty. \quad (3.13)$$

Note that such a sequence $\{\phi_n\}$ satisfying (3.13) exists by Steps 1–3 above. Moreover, by (3.11) the limit in (3.12) exists and does not depend on the actual choice of $\{\phi_n\}$, as long as (3.13) holds. Furthermore, from (3.11) and (3.12) we get the following important

COROLLARY 3.7. (The Ito isometry)

$$E\left[\left(\int_S^T f(t, \omega) dB_t\right)^2\right] = E\left[\int_S^T f^2(t, \omega) dt\right] \quad \text{for all } f \in \mathcal{V}(S, T). \quad (3.14)$$

We illustrate this integral with an example:

EXAMPLE 3.8. Assume $B_0 = 0$. Then

$$\int_0^t B_s dB_s = \frac{1}{2} B_t^2 - \frac{1}{2} t.$$

Proof. Put $\phi_n(s, \omega) = \sum B_j(\omega) \cdot \mathcal{X}_{[t_j, t_{j+1})}(s)$, where $B_j = B_{t_j}$. Then

$$\begin{aligned} E\left[\int_0^t (\phi_n - B_s)^2 ds\right] &= E\left[\sum_j \int_{t_j}^{t_{j+1}} (B_j - B_s)^2 ds\right] \\ &= \sum_j \int_{t_j}^{t_{j+1}} (s - t_j) ds = \sum_j \frac{1}{2} (t_{j+1} - t_j)^2 \rightarrow 0 \text{ as } \Delta t_j \rightarrow 0. \end{aligned}$$

So

$$\int_0^t B_s dB_s = \lim_{\Delta t_j \rightarrow 0} \int_0^t \phi_n dB_s = \lim_{\Delta t_j \rightarrow 0} \sum_j B_j \Delta B_j.$$

(See also Exercise 3.13.) Now

$$\begin{aligned}\Delta(B_j^2) &= B_{j+1}^2 - B_j^2 = (B_{j+1} - B_j)^2 + 2B_j(B_{j+1} - B_j) \\ &= (\Delta B_j)^2 + 2B_j \Delta B_j,\end{aligned}$$

and therefore, since $B_0 = 0$,

$$B_t^2 = \sum_j \Delta(B_j^2) = \sum_j (\Delta B_j)^2 + 2 \sum_j B_j \Delta B_j$$

or

$$\sum_j B_j \Delta B_j = \frac{1}{2} B_t^2 - \frac{1}{2} \sum_j (\Delta B_j)^2.$$

Since $\sum_j (\Delta B_j)^2 \rightarrow t$ in $L^2(P)$ as $\Delta t_j \rightarrow 0$ (Exercise 2.15), the result follows.

The extra term $-\frac{1}{2}t$ shows that the Ito stochastic integral does not behave like ordinary integrals. In the next chapter we will establish the Ito formula, which explains the result in this example and which makes it easy to calculate many stochastic integrals.

Some properties of the Ito integral

First we observe the following:

THEOREM 3.9. Let $f, g \in \mathcal{V}(0, T)$ and let $0 \leq S < U < T$. Then

- (i) $\int_S^T f dB_t = \int_S^U f dB_t + \int_U^T f dB_t$ for a.a. ω
- (ii) $\int_S^T (cf + g) dB_t = c \cdot \int_S^T f dB_t + \int_S^T g dB_t$ (c constant) for a.a. ω
- (iii) $E[\int_S^T f dB_t] = 0$
- (iv) $\int_S^T f dB_t$ is \mathcal{F}_T -measurable.

Proof. This clearly holds for all elementary functions, so by taking limits we obtain this for all $f, g \in \mathcal{V}(0, T)$.

An important property of the Ito integral is that it is a *martingale*:

DEFINITION 3.10. A filtration (on (Ω, \mathcal{F})) is a family $\{\mathcal{M}_t\}_{t \geq 0}$ of σ -algebras $\mathcal{M}_t \subset \mathcal{F}$ such that

$$0 \leq s < t \Rightarrow \mathcal{M}_s \subset \mathcal{M}_t$$

(i.e. $\{\mathcal{M}_t\}$ is *increasing*). An n -dimensional stochastic process $\{M_t\}_{t \geq 0}$ on (Ω, \mathcal{F}, P) is called a *martingale* with respect to a filtration $\{\mathcal{M}_t\}_{t \geq 0}$ (and with respect to P^0) if

- (i) M_t is \mathcal{M}_t -measurable for all t ,
- (ii) $E[|M_t|] < \infty$ for all t

and

$$(iii) E[M_s | \mathcal{M}_t] = M_t \text{ for all } s \geq t.$$

Here the expectation in (ii) and the conditional expectation in (iii) is taken with respect to P^0 . (See Appendix B for a survey of conditional expectation).

EXAMPLE 3.11. Brownian motion B_t in \mathbf{R}^n is a martingale w.r.t. the σ -algebras \mathcal{F}_t generated by $\{B_s; s \leq t\}$, because

$$\begin{aligned} E[|B_t|^2] &\leq E[|B_t|^2] = |B_0|^2 + nt \quad \text{and if } s \geq t \text{ then} \\ E[B_s | \mathcal{F}_t] &= E[B_s - B_t + B_t | \mathcal{F}_t] \\ &= E[B_s - B_t | \mathcal{F}_t] + E[B_t | \mathcal{F}_t] = 0 + B_t = B_t. \end{aligned}$$

Here we have used that $E[B_s - B_t | \mathcal{F}_t] = E[B_s - B_t] = 0$ since $B_s - B_t$ is independent of \mathcal{F}_t (see (2.11) and Theorem B.2.d)) and we have used that $E[B_t | \mathcal{F}_t] = B_t$ since B_t is \mathcal{F}_t -measurable (see Theorem B.2.c)).

For continuous martingales we have the following important inequality due to Doob: (See e.g. Stroock and Varadhan (1979), Th. 1.2.3 or Revuz and Yor (1991), Th. II.1.7)

THEOREM 3.12. (Doob's martingale inequality) If M_t is a martingale such that $t \rightarrow M_t(\omega)$ is continuous a.s., then for all $p \geq 1, T \geq 0$ and all $\lambda > 0$

$$P\left[\sup_{0 \leq t \leq T} |M_t| \geq \lambda\right] \leq \frac{1}{\lambda^p} \cdot E[|M_T|^p].$$

We now use this inequality to prove that the Ito integral

$$\int_0^t f(s, \omega) dB_s$$

can be chosen to depend continuously on t :

THEOREM 3.13. Let $f \in \mathcal{V}(0, T)$. Then there exists a t -continuous version of

$$\int_0^t f(s, \omega) dB_s(\omega); 0 \leq t \leq T,$$

i.e. there exists a t -continuous stochastic process J_t on (Ω, \mathcal{F}, P) such that

$$P[J_t = \int_0^t f dB] = 1 \text{ for all } t, 0 \leq t \leq T. \quad (3.15)$$

Proof. Let $\phi_n = \phi_n(t, \omega) = \sum_j e_j^{(n)}(\omega) \chi_{[t_j^{(n)}, t_{j+1}^{(n)})}(t)$ be elementary functions such that

$$E\left[\int_0^T (f - \phi_n)^2 dt\right] \rightarrow 0 \text{ when } n \rightarrow \infty.$$

Put

$$\begin{aligned} I_n(t, \omega) &= \int_0^t \phi_n(s, \omega) dB_s(\omega) \text{ and} \\ I_t &= I(t, \omega) = \int_0^t f(s, \omega) dB_s(\omega); 0 \leq t \leq T. \end{aligned}$$

Then $I_n(\cdot, \omega)$ is continuous, for all n . Moreover, $I_n(t, \omega)$ is a martingale with respect to \mathcal{F}_t , for all n :

$$\begin{aligned} E[I_n(s, \omega) | \mathcal{F}_t] &= E\left[\left(\int_0^t \phi_n dB + \int_t^s \phi_n dB\right) | \mathcal{F}_t\right] \\ &= \int_0^t \phi_n dB + E\left[\sum_{t \leq t_j^{(n)} \leq t_{j+1}^{(n)} \leq s} e_j^{(n)} dB_j | \mathcal{F}_t\right] \\ &= \int_0^t \phi_n dB + \sum_j E[E[e_j^{(n)} \Delta B_j | \mathcal{F}_{t_j}] | \mathcal{F}_t] \\ &= \int_0^t \phi_n dB + \sum_j E[e_j^{(n)} E[\Delta B_j | \mathcal{F}_{t_j}] | \mathcal{F}_t] \\ &= \int_0^t \phi_n dB = I_n(t, \omega) \end{aligned} \quad (3.16)$$

when $t < s$, using Theorem B.3. and Theorem B.2.d).

Hence $I_n - I_m$ is also an \mathcal{F}_t -martingale, so by the martingale inequality (Theorem 3.10) it follows that

$$\begin{aligned} P\left[\sup_{0 \leq t \leq T} |I_n(t, \omega) - I_m(t, \omega)| > \epsilon\right] &\leq \frac{1}{\epsilon^2} \cdot E[|I_n(T, \omega) - I_m(T, \omega)|^2] \\ &= \frac{1}{\epsilon^2} E\left[\int_0^T (\phi_n - \phi_m)^2 ds\right] \rightarrow 0 \text{ as } m, n \rightarrow \infty. \end{aligned}$$

Hence we may choose a subsequence $n_k \uparrow \infty$ s.t.

$$P\left[\sup_{0 \leq t \leq T} |I_{n_{k+1}}(t, \omega) - I_{n_k}(t, \omega)| > 2^{-k}\right] < 2^{-k}.$$

By the Borel-Cantelli lemma

$$P\left[\sup_{0 \leq t \leq T} |I_{n_{k+1}}(t, \omega) - I_{n_k}(t, \omega)| > 2^{-k} \text{ for infinitely many } k\right] = 0.$$

So for a.a. ω there exists $k_1(\omega)$ such that

$$\sup_{0 \leq t \leq T} |I_{n_{k+1}}(t, \omega) - I_{n_k}(t, \omega)| < 2^{-k} \text{ for } k \geq k_1(\omega).$$

Therefore $I_{n_k}(t, \omega)$ is uniformly convergent for $t \in [0, T]$, for a.a. ω and so the limit, denoted by $J_t(\omega)$, is t-continuous for $t \in [0, T]$, a.s. Since $I_{n_k}(t, \cdot) \rightarrow I(t, \cdot)$ in $L^2[P]$ for all t , we must have

$$I_t = J_t \text{ a.s., for all } t \in [0, T].$$

That completes the proof.

From now on we shall always assume that $\int_0^t f(s, \omega) dB_s(\omega)$ means a t-continuous version of the integral.

COROLLARY 3.14. Let $f(t, \omega) \in \mathcal{V}(0, T)$ for all T . Then

$$M_t(\omega) = \int_0^t f(s, \omega) dB_s$$

is a martingale w.r.t. \mathcal{F}_t and

$$P\left[\sup_{0 \leq t \leq T} |M_t| \geq \lambda\right] \leq \frac{1}{\lambda^2} \cdot E\left[\int_0^T f(s, \omega)^2 ds\right]; \lambda, T > 0. \quad (3.17)$$

Proof. This follows from (3.16), the a.s. t-continuity of M_t and the martingale inequality (Theorem 3.12), combined with the Ito isometry (3.14).

Extensions of the Ito integral

The Ito integral $\int f dB$ can be defined for a larger class of integrands f than \mathcal{V} . First, the measurability condition (ii) of Definition 3.4 can be relaxed to the following:

- (ii)'. There exists an increasing family of σ -algebras $\mathcal{H}_t; t \geq 0$ such that
 - a) B_t is a martingale with respect to \mathcal{H}_t and
 - b) f_t is \mathcal{H}_t -adapted

Note that a) implies that $\mathcal{F}_t \subset \mathcal{H}_t$. The essence of this extension is that we can allow f_t to depend on more than \mathcal{F}_t as long as B_t remains a martingale with respect to the “history” of $f_s; s \leq t$. If (ii)' holds, then $E[B_s - B_t | \mathcal{H}_t] = 0$ for all $s > t$ and if we inspect our proofs above, we see that this is sufficient to carry out the construction of the Ito integral as before.

The most important example of a situation where (ii)' applies (and (ii) doesn't) is the following:

Suppose $B_t(\omega) = B_k(t, \omega)$ is the k 'th coordinate of n -dimensional Brownian motion (B_1, \dots, B_n) . Let $\mathcal{F}_t^{(n)}$ be the σ -algebra generated by $B_1(s_1, \cdot), \dots, B_n(s_n, \cdot); s_k \leq t$. Then $B_k(t, \omega)$ is a martingale with respect to $\mathcal{F}_t^{(n)}$ because $B_k(s, \cdot) - B_k(t, \cdot)$ is independent of $\mathcal{F}_t^{(n)}$ when $s > t$. Thus we have now defined $\int_0^t f(s, \omega) dB_k(s, \omega)$ for $\mathcal{F}_t^{(n)}$ -adapted integrands f_t . That includes integrals like

$$\int B_2 dB_1 \quad \text{or} \quad \int \sin(B_1^2 + B_2^2) dB_2$$

involving several components of n -dimensional Brownian motion. (Here we have used the notation $dB_1 = dB_1(t, \omega)$ etc.)

This allows us to define the *multi-dimensional Ito integral* as follows:

Let $B = (B_1, B_2, \dots, B_n)$ be n -dimensional Brownian motion and let $v = [v_{ij}(t, \omega)]$ be an $m \times n$ matrix where each entry $v_{ij}(t, \omega)$ satisfies (i) and (iii) of Definition 3.4 and is $\mathcal{F}_t^{(n)}$ -adapted (or more generally satisfies (ii)' with respect to all B_k). Then, using matrix notation, we define

$$\int_0^t v dB = \int_0^t \begin{pmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & & \\ v_{m1} & \cdots & v_{mn} \end{pmatrix} \begin{pmatrix} dB_1 \\ \vdots \\ dB_n \end{pmatrix}$$

to be the $m \times 1$ matrix (column vector) whose i 'th component is the following sum of 1-dimensional Ito integrals:

$$\sum_{j=1}^n \int_0^t v_{ij}(s, \omega) dB_j(s, \omega)$$

The next extension of the Ito integral consists of weakening condition (iii) of Definition 3.4 to

$$(iii)' \quad P\left[\int_0^t f(s, \omega)^2 ds < \infty \text{ for all } t \geq 0\right] = 1$$

DEFINITION 3.15. \mathcal{W} denotes the class of processes $f(t, \omega)$ satisfying (i) of Definition 3.4 and (ii)', (iii)' above.

Let B_t denote 1-dimensional Brownian motion. If $f \in \mathcal{W}$ one can show that for all t there exist step functions $f_n \in \mathcal{W}[0, t]$ such that $\int |f_n - f|^2 ds \rightarrow 0$ in probability, i.e. in measure with respect to P . For such a sequence one has that $\int_0^t f_n(s, \omega) dB_s$ converges in probability to some random variable and the limit only depends on f , not on the sequence $\{f_n\}$. Thus we may define

$$\int_0^t f(s, \omega) dB_s(\omega) = \lim_{n \rightarrow \infty} \int_0^t f_n(s, \omega) dB_s(\omega) \quad (\text{limit in probability}) \text{ for } f \in \mathcal{W}. \quad (3.18)$$

As before there exists a t-continuous version of this integral. See Friedman (1975, Ch. 4) or McKean (1969, Ch. 2) for details.

A comparison of Ito and Stratonovich integrals

Let us now return to our original question in this chapter: We have argued that the mathematical interpretation of the white noise equation

$$\frac{dX}{dt} = b(t, X_t) + \sigma(t, X_t) \cdot W_t \quad (3.2)$$

is that X_t is a solution of the integral equation

$$X_t = X_0 + \int_0^t b(s, X_s) ds + \left\langle \int_0^t \sigma(s, X_s) dB_s \right\rangle, \quad (3.5)$$

for some suitable interpretation of the last integral in (3.5). However, as indicated earlier, the Ito interpretation of an integral of the form

$$\left\langle \int_0^t f(s, \omega) dB_s(\omega) \right\rangle \quad (*)$$

is just one of several reasonable choices. For example, the Stratonovich integral is another possibility, leading (in general) to a different result. So the question still remains: *Which interpretation of (*) makes (3.5) the “right” mathematical*

model for the equation (3.2)? Here is an argument that indicates that the Stratonovich interpretation in some situations may be the most appropriate: Choose t-continuously differentiable processes $B_t^{(n)}$ such that for a.a. ω

$$B^{(n)}(t, \omega) \rightarrow B(t, \omega) \text{ as } n \rightarrow \infty$$

uniformly (in t) in bounded intervals. For each ω let $X_t^{(n)}(\omega)$ be the solution of the corresponding (deterministic) differential equation

$$\frac{dX_t}{dt} = b(t, X_t) + \sigma(t, X_t) \frac{dB_t^{(n)}}{dt} \quad (3.19)$$

Then $X_t^{(n)}(\omega)$ converges to some function $X_t(\omega)$ in the same sense: For a.a. ω we have that $X_t^{(n)}(\omega) \rightarrow X_t(\omega)$ as $n \rightarrow \infty$, uniformly (in t) in bounded intervals.

It turns out (see Wong and Zakai (1969) and Sussman (1978)) that this solution X_t coincides with the solution of (3.5) obtained by using Stratonovich integrals, i.e.

$$X_t = X_0 + \int_0^t b(s, X_s) ds + \int_0^t \sigma(s, X_s) \circ dB_s \quad (3.20)$$

This implies that X_t is the solution of the following modified Ito equation:

$$X_t = X_0 + \int_0^t b(s, X_s) ds + \frac{1}{2} \int_0^t \sigma'(s, X_s) \sigma(s, X_s) ds + \int_0^t \sigma(s, X_s) dB_s, \quad (3.21)$$

where σ' denotes the derivative of $\sigma(t, x)$ with respect to x . (See Stratonovich (1966)).

Therefore, from this point of view it seems reasonable to use (3.21) (i.e. the Stratonovich interpretation) - and not the Ito interpretation

$$X_t = X_0 + \int_0^t b(s, X_s) ds + \int_0^t \sigma(s, X_s) dB_s \quad (3.22)$$

as the model for the original white noise equation (3.2).

On the other hand, the specific feature of the Ito model of “not looking into the future” (as explained after Example 3.1) seems to be a reason for choosing the Ito interpretation in many cases, for example in biology (see the discussion in Turelli (1977)). The difference between the two interpretations is illustrated in Example 5.1. Note that (3.21) and (3.22) coincide if $\sigma(t, x)$ does not depend on x . For example, this is the situation in the linear case handled in the filtering problem in Chapter VI.

In any case, because of the explicit connection (3.21) between the two models (and a similar connection in higher dimensions - see (6.2)), it will for many purposes suffice to do the general mathematical treatment for one of the two types of integrals. In general one can say that the Stratonovich integral has the advantage of leading to ordinary chain rule formulas under a transformation (change of variable), i.e. there are no second order terms in the Stratonovich analogue of the Ito transformation formula (see Theorems 4.2 and 4.6). This property makes the Stratonovich integral natural to use for example in connection with stochastic differential equations on manifolds (see Elworthy (1982) or Ikeda and Watanabe (1989)).

However, Stratonovich integrals are not martingales, as we have seen that Ito integrals are. This gives the Ito integral an important computational advantage, even though it does not behave so nicely under transformations (as Example 3.8 shows). For our purposes the Ito integral will be most convenient, so we will base our discussion on that from now on.

EXERCISES

Unless otherwise stated B_t denotes Brownian motion in \mathbf{R} , $B_0 = 0$.

3.1.

Prove directly from the definition of Ito integrals (Definition 3.6) that

$$\int_0^t s dB_s = t B_t - \int_0^t B_s ds$$

(Hint: Note that

$$\sum_j \Delta(s_j B_j) = \sum_j s_j \Delta B_j + \sum_j B_{j+1} \Delta s_j$$

3.2.

Prove directly from the definition of Ito integrals that

$$\int_0^t B_s^2 dB_s = \frac{1}{3} B_t^3 - \int_0^t B_s ds$$

3.3.

If $X_t : \Omega \rightarrow \mathbf{R}^n$ is a stochastic process, let $\mathcal{H}_t = \mathcal{H}_t^{(X)}$ denote the σ -algebra generated by $\{X_s(\cdot); s \leq t\}$ (i.e. $\{\mathcal{H}_t^{(X)}\}_{t \geq 0}$ is the filtration of the process $\{X_t\}_{t \geq 0}$).

a) Show that if X_t is a martingale w.r.t. some filtration $\{\mathcal{N}_t\}_{t \geq 0}$, then X_t is also a martingale w.r.t. its own filtration $\{\mathcal{H}_t^{(X)}\}_{t \geq 0}$.

b) Show that if X_t is a martingale w.r.t $\mathcal{H}_t^{(X)}$, then

$$E[X_t] = E[X_0] \quad \text{for all } t \geq 0 \quad (*)$$

c) Give an example of a stochastic process X_t satisfying $(*)$ and which is *not* a martingale w.r.t. its own filtration.

3.4.

Check whether the following processes X_t are martingales w.r.t. $\{\mathcal{F}_t\}$:

(i) $X_t = B_t + 4t$

(ii) $X_t = B_t^2$

(iii) $X_t = t^2 B_t - 2 \int_0^t s B_s ds$

(iv) $X_t = B_1(t)B_2(t)$, where $(B_1(t), B_2(t))$ is 2-dimensional Brownian motion.

3.5.

Prove directly (without using Example 3.8) that

$$M_t = B_t^2 - t$$

is an \mathcal{F}_t -martingale.

3.6.

Prove that $N_t = B_t^3 - 3tB_t$ is a martingale

3.7.

A famous result of Ito (1951) gives the following formula for n times iterated Ito integrals:

$$n! \int_{0 \leq u_1 \leq \dots \leq u_n \leq t} \left(\int \left(\int dB_{u_1} \right) dB_{u_2} \right) \dots dB_{u_n} = t^{\frac{n}{2}} h_n \left(\frac{B_t}{\sqrt{t}} \right) \quad (3.23)$$

where h_n is the Hermite polynomial of degree n , defined by

$$h_n(x) = (-1)^n e^{\frac{x^2}{2}} \frac{d^n}{dx^n} (e^{-\frac{x^2}{2}}); n = 0, 1, 2, \dots$$

(Thus $h_0(x) = 1, h_1(x) = x, h_2(x) = x^2 - 1, h_3(x) = x^3 - 3x$)

a) Verify that in each of these n Ito integrals the integrand satisfies the requirements in Definition 3.4.

- b) Verify formula (3.23) for $n = 1, 2, 3$ by combining Example 3.8 and Exercise 3.2.
- c) Use b) to give a new proof of the statement in Exercise 3.6.

3.8.

- a) Let Y be a real valued random variable on (Ω, \mathcal{F}, P) such that

$$E[|Y|] < \infty$$

Define

$$M_t = E[Y | \mathcal{F}_t]; \quad t \geq 0.$$

Show that M_t is an \mathcal{F}_t -martingale.

- b) Conversely, let $M_t; t \geq 0$ be a real valued \mathcal{F}_t -martingale such that

$$\sum_{t \geq 0} E[|M_t|^p] < \infty \quad \text{for some } p > 1.$$

Show that there exists $Y \in L^1(P)$ such that

$$M_t = E[Y | \mathcal{F}_t].$$

(Hint: Use Corollary C.7.)

3.9.

Suppose $f \in \mathcal{V}(0, T)$ and that $t \rightarrow f(t, \omega)$ is continuous for a.a. ω . Then we have shown that

$$\int_0^T f(t, \omega) dB_t(\omega) = \lim_{\Delta t_j \rightarrow 0} \sum_j f(t_j, \omega) \Delta B_j \quad \text{in } L^2(P^0)$$

Similarly we define the *Stratonovich integral* of f by

$$\int_0^T f(t, \omega) \circ dB_t(\omega) = \lim_{\Delta t_j \rightarrow 0} \sum_j f(t_j^*, \omega) \Delta B_j, \quad \text{where } t_j^* = \frac{1}{2}(t_j + t_{j+1}),$$

whenever the limit exists in $L^2(P^0)$. In general these integrals are different. For example, compute

$$\int_0^T B_t \circ dB_t$$

and compare with Example 3.8.

3.10.

If the function f in Exercise 3.9 varies “smoothly” with t then in fact the Ito and Stratonovich integrals of f coincide. More precisely, assume that there exists $K < \infty$ and $\epsilon > 0$ such that

$$E[|f(s, \cdot) - f(t, \cdot)|^2] \leq K|s - t|^{1+\epsilon}; \quad 0 \leq s, t \leq T$$

Prove that then we have

$$\int_0^T f(t, \omega) dB_t = \lim_{\Delta t_j \rightarrow 0} \sum_j f(t'_j, \omega) \Delta B_j \quad (\text{limit in } L^1(P^0))$$

for any choice of $t'_j \in [t_j, t_{j+1}]$. In particular,

$$\int_0^T f(t, \omega) dB_t = \int_0^T f(t, \omega) \circ dB_t$$

(Hint: Consider $E[|\sum_j f(t_j, \omega) \Delta B_j - \sum_j f(t'_j, \omega) \Delta B_j|]$).

3.11.

Let W_t be a stochastic process satisfying (i), (ii) and (iii) (below (3.2)). Prove that W_t cannot have continuous paths. (Hint: Consider $E[(W_t^{(N)} - W_s^{(N)})^2]$, where

$$W_t^{(N)} = (-N) \vee (N \wedge W_t), \quad N = 1, 2, 3, \dots$$

3.12.

As in Exercise 3.9 we let $\circ dB_t$ denote Stratonovich differentials.

- (i) Use (3.21) to transform the following Stratonovich differential equations into Ito differential equations:

- (a) $dX_t = \gamma X_t dt + \alpha X_t \circ dB_t$
- (b) $dX_t = \sin X_t \cos X_t dt + (t^2 + \cos X_t) \circ dB_t$

- (ii) Transform the following Ito differential equations into Stratonovich differential equations:

- (a) $dX_t = r X_t dt + \alpha X_t dB_t$
- (b) $dX_t = 2e^{-X_t} dt + X_t^2 dB_t$

3.13.

A stochastic process $X_t(\cdot) : \Omega \rightarrow \mathbf{R}$ is *continuous in mean square* if $E[X_t^2] < \infty$ for all t and

$$\lim_{s \rightarrow t} E[(X_s - X_t)^2] = 0 \quad \text{for all } t \geq 0$$

- a) Prove that Brownian motion B_t is continuous in mean square.
- b) Let $f : \mathbf{R} \rightarrow \mathbf{R}$ be a Lipschitz continuous function, i.e. there exists $C < \infty$ such that

$$|f(x) - f(y)| \leq C|x - y| \quad \text{for all } x, y \in \mathbf{R}$$

Prove that

$$Y_t := f(B_t)$$

is continuous in mean square.

- c) Let X_t be a stochastic process which is continuous in mean square and assume that $X_t \in \mathcal{V}[S, T]$. Show that

$$\int_S^T X_t dB_t = \lim_{n \rightarrow \infty} \int_S^T \phi_n(t, \omega) dB_t(\omega) \quad (\text{limit in } L^2(P^0))$$

where

$$\phi_n(t, \omega) = \sum_j X_{t_j^{(n)}}(\omega) \mathcal{X}_{[t_j^{(n)}, t_{j+1}^{(n)})}(t), \quad T < \infty.$$

(Hint: Consider

$$E\left[\int_S^T (X_t - \phi_n(t))^2 dt\right] = E\left[\sum_j \int_{t_j^{(n)}}^{t_{j+1}^{(n)}} (X_t - X_{t_j^{(n)}})^2 dt\right].$$

3.14.

Show that a function $h(\omega)$ is \mathcal{F}_t -measurable if and only if h is a pointwise limit (for a.a. ω) of sums of functions of the form

$$g_1(B_{t_1}) \cdot g_2(B_{t_2}) \cdots g_k(B_{t_k})$$

where g_1, \dots, g_k are bounded continuous functions and $t_j \leq t$ for $j \leq k$, $k = 1, 2, \dots$

Hint: Complete the following steps:

- a) We may assume that h is bounded.
- b) For $n = 1, 2, \dots$ and $j = 1, 2, \dots$ put $t_j = t_j^{(n)} = j \cdot 2^{-n}$. For fixed n let \mathcal{H}_n be the σ -algebra generated by $\{B_{t_j}(\cdot)\}_{t_j \leq t}$. Then by Corollary C.9

$$h = E[h|\mathcal{F}_t] = \lim_{n \rightarrow \infty} E[h|\mathcal{H}_n] \quad (\text{pointwise a.e. limit})$$

- c) Define $h_n := E[h|\mathcal{H}_n]$. Then by the Doob-Dynkin lemma (Lemma 2.1) we have

$$h_n(\omega) = G_n(B_{t_1}(\omega), \dots, B_{t_k}(\omega))$$

for some Borel function $G_n : \mathbf{R}^k \rightarrow \mathbf{R}$, where $k = \max\{j; j \cdot 2^{-n} \leq t\}$.

Now use that any Borel function $G : \mathbf{R}^k \rightarrow \mathbf{R}$ can be approximated pointwise a.e. by a continuous function $F : \mathbf{R}^k \rightarrow \mathbf{R}$ and complete the proof by applying the Stone-Weierstrass theorem.

IV. Ito Processes and the Ito Formula

The 1-dimensional Ito formula

Example 3.8 illustrates that the basic definition of Ito integrals is not very useful when we try to evaluate a given integral. This is similar to the situation for ordinary Riemann integrals, where we do not use the basic definition but rather the fundamental theorem of calculus plus the chain rule in the explicit calculations.

In this context, however, we have no differentiation theory, only integration theory. Nevertheless it turns out that it is possible to establish an Ito integral version of the chain rule, called the Ito formula. The Ito formula is, as we will show by examples, very useful for evaluating Ito integrals.

From the example

$$\int_0^t B_s dB_s = \frac{1}{2} B_t^2 - \frac{1}{2} t \quad \text{or} \quad \frac{1}{2} B_t^2 = \frac{1}{2} t + \int_0^t B_s dB_s, \quad (4.1)$$

we see that the image of the Ito integral $B_t = \int_0^t dB_s$ by the map $g(x) = \frac{1}{2}x^2$ is not again an Ito integral of the form

$$\int_0^t f(s, \omega) dB_s(\omega)$$

but a combination of a dB_s -and a ds-integral:

$$\frac{1}{2} B_t^2 = \int_0^t \frac{1}{2} ds + \int_0^t B_s dB_s. \quad (4.2)$$

It turns out that if we introduce *Ito processes* (also called *stochastic integrals*) as sums of a dB_s -and a ds-integral then this family of integrals is stable under smooth maps. Thus we define

DEFINITION 4.1. Let B_t be 1-dimensional Brownian motion on (Ω, \mathcal{F}, P) . A (1-dimensional) *Ito process* (or *stochastic integral*) is a stochastic process X_t on (Ω, \mathcal{F}, P) of the form

$$X_t = X_0 + \int_0^t u(s, \omega) ds + \int_0^t v(s, \omega) dB_s, \quad (4.3)$$

where $v \in \mathcal{W}$, so that

$$P\left[\int_0^t v(s, \omega)^2 ds < \infty \text{ for all } t \geq 0\right] = 1 \quad (4.4)$$

(see Definition 3.15). We also assume that u is \mathcal{H}_t -adapted (where \mathcal{H}_t is as in (ii)', Definition 3.15) and

$$P\left[\int_0^t |u(s, \omega)| ds < \infty \text{ for all } t \geq 0\right] = 1. \quad (4.5)$$

If X_t is an Ito process of the form (4.3) the equation (4.3) is sometimes written in the shorter differential form

$$dX_t = u dt + v dB_t. \quad (4.6)$$

For example, (4.1) (or (4.2)) may be represented by

$$d\left(\frac{1}{2}B_t^2\right) = \frac{1}{2}dt + B_t dB_t.$$

We are now ready to state the first main result in this chapter:

THEOREM 4.2. (THE 1-DIMENSIONAL ITO FORMULA). Let X_t be an Ito process given by

$$dX_t = u dt + v dB_t.$$

Let $g(t, x) \in C^2([0, \infty) \times \mathbf{R})$ (i.e. g is twice continuously differentiable on $[0, \infty) \times \mathbf{R}$). Then

$$Y_t = g(t, X_t)$$

is again an Ito process, and

$$dY_t = \frac{\partial g}{\partial t}(t, X_t) dt + \frac{\partial g}{\partial x}(t, X_t) dX_t + \frac{1}{2} \frac{\partial^2 g}{\partial x^2}(t, X_t) \cdot (dX_t)^2, \quad (4.7)$$

where $(dX_t)^2 = (dX_t) \cdot (dX_t)$ is computed according to the rules

$$dt \cdot dt = dt \cdot dB_t = dB_t \cdot dt = 0, dB_t \cdot dB_t = dt. \quad (4.8)$$

Before we prove Ito's formula let us look at some examples.

EXAMPLE 4.3. Let us return to the integral

$$I = \int_0^t B_s dB_s \quad \text{from Chapter III.}$$

Choose $X_t = B_t$ and $g(t, x) = \frac{1}{2}x^2$. Then

$$Y_t = g(t, B_t) = \frac{1}{2}B_t^2.$$

Then by Ito's formula,

$$dY_t = \frac{\partial g}{\partial t} dt + \frac{\partial g}{\partial x} dB_t + \frac{1}{2} \frac{\partial^2 g}{\partial x^2} (dB_t)^2 = B_t dB_t + \frac{1}{2}(dB_t)^2 = B_t dB_t + \frac{1}{2}dt.$$

Hence

$$d\left(\frac{1}{2}B_t^2\right) = B_t dB_t + \frac{1}{2}dt.$$

In other words,

$$\frac{1}{2}B_t^2 = \int_0^t B_s dB_s + \frac{1}{2}t, \quad \text{as in Chapter III.}$$

EXAMPLE 4.4. What is

$$\int_0^t s dB_s?$$

From classical calculus it seems reasonable that a term of the form tB_t should appear, so we put

$$g(t, x) = tx$$

and

$$Y_t = g(t, B_t) = tB_t.$$

Then by Ito's formula,

$$dY_t = B_t dt + t dB_t + 0$$

i.e.

$$d(tB_t) = B_t dt + t dB_t$$

or

$$tB_t = \int_0^t B_s ds + \int_0^t s dB_s$$

or

$$\int_0^t s dB_s = t B_t - \int_0^t B_s ds,$$

which is reasonable from an integration-by-parts point of view.

More generally, the same method gives

THEOREM 4.5. (INTEGRATION BY PARTS). Suppose $f(s, \omega) = f(s)$ only depends on s and that f is continuous and of bounded variation in $[0, t]$. Then

$$\int_0^t f(s) dB_s = f(t) B_t - \int_0^t B_s df_s$$

Note that it is crucial for the result to hold that f does not depend on ω . (See Exercise 4.3.)

Proof of the Ito formula. First observe that if we substitute

$$dX_t = u dt + v dB_t$$

and use (4.8) we get the equivalent expression

$$\begin{aligned} g(t, X_t) &= g(0, X_0) + \int_0^t \left(\frac{\partial g}{\partial s}(s, X_s) + u_s \frac{\partial g}{\partial x}(s, X_s) + \frac{1}{2} v_s^2 \cdot \frac{\partial^2 g}{\partial x^2}(s, X_s) \right) ds \\ &\quad + \int_0^t v_s \cdot \frac{\partial g}{\partial x}(s, X_s) dB_s \quad \text{where } u_s = u(s, \omega), v_s = v(s, \omega). \end{aligned} \quad (4.9)$$

Note that (4.9) is an Ito process in the sense of Definition 4.1.

We may assume that g , $\frac{\partial g}{\partial t}$, $\frac{\partial g}{\partial x}$ and $\frac{\partial^2 g}{\partial x^2}$ are bounded, for if (4.9) is proved in this case we obtain the general case by approximating by C^2 functions g_n such that g_n , $\frac{\partial g_n}{\partial t}$, $\frac{\partial g_n}{\partial x}$ and $\frac{\partial^2 g_n}{\partial x^2}$ are bounded for each n and converge uniformly on compact subsets of $[0, \infty) \times \mathbf{R}$ to g , $\frac{\partial g}{\partial t}$, $\frac{\partial g}{\partial x}$, $\frac{\partial^2 g}{\partial x^2}$, respectively. (See Exercise 4.9.) Moreover, from (3.12)–(3.13) we see that we may assume that $u(t, \omega)$ and $v(t, \omega)$ are elementary functions. Using Taylor's theorem we get

$$\begin{aligned} g(t, X_t) &= g(0, X_0) + \sum_j \Delta g(t_j, X_j) = g(0, X_0) + \sum_j \frac{\partial g}{\partial t} \Delta t_j + \sum_j \frac{\partial g}{\partial x} \Delta X_j \\ &\quad + \frac{1}{2} \sum_j \frac{\partial^2 g}{\partial t^2} (\Delta t_j)^2 + \sum_j \frac{\partial^2 g}{\partial t \partial x} (\Delta t_j)(\Delta X_j) + \frac{1}{2} \sum_j \frac{\partial^2 g}{\partial x^2} (\Delta X_j)^2 + \sum_j R_j, \end{aligned}$$

where $\frac{\partial g}{\partial t}, \frac{\partial g}{\partial x}$ etc. are evaluated at the points (t_j, X_{t_j}) ,

$$\Delta t_j = t_{j+1} - t_j, \Delta X_j = X_{t_{j+1}} - X_{t_j}, \Delta g(t_j, X_j) = g(t_{j+1}, X_{t_{j+1}}) - g(t_j, X_j)$$

and $R_j = o(|\Delta t_j|^2 + |\Delta X_j|^2)$ for all j .

If $\Delta t_j \rightarrow 0$ then

$$\sum_j \frac{\partial g}{\partial t} \Delta t_j = \sum_j \frac{\partial g}{\partial t}(t_j, X_j) \Delta t_j \rightarrow \int_0^t \frac{\partial g}{\partial t}(s, X_s) ds \quad (4.10)$$

$$\sum_j \frac{\partial g}{\partial x} \Delta X_j = \sum_j \frac{\partial g}{\partial x}(t_j, X_j) \Delta X_j \rightarrow \int_0^t \frac{\partial g}{\partial x}(s, X_s) dX_s. \quad (4.11)$$

Moreover, since u and v are elementary we get

$$\begin{aligned} \sum_j \frac{\partial^2 g}{\partial x^2} (\Delta X_j)^2 &= \sum_j \frac{\partial^2 g}{\partial x^2} u_j^2 (\Delta t_j)^2 + 2 \sum_j \frac{\partial^2 g}{\partial x^2} u_j v_j (\Delta t_j) (\Delta B_j) \\ &\quad + \sum_j \frac{\partial^2 g}{\partial x^2} v_j^2 \cdot (\Delta B_j)^2, \quad \text{where } u_j = u(t_j, \omega), v_j = v(t_j, \omega). \end{aligned} \quad (4.12)$$

The first two terms here tend to 0 as $\Delta t_j \rightarrow 0$. For example,

$$E[(\sum_j \frac{\partial^2 g}{\partial x^2} u_j v_j (\Delta t_j) (\Delta B_j))^2] = \sum_j E[(\frac{\partial^2 g}{\partial x^2} u_j v_j)^2] (\Delta t_j)^3 \rightarrow 0 \text{ as } \Delta t_j \rightarrow 0.$$

We claim that the last term tends to

$$\int_0^t \frac{\partial^2 g}{\partial x^2} v^2 ds \quad \text{in } L^2(P^0), \text{ as } \Delta t_j \rightarrow 0.$$

To prove this, put $a(t) = \frac{\partial^2 g}{\partial x^2}(t, X_t) v^2(t, \omega)$, $a_j = a(t_j)$ and consider

$$E[(\sum_j a_j (\Delta B_j)^2 - \sum_j a_j \Delta t_j)^2] = \sum_{i,j} E[a_i a_j ((\Delta B_i)^2 - \Delta t_i)((\Delta B_j)^2 - \Delta t_j)]$$

If $i < j$ then $a_i a_j ((\Delta B_i)^2 - \Delta t_i)$ and $(\Delta B_j)^2 - \Delta t_j$ are independent so the terms vanish in this case, and similarly if $i > j$. So we are left with

$$\begin{aligned} \sum_j E[a_j^2 ((\Delta B_j)^2 - \Delta t_j)^2] &= \sum_j E[a_j^2] \cdot E[(\Delta B_j)^4 - 2(\Delta B_j)^2 \Delta t_j + (\Delta t_j)^2] \\ &= \sum_j E[a_j^2] \cdot (3(\Delta t_j)^2 - 2(\Delta t_j)^2 + (\Delta t_j)^2) = 2 \sum_j E[a_j^2] \cdot (\Delta t_j)^2 \\ &\rightarrow 0 \text{ as } \Delta t_j \rightarrow 0. \end{aligned}$$

In other words, we have established that

$$\sum_j a_j (\Delta B_j)^2 \rightarrow \int_0^t a(s) ds \quad \text{in } L^2(P^0) \text{ as } \Delta t_j \rightarrow 0$$

and this is often expressed shortly by the striking formula

$$(dB_t)^2 = dt. \quad (4.13)$$

The argument above also proves that $\sum R_j \rightarrow 0$ as $\Delta t_j \rightarrow 0$. That completes the proof of the Ito formula.

REMARK. Note that it is enough that $g(t, x)$ is C^2 on $[0, \infty) \times U$, if $U \subset \mathbf{R}$ is an open set such that $X_t(\omega) \in U$ for all $t \geq 0, \omega \in \Omega$. Moreover, it is sufficient that $g(t, x)$ is C^1 w.r.t. t and C^2 w.r.t. x .

The multi-dimensional Ito formula

We now turn to the situation in higher dimensions: Let $B(t, \omega) = (B_1(t, \omega), \dots, B_m(t, \omega))$ denote m -dimensional Brownian motion. If each of the processes $u_i(t, \omega)$ and $v_{ij}(t, \omega)$ satisfies the conditions given in (4.3)–(4.5) above ($1 \leq i \leq n$, $1 \leq j \leq m$) then we can form the following n Ito processes

$$\begin{cases} dX_1 = u_1 dt + v_{11} dB_1 + \dots + v_{1m} dB_m \\ \vdots \quad \vdots \quad \vdots \\ dX_n = u_n dt + v_{n1} dB_1 + \dots + v_{nm} dB_m \end{cases} \quad (4.14)$$

Or, in matrix notation simply

$$dX(t) = u dt + v dB(t), \quad (4.15)$$

where

$$X(t) = \begin{pmatrix} X_1(t) \\ \vdots \\ X_n(t) \end{pmatrix}, \quad u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}, \quad v = \begin{pmatrix} v_{11} & \dots & v_{1m} \\ \vdots & & \vdots \\ v_{n1} & \dots & v_{nm} \end{pmatrix}, \quad dB(t) = \begin{pmatrix} dB_1(t) \\ \vdots \\ dB_m(t) \end{pmatrix} \quad (4.16)$$

We now ask: What is the result of applying a smooth function to X ? The answer is given by

THEOREM 4.6. (THE GENERAL ITO FORMULA)

Let

$$dX(t) = u dt + v dB(t)$$

be an n -dimensional Ito process as above. Let $g(t, x) = (g_1(t, x), \dots, g_p(t, x))$ be a C^2 map from $[0, \infty) \times \mathbf{R}^n$ into \mathbf{R}^p . Then the process

$$Y(t, \omega) = g(t, X(t))$$

is again an Ito process, whose component nr. k , Y_k , is given by

$$dY_k = \frac{\partial g_k}{\partial t}(t, X)dt + \sum_i \frac{\partial g_k}{\partial x_i}(t, X)dX_i + \frac{1}{2} \sum_{i,j} \frac{\partial^2 g_k}{\partial x_i \partial x_j}(t, X)dX_i dX_j$$

where $dB_i dB_j = \delta_{ij} dt$, $dB_i dt = dt dB_i = 0$.

The proof is similar to the 1-dimensional version (Theorem 4.2) and is omitted.

EXAMPLE 4.7. Let $B = (B_1, \dots, B_n)$ be Brownian motion in \mathbf{R}^n , $n \geq 2$, and consider

$$R(t, \omega) = |B(t, \omega)| = (B_1^2(t, \omega) + \dots + B_n^2(t, \omega))^{\frac{1}{2}},$$

i.e. the distance to the origin of $B(t, \omega)$. The function $g(t, x) = |x|$ is not C^2 at the origin, but since B_t never hits the origin, a.s. when $n \geq 2$ (see Exercise 9.7) Ito's formula still works and we get

$$dR = \sum_{i=1}^n \frac{B_i dB_i}{R} + \frac{n-1}{2R} dt.$$

The process R is called the *n-dimensional Bessel process* because its generator (Chapter VII) is the Bessel differential operator $Af(x) = \frac{1}{2}f''(x) + \frac{n-1}{2x}f'(x)$. See Example 8.9.

The martingale representation theorem

In Chapter III (Corollary 3.14) we proved that an Ito integral

$$X_t = X_0 + \int_0^t v(s, \omega) dB_s \quad v \in \mathcal{V}[0, \infty]; t \geq 0$$

is always a martingale w.r.t. the filtration \mathcal{F}_t generated by $\{B_s(\cdot)\}_{s \leq t}$ (and w.r.t. the probability measure P^0). In this section we will prove that the converse is also true: Any \mathcal{F}_t -martingale (w.r.t. P^0) can be represented as an Ito integral. This result, called the *martingale representation theorem*, is important for many applications, for example in mathematical finance. See e.g. Duffie (1992).

We first establish some auxiliary results.

LEMMA 4.8. Fix $T > 0$. The set of random variables

$$\{\phi(B_{t_1}, \dots, B_{t_n}); t_i \in [0, T], \phi \in C_0^\infty(\mathbf{R}^n), n = 1, 2, \dots\}$$

is dense in $L^2(\mathcal{F}_T, P^0)$.

Proof. Let $\{t_i\}_{i=1}^\infty$ be a dense subset of $[0, T]$ and for each $n = 1, 2, \dots$ let \mathcal{H}_n be the σ -algebra generated by $B_{t_1}(\cdot), \dots, B_{t_n}(\cdot)$. Then clearly

$$\mathcal{H}_n \subset \mathcal{H}_{n+1}$$

and \mathcal{F}_T is the smallest σ -algebra containing all the \mathcal{H}_n 's. Choose $g \in L^2(\mathcal{F}_T, P^0)$. Then by the martingale convergence theorem Corollary C.9 (Appendix C) we have that

$$g = E[g|\mathcal{F}_T] = \lim_{n \rightarrow \infty} E[g|\mathcal{H}_n].$$

The limit is pointwise a.e. (P^0) and in $L^2(\mathcal{F}_T, P^0)$. By the Doob-Dynkin Lemma we can write, for each n ,

$$E[g|\mathcal{H}_n] = g_n(B_{t_1}, \dots, B_{t_n})$$

for some Borel measurable function $g_n : \mathbf{R}^n \rightarrow \mathbf{R}$. Each such $g_n(B_{t_1}, \dots, B_{t_n})$ can be approximated in $L^2(\mathcal{F}_T, P)$ by functions $\phi_n(B_{t_1}, \dots, B_{t_n})$ where $\phi_n \in C_0^\infty(\mathbf{R}^n)$ and the result follows. \square

LEMMA 4.9. The linear span of random variables of the type

$$\exp\left\{\int_0^T h(t)dB_t(\omega) - \frac{1}{2}\int_0^t h^2(t)dt\right\}; \quad h \in L^2[0, T] \text{ (deterministic)} \quad (4.17)$$

is dense in $L^2(\mathcal{F}_T, P^0)$.

Proof. Suppose $g \in L^2(\mathcal{F}_T, P^0)$ is orthogonal (in $L^2(\mathcal{F}_T, P^0)$) to all functions of the form (4.17). Then in particular

$$G(\lambda) := \int_{\Omega} \exp\{\lambda_1 B_{t_1}(\omega) + \dots + \lambda_n B_{t_n}(\omega)\} g(\omega) dP^0(\omega) = 0 \quad (4.18)$$

for all $\lambda = (\lambda_1, \dots, \lambda_n) \in \mathbf{R}^n$ and all $t_1, \dots, t_n \in [0, T]$. The function $G(\lambda)$ is real analytic in $\lambda \in \mathbf{R}^n$ and hence G has an analytic extension to the complex space \mathbf{C}^n given by

$$G(z) = \int_{\Omega} \exp\{z_1 B_{t_1}(\omega) + \dots + z_n B_{t_n}(\omega)\} g(\omega) dP^0(\omega) \quad (4.19)$$

for all $z = (z_1, \dots, z_n) \in \mathbf{C}^n$. (See the estimates in Exercise 2.7 b.) Since $G = 0$ on \mathbf{R}^n and G is analytic, $G = 0$ on \mathbf{C}^n . In particular, $G(iy_1, iy_2, \dots, iy_n) = 0$ for all $y = (y_1, \dots, y_n) \in \mathbf{R}^n$. But then we get, for $\phi \in C_0^\infty(\mathbf{R}^n)$,

$$\begin{aligned}
& \int_{\Omega} \phi(B_{t_1}, \dots, B_{t_n}) g(\omega) dP^0(\omega) \\
&= \int_{\Omega} (2\pi)^{-n/2} \left(\int_{\mathbf{R}^n} \hat{\phi}(y) e^{i(y_1 B_{t_1} + \dots + y_n B_{t_n})} dy \right) g(\omega) dP^0(\omega) \\
&= (2\pi)^{-n/2} \int_{\mathbf{R}^n} \hat{\phi}(y) \left(\int_{\Omega} e^{i(y_1 B_{t_1} + \dots + y_n B_{t_n})} g(\omega) dP^0(\omega) \right) dy \\
&= (2\pi)^{-n/2} \int_{\mathbf{R}^n} \hat{\phi}(y) G(iy) dy = 0,
\end{aligned} \tag{4.20}$$

where

$$\hat{\phi}(y) = (2\pi)^{-n/2} \int_{\mathbf{R}^n} \hat{\phi}(x) e^{-i(x,y)} dy$$

is the Fourier transform of ϕ and we have used the inverse Fourier transform theorem

$$\phi(x) = (2\pi)^{-n/2} \int_{\mathbf{R}^n} \hat{\phi}(y) e^{i(x,y)} dy$$

(see e.g. Folland (1984)).

By (4.20) and Lemma 4.9 g is orthogonal to a dense subset of $L^2(\mathcal{F}_T, P^0)$ and we conclude that $g = 0$. Therefore the linear span of the functions in (4.17) must be dense in $L^2(\mathcal{F}_T, P^0)$ as claimed. \square

If $v(s, \omega) \in \mathcal{V}[0, T]$ then the random variable

$$V(\omega) := \int_0^T v(t, \omega) dB_t(\omega) \tag{4.21}$$

is \mathcal{F}_T -measurable and by the Ito isometry

$$E[V^2] = \int_0^T E[v^2(t, \cdot)] dt < \infty \quad \text{so } V \in L^2(\mathcal{F}_T, P^0).$$

The next result states that any $F \in L^2(\mathcal{F}_T, P^0)$ can be represented this way:

THEOREM 4.10. (The Ito representation theorem)

Let $F \in L^2(\mathcal{F}_T, P^0)$. Then there exists a unique stochastic process $f(t, \omega) \in \mathcal{V}[0, T]$ such that

$$F(\omega) = E[F] + \int_0^T f(t, \omega) dB_t(\omega). \tag{4.22}$$

Proof. First assume that F has the form (4.17), i.e.

$$F(\omega) = \exp\left\{\int_0^T h(t)dB_t(\omega) - \frac{1}{2}\int_0^T h^2(t)dt\right\}$$

for some $h(t) \in L^2[0, T]$.

Define

$$Y_t(\omega) = \exp\left\{\int_0^t h(s)dB_s(\omega) - \frac{1}{2}\int_0^t h^2(s)ds\right\}; \quad 0 \leq t \leq T.$$

Then by Ito's formula

$$dY_t = Y_t(h(t)dB_t - \frac{1}{2}h^2(t)dt) + \frac{1}{2}Y_t(h(t)dB_t)^2 = Y_t h(t)dB_t$$

so that

$$Y_t = 1 + \int_0^t Y_s h(s)dB_s; \quad t \in [0, T].$$

Therefore

$$F = Y_T = 1 + \int_0^T Y_s h(s)dB_s$$

and hence $E[F] = 1$. So (4.22) holds in this case. By linearity (4.22) also holds for linear combinations of functions of the form (4.17). So if $F \in L^2(\mathcal{F}_T, P^0)$ is arbitrary, we approximate F in $L^2(\mathcal{F}_T, P^0)$ by linear combinations F_n of functions of the form (4.17). Then for each n we have

$$F_n(\omega) = E[F_n] + \int_0^T f_n(s, \omega)dB_s(\omega), \quad \text{where } f_n \in \mathcal{V}[0, T].$$

By the Ito isometry

$$\begin{aligned} E[(F_n - F_m)^2] &= E[(E[F_n - F_m] + \int_0^T (f_n - f_m)dB_s)^2] \\ &= (E[F_n - F_m])^2 + \int_0^T E[(f_n - f_m)^2]dt \rightarrow 0 \quad \text{as } n, m \rightarrow \infty \end{aligned}$$

so $\{f_n\}$ is a Cauchy sequence in $L^2([0, T] \times \Omega)$ and hence converges to some $f \in L^2([0, T] \times \Omega)$. Since $f_n \in \mathcal{V}[0, T]$ we have $f \in \mathcal{V}[0, T]$ and, again using the Ito isometry we see that

$$F = \lim_{n \rightarrow \infty} F_n = \lim_{n \rightarrow \infty} (E[F_n] + \int_0^T f_n dB) = E[F] + \int_0^T f dB,$$

the limit being taken in $L^2(\mathcal{F}_T, P^0)$. Hence the representation (4.22) holds for all $F \in L^2(\mathcal{F}_T, P^0)$. \square

The uniqueness follows from the Ito isometry: Suppose

$$F(\omega) = E[F] + \int_0^T f_1(t, \omega) dB_t(\omega) = E[F] + \int_0^T f_2(t, \omega) dB_t(\omega)$$

with $f_1, f_2 \in \mathcal{V}[0, T]$. Then

$$0 = E[(\int_0^T (f_1(t, \omega) - f_2(t, \omega)) dB_t(\omega))^2] = \int_0^T E[(f_1(t, \omega) - f_2(t, \omega))^2] dt$$

and therefore $f_1(t, \omega) = f_2(t, \omega)$ for a.a. $(t, \omega) \in [0, T] \times \Omega$. \square

REMARK. The process $f(t, \omega)$ can be expressed in terms of the Frechet derivative and also in terms of the Malliavin derivative of $F(\omega)$. See Clark (1970/71) and Ocone (1984).

THEOREM 4.11. (The martingale representation theorem)

Suppose M_t is an \mathcal{F}_t -martingale (w.r.t. P^0) and that $M_t \in L^2(P^0)$ for all $t \geq 0$. Then there exists a unique stochastic process $g(s, \omega)$ such that $g \in \mathcal{V}[0, t]$ for all $t \geq 0$ and

$$M_t(\omega) = E[M_0] + \int_0^t g(s, \omega) dB_s(\omega); \quad t \geq 0$$

Proof. By Theorem 4.10 applied to $t = T$, $F = M_t$, we have that for all t there exists a unique $f^{(t)}(s, \omega) \in L^2(\mathcal{F}_t, P^0)$ such that

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

Now assume $0 \leq t_1 < t_2$. Then

$$\begin{aligned} M_{t_1} &= E[M_{t_2} | \mathcal{F}_{t_1}] = E[M_0] + E[\int_0^{t_2} f^{(t_2)}(s, \omega) dB_s(\omega) | \mathcal{F}_{t_1}] \\ &= E[M_0] + \int_0^{t_1} f^{(t_2)}(s, \omega) dB_s(\omega). \end{aligned} \tag{4.23}$$

But we also have

$$M_{t_1} = E[M_0] + \int_0^{t_1} f^{(t_1)}(s, \omega) dB_s(\omega). \quad (4.24)$$

Hence, comparing (4.23) and (4.24) we get that

$$0 = E\left[\left(\int_0^{t_2} (f^{(t_2)} - f^{(t_1)}) dB_s\right)^2\right] = \int_0^{t_2} E[(f^{(t_2)} - f^{(t_1)})^2] ds$$

and therefore

$$f^{(t_1)}(s, \omega) = f^{(t_2)}(s, \omega) \quad \text{for a.a. } (s, \omega) \in [0, t_1] \times \Omega.$$

So we can define $f(s, \omega)$ for a.a. $s \in [0, \infty) \times \Omega$ by setting

$$f(s, \omega) = f^{(N)}(s, \omega) \quad \text{if } s \in [0, N]$$

and then we get

$$M_t = E[M_0] + \int_0^t f^{(t)}(s, \omega) dB_s(\omega) = E[M_0] + \int_0^t f(s, \omega) dB_s(\omega) \quad \text{for all } t \geq 0. \quad \square$$

EXERCISES

4.1.

Use Ito's formula to write the following stochastic processes X_t on the standard form

$$dX_t = u(t, \omega) dt + v(t, \omega) dB_t$$

for suitable choices of $u \in \mathbf{R}^n$, $v \in \mathbf{R}^{n \times m}$ and dimensions n, m :

- a) $X_t = B_t^2$, where B_t is 1-dimensional
- b) $X_t = 2 + t + e^{B_t}$ (B_t is 1-dimensional)
- c) $X_t = B_1^2(t) + B_2^2(t)$ where (B_1, B_2) is 2-dimensional
- d) $X_t = (t_0 + t, B_t)$ (B_t is 1-dimensional)
- e) $X_t = (B_1(t) + B_2(t) + B_3(t), B_2^2(t) - B_1(t)B_3(t))$, where (B_1, B_2, B_3) is 3-dimensional.

4.2.

Use Ito's formula to prove that

$$\int_0^t B_s^2 dB_s = \frac{1}{3} B_t^3 - \int_0^t B_s ds.$$

4.3.

Let X_t, Y_t be Ito processes in \mathbf{R} . Prove that

$$d(X_t Y_t) = X_t dY_t + Y_t dX_t + dX_t \cdot dY_t.$$

Deduce the following general *integration by parts formula*

$$\int_0^t X_s dY_s = X_t Y_t - X_0 Y_0 - \int_0^t Y_s dX_s - \int_0^t dX_s \cdot dY_s.$$

4.4. (Exponential martingales)

Suppose $\theta(t, \omega) = (\theta_1(t, \omega), \dots, \theta_n(t, \omega)) \in \mathbf{R}^n$ with $\theta_k(t, \omega) \in \mathcal{V}[0, T]$ for $k = 1, \dots, n$, where $T \leq \infty$. Define

$$Z_t = \exp\left\{\int_0^t \theta(s, \omega) dB(s) - \frac{1}{2} \int_0^t \theta^2(s, \omega) ds\right\}; \quad 0 \leq t \leq T$$

where $B(s) \in \mathbf{R}^n$ and $\theta^2 = \theta \cdot \theta$ (dot product).

a) Use Ito's formula to prove that

$$dZ_t = Z_t \theta(t, \omega) dB(t).$$

b) Deduce that Z_t is a martingale for $t \leq T$, provided that

$$Z_t \theta_k(t, \omega) \in \mathcal{V}[0, T] \quad \text{for } 1 \leq k \leq n.$$

Remark.

A sufficient condition that Z_t be a martingale is the *Kazamaki condition*

$$E\left[\exp\left(\frac{1}{2} \int_0^t \theta(s, \omega) dB(s)\right)\right] < \infty \quad \text{for all } t \leq T. \quad (4.25)$$

This is implied by the following (stronger)
Novikov condition

$$E\left[\exp\left(\frac{1}{2} \int_0^T \theta^2(s, \omega) ds\right)\right] < \infty. \quad (4.26)$$

See e.g. Ikeda & Watanabe (1989), Section III.5, and the references therein.

4.5.

Let $B_t \in \mathbf{R}$, $B_0 = 0$. Define

$$\beta_k(t) = E[B_t^k]; k = 0, 1, 2, \dots; t \geq 0.$$

Use Ito's formula to prove that

$$\beta_k(t) = \frac{1}{2}k(k-1) \int_0^t \beta_{k-2}(s) ds; k \geq 2.$$

Deduce that

$$E[B_t^4] = 3t^2 \quad (\text{see (2.14)})$$

and find

$$E[B_t^6].$$

4.6.

- a) For c, α constants, $B_t \in \mathbf{R}$ define

$$X_t = e^{ct + \alpha B_t}.$$

Prove that

$$dX_t = (c + \frac{1}{2}\alpha^2)X_t dt + \alpha X_t dB_t.$$

- b) For $c, \alpha_1, \dots, \alpha_n$ constants, $B_t = (B_1(t), \dots, B_n(t)) \in \mathbf{R}^n$ define

$$X_t = \exp(ct + \sum_{j=1}^n \alpha_j B_j(t)).$$

Prove that

$$dX_t = (c + \frac{1}{2} \sum_{j=1}^n \alpha_j^2)X_t dt + X_t \left(\sum_{j=1}^n \alpha_j dB_j \right).$$

4.7.

Let X_t be an Ito integral

$$dX_t = v(t, \omega) dB_t(\omega) \quad \text{where } v \in \mathbf{R}^n, B_t \in \mathbf{R}^n.$$

- a) Give an example to show that X_t^2 is not in general a martingale.
 b) Prove that

$$M_t := X_t^2 - \int_0^t |v_s|^2 ds \quad \text{is a martingale.}$$

The process $\langle X, X \rangle_t := \int_0^t |v_s|^2 ds$ is called the *quadratic variation process* of the martingale X_t . For general processes X_t it is defined by

$$\langle X, X \rangle_t = \lim_{\Delta t_k \rightarrow 0} \sum_{t_k \leq t} |X_{t_{k+1}} - X_{t_k}|^2 \quad (4.27)$$

where $0 = t_1 < t_2 \cdots < t_n = t$ and $\Delta t_k = t_{k+1} - t_k$. The limit can be shown to exist for continuous square integrable martingales X_t . See e.g. Karatzas and Shreve (1988).

4.8.

- a) Let B_t denote n-dimensional Brownian motion and let $f : \mathbf{R}^n \rightarrow \mathbf{R}$ be C^2 . Use Ito's formula to prove that

$$f(B_t) = f(B_0) + \int_0^t \nabla f(B_s) dB_s + \frac{1}{2} \int_0^t \Delta f(B_s) ds,$$

where $\Delta = \sum_{i=1}^n \frac{\partial^2}{\partial x_i^2}$ is the Laplace operator.

- b) Assume that $g : \mathbf{R} \rightarrow \mathbf{R}$ is C^1 everywhere and C^2 outside finitely many points z_1, \dots, z_N with $|g''(x)| \leq M$ for $x \notin \{z_1, \dots, z_N\}$. Let B_t be 1-dimensional Brownian motion. Prove that the 1-dimensional version of a) still holds, i.e.

$$g(B_t) = g(B_0) + \int_0^t g'(B_s) dB_s + \frac{1}{2} \int_0^t g''(B_s) ds.$$

(Hint: Choose $f_k \in C^2(\mathbf{R})$ s.t. $f_k \rightarrow g$ uniformly, $f'_k \rightarrow g'$ uniformly and $|f''_k| \leq M$, $f''_k \rightarrow g''$ outside z_1, \dots, z_N . Apply a) to f_k and let $k \rightarrow \infty$).

4.9.

Prove the first statement in the proof of the Ito formula (Theorem 4.2) by proceeding as follows: For fixed $t \geq 0$ and $n = 1, 2, \dots$ choose g_n as in the statement such that $g_n(s, x) = g(s, x)$ for all $s \leq t$ and all $|x| \leq n$. Suppose we have proved that (4.9) holds for each g_n . Define the stochastic time

$$\tau_n = \tau_n(\omega) = \inf\{s > 0; |X_s(\omega)| \geq n\}$$

(τ_n is called a *stopping time* (See Ch. VII)) and prove that

$$\left(\int_0^t v \frac{\partial g_n}{\partial x}(s, X_s) \mathcal{X}_{s \leq \tau_n} dB_s := \right)$$

$$\int_0^{t \wedge \tau_n} v \frac{\partial g_n}{\partial x}(s, X_s) dB_s = \int_0^{t \wedge \tau_n} v \frac{\partial g}{\partial x}(s, X_s) dB_s$$

for each n . This gives that

$$g(t \wedge \tau_n, X_{t \wedge \tau_n}) = g(0, X_0)$$

$$+ \int_0^{t \wedge \tau_n} \left(\frac{\partial g}{\partial s} + u \frac{\partial g}{\partial x} + \frac{1}{2} v^2 \frac{\partial^2 g}{\partial x^2} \right) ds + \int_0^{t \wedge \tau_n} v \frac{\partial g}{\partial x} dB_s$$

and since

$$P[\tau_n > t] \rightarrow 1 \text{ as } n \rightarrow \infty$$

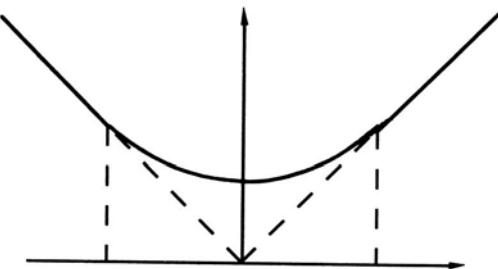
we can conclude that (4.9) holds (a.s.) for g .

4.10. (Tanaka's formula and local time).

What happens if we try to apply the Ito formula to $g(B_t)$ when B_t is 1-dimensional and $g(x) = |x|$? In this case g is not C^2 at $x = 0$, so we modify $g(x)$ near $x = 0$ to $g_\epsilon(x)$ as follows:

$$g_\epsilon(x) = \begin{cases} |x| & \text{if } |x| \geq \epsilon \\ \frac{1}{2}(\epsilon + \frac{x^2}{\epsilon}) & \text{if } |x| < \epsilon \end{cases}$$

where $\epsilon > 0$.



- a) Apply Exercise 4.8 b) to show that

$$g_\epsilon(B_t) = g_\epsilon(B_0) + \int_0^t g'_\epsilon(B_s) dB_s + \frac{1}{2\epsilon} \cdot |\{s \in [0, t]; B_s \in (-\epsilon, \epsilon)\}|$$

where $|F|$ denotes the Lebesgue measure of the set F .

b) Prove that

$$\int_0^t g'_\epsilon(B_s) \cdot \mathcal{X}_{B_s \in (-\epsilon, \epsilon)} dB_s = \int_0^t \frac{B_s}{\epsilon} \cdot \mathcal{X}_{B_s \in (-\epsilon, \epsilon)} dB_s \rightarrow 0$$

in $L^2(P)$ as $\epsilon \rightarrow 0$.

(Hint: Apply the Ito isometry to

$$E\left[\left(\int_0^t \frac{B_s}{\epsilon} \cdot \mathcal{X}_{B_s \in (-\epsilon, \epsilon)} dB_s\right)^2\right]$$

c) By letting $\epsilon \rightarrow 0$ prove that

$$|B_t| = |B_0| + \int_0^t \text{sign}(B_s) dB_s + L_t(\omega), \quad (4.28)$$

where

$$L_t = \lim_{\epsilon \rightarrow 0} \frac{1}{2\epsilon} \cdot |\{s \in [0, t]; B_s \in (-\epsilon, \epsilon)\}| \quad (\text{limit in } L^2(P))$$

and

$$\text{sign}(x) = \begin{cases} -1 & \text{for } x \leq 0 \\ 1 & \text{for } x > 0 \end{cases}$$

L_t is called the *local time* for Brownian motion at 0 and (4.28) is the *Tanaka formula* (for Brownian motion). (See e.g. Rogers and Williams (1987)).

4.11.

Use Ito's formula (for example in the form of Exercise 4.3) to prove that the following stochastic processes are $\{\mathcal{F}_t\}$ -martingales:

- a) $X_t = e^{\frac{1}{2}t} \cos B_t \quad (B_t \in \mathbf{R})$
- b) $X_t = e^{\frac{1}{2}t} \sin B_t \quad (B_t \in \mathbf{R})$
- c) $X_t = (B_t + t)\exp(-B_t - \frac{1}{2}t) \quad (B_t \in \mathbf{R}).$

4.12.

Let $dX_t = u(t, \omega) + v(t, \omega)dB_t$ be an Ito process in \mathbf{R}^n such that

$$E\left[\int_0^t |u(r, \omega)| dr\right] + E\left[\int_0^t |vv^T(r, \omega)| dr\right] < \infty \quad \text{for all } t \geq 0.$$

Suppose X_t is an $\{\mathcal{F}_t^{(n)}\}$ -martingale. Prove that

$$u(s, \omega) = 0 \quad \text{for a.a. } (s, \omega) \in [0, \infty) \times \Omega. \quad (4.29)$$

Remarks:

- 1) This result may be regarded as a special case of the Martingale Representation Theorem.
- 2) The conclusion (4.29) does not hold if the filtration $\mathcal{F}_t^{(n)}$ is replaced by the σ -algebras \mathcal{M}_t generated by $X_s(\cdot); s \leq t$, i.e. if we only assume that X_t is a martingale w.r.t. its own filtration. See e.g. the Brownian motion characterization in Chapter VIII.

Hint for the solution:

If X_t is an $\mathcal{F}_t^{(n)}$ -martingale, then deduce that

$$E\left[\int_t^s u(r, \omega) dr \mid \mathcal{F}_t^{(n)}\right] = 0 \quad \text{for all } s \geq t.$$

Differentiate w.r.t. s to deduce that

$$E[u(s, \omega) \mid \mathcal{F}_t^{(n)}] = 0 \quad \text{a.s., for a.a. } s > t.$$

Then let $t \uparrow s$ and apply Corollary C.9.

4.13.

Let $dX_t = u(t, \omega)dt + dB_t$ ($u \in \mathbf{R}$, $B_t \in \mathbf{R}$) be an Ito process and assume for simplicity that u is bounded. Then from Exercise 4.12 we know that unless $u = 0$ the process X_t is not an \mathcal{F}_t -martingale. However, it turns out that we can construct an \mathcal{F}_t -martingale from X_t by multiplying by a suitable exponential martingale. More precisely, define

$$Y_t = X_t M_t$$

where

$$M_t = \exp\left(-B_t - \frac{1}{2} \int_0^t u^2(r, \omega) dr\right).$$

Use Ito's formula to prove that

$$Y_t \quad \text{is an } \mathcal{F}_t\text{-martingale.}$$

Remarks.

- a) Compare with Exercise 4.11 c).
- b) This result is a special case of the important *Girsanov Theorem*. It can be interpreted as follows: $\{X_t\}_{t \leq T}$ is a martingale w.r.t the measure Q defined on \mathcal{F}_T by

$$dQ = M_T dP^0 \quad (T < \infty).$$

See Chapter VIII.

4.14.

In each of the cases below find the process $f(t, \omega) \in \mathcal{V}[0, T]$ such that (4.22) holds, i.e.

$$F(\omega) = E[F] + \int_0^T f(t, \omega) dB_t(\omega).$$

- | | |
|----------------------------------|--|
| a) $F(\omega) = B_T(\omega)$ | b) $F(\omega) = \int_0^T B_t(\omega) dt$ |
| c) $F(\omega) = B_T^2(\omega)$ | d) $F(\omega) = B_T^3(\omega)$ |
| e) $F(\omega) = e^{B_T(\omega)}$ | f) $F(\omega) = \sin B_T(\omega)$ |

V. Stochastic Differential Equations

We now return to the possible solutions $X_t(\omega)$ of the stochastic differential equation

$$\frac{dX_t}{dt} = b(t, X_t) + \sigma(t, X_t)W_t, \quad b(t, x) \in \mathbf{R}, \sigma(t, x) \in \mathbf{R} \quad (5.1)$$

where W_t is 1-dimensional “white noise”. As discussed in Chapter III the Ito interpretation of (5.1) is that X_t satisfies the stochastic integral equation

$$X_t = X_0 + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dB_s$$

or in differential form

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t. \quad (5.2)$$

Therefore, to get from (5.1) to (5.2) we formally just replace the white noise W_t by $\frac{dB_t}{dt}$ in (5.1) and multiply by dt . It is natural to ask:

- (A) Can one obtain existence and uniqueness theorems for such equations? What are the properties of the solutions?
- (B) How can one solve a given such equation?

We will first consider question (B) by looking at some simple examples, and then discuss (A).

It is the Ito formula that is the key to the solution of many stochastic differential equations. The method is illustrated in the following examples.

EXAMPLE 5.1. Let us return to the population growth model in Chapter I:

$$\frac{dN_t}{dt} = a_t N_t, \quad N_0 \text{ given} \quad (5.3)$$

where $a_t = r_t + \alpha W_t$, W_t = white noise, α = constant.

Let us assume that $r_t = r = \text{constant}$. By the Ito interpretation (5.2) this equation is equivalent to (here $\sigma(t, x) = \alpha x$)

$$dN_t = rN_t dt + \alpha N_t dB_t$$

or

$$\frac{dN_t}{N_t} = rdt + \alpha dB_t.$$

Hence

$$\int_0^t \frac{dN_s}{N_s} = rt + \alpha B_t \quad (B_0 = 0). \quad (5.4)$$

To evaluate the integral on the left hand side we use the Ito formula for the function

$$g(t, x) = \ln x; x > 0$$

and obtain

$$\begin{aligned} d(\ln N_t) &= \frac{1}{N_t} \cdot dN_t + \frac{1}{2} \left(-\frac{1}{N_t^2} \right) (dN_t)^2 \\ &= \frac{dN_t}{N_t} - \frac{1}{2N_t^2} \cdot \alpha^2 N_t^2 dt = \frac{dN_t}{N_t} - \frac{1}{2} \alpha^2 dt. \end{aligned}$$

Hence

$$\frac{dN_t}{N_t} = d(\ln N_t) + \frac{1}{2} \alpha^2 dt$$

so from (5.4) we conclude

$$\ln \frac{N_t}{N_0} = (r - \frac{1}{2} \alpha^2)t + \alpha B_t$$

or

$$N_t = N_0 \exp((r - \frac{1}{2} \alpha^2)t + \alpha B_t) \quad (5.5)$$

For comparison, referring to the discussion at the end of Chapter III, the *Stratonovich* interpretation of (5.3),

$$d\bar{N}_t = r\bar{N}_t dt + \alpha \bar{N}_t \circ dB_t,$$

would have given the solution

$$\bar{N}_t = N_0 \exp(rt + \alpha B_t). \quad (5.6)$$

The solutions N_t, \bar{N}_t are both processes of the type

$$X_t = X_0 \exp(\mu t + \alpha B_t) \quad (\mu, \alpha \text{ constants}).$$

Such processes are called *geometric Brownian motions*. They are important as models for stochastic prices in economics. See Chapters X, XI.

REMARK. It seems reasonable that if B_t is independent of N_0 we should have

$$E[N_t] = E[N_0]e^{rt}, \quad (*)$$

i.e. the same as when there is no noise in a_t . To see if this is indeed the case, we let

$$Y_t = e^{\alpha B_t}$$

and apply Ito's formula:

$$dY_t = \alpha e^{\alpha B_t} dB_t + \frac{1}{2} \alpha^2 e^{\alpha B_t} dt$$

or

$$Y_t = Y_0 + \alpha \int_0^t e^{\alpha B_s} dB_s + \frac{1}{2} \alpha^2 \int_0^t e^{\alpha B_s} ds.$$

Since $E[\int_0^t e^{\alpha B_s} dB_s] = 0$ (Theorem 3.7 (iii)), we get

$$E[Y_t] = E[Y_0] + \frac{1}{2} \alpha^2 \int_0^t E[Y_s] ds$$

i.e.

$$\frac{d}{dt} E[Y_t] = \frac{1}{2} \alpha^2 E[Y_t], E[Y_0] = 1.$$

So

$$E[Y_t] = e^{\frac{1}{2} \alpha^2 t},$$

and therefore – as anticipated – we obtain

$$E[N_t] = E[N_0]e^{rt}.$$

For the Stratonovich solution, however, the same calculation gives

$$E[\bar{N}_t] = E[N_0]e^{(r+\frac{1}{2}\alpha^2)t}.$$

Now that we have found the explicit solutions N_t and \bar{N}_t in (5.5), (5.6) we can use our knowledge about the behaviour of B_t to gain information on these solutions. For example, for the Ito solution N_t we get the following:

- (i) If $r > \frac{1}{2}\alpha^2$ then $N_t \rightarrow \infty$ as $t \rightarrow \infty$, a.s.
- (ii) If $r < \frac{1}{2}\alpha^2$ then $N_t \rightarrow 0$ as $t \rightarrow \infty$, a.s.
- (iii) If $r = \frac{1}{2}\alpha^2$ then N_t will fluctuate between arbitrary large and arbitrary small values as $t \rightarrow \infty$, a.s.

These conclusions are direct consequences of the formula (5.5) for N_t together with the following basic result about 1-dimensional Brownian motion B_t :

THEOREM 5.2 (The law of iterated logarithm)

$$\limsup_{t \rightarrow \infty} \frac{B_t}{\sqrt{2t \log \log t}} = 1 \text{ a.s.}$$

For a proof we refer to Lamperti (1977), §22.

For the *Stratonovich* solution \bar{N}_t we get by the same argument that $\bar{N}_t \rightarrow 0$ a.s. if $r < 0$ and $\bar{N}_t \rightarrow \infty$ a.s. if $r > 0$.

Thus the two solutions have fundamentally different properties and it is an interesting question what solution gives the best description of the situation.

EXAMPLE 5.3. Let us return to the equation in Problem 2 of Chapter I:

$$LQ''_t + RQ'_t + \frac{1}{C}Q_t = F_t = G_t + \alpha W_t. \quad (5.7)$$

We introduce the vector

$$X = X(t, \omega) = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} Q_t \\ Q'_t \end{pmatrix} \text{ and obtain}$$

$$\begin{cases} X'_1 = X_2 \\ LX'_2 = -RX_2 - \frac{1}{C}X_1 + G_t + \alpha W_t \end{cases} \quad (5.8)$$

or, in matrix notation,

$$dX = AXdt + Hdt + KdB_t \quad (5.9)$$

where

$$dX = \begin{pmatrix} dX_1 \\ dX_2 \end{pmatrix}, A = \begin{pmatrix} 0 & 1 \\ -\frac{1}{CL} & -\frac{R}{L} \end{pmatrix}, H_t = \begin{pmatrix} 0 \\ \frac{1}{L}G_t \end{pmatrix}, K = \begin{pmatrix} 0 \\ \frac{\alpha}{L} \end{pmatrix}, \quad (5.10)$$

and B_t is a 1-dimensional Brownian motion.

Thus we are led to a 2-dimensional stochastic differential equation. We rewrite (5.9) as

$$\exp(-At)dX - \exp(-At)AXdt = \exp(-At)[H_tdt + KdB_t], \quad (5.11)$$

where for a general $n \times n$ matrix F we define $\exp(F)$ to be the $n \times n$ matrix given by $\exp(F) = \sum_{n=0}^{\infty} \frac{1}{n!}F^n$. Here it is tempting to relate the left hand side to

$$d(\exp(-At)X).$$

To do this we use a 2-dimensional version of the Ito formula (Theorem 4.6).

Applying this result to the two coordinate functions g_1, g_2 of

$$g : [0, \infty) \times \mathbf{R}^2 \rightarrow \mathbf{R}^2 \text{ given by } g(t, x_1, x_2) = \exp(-At) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

we obtain that

$$d(\exp(-At)X) = (-A)\exp(-At)Xdt + \exp(-At)dX.$$

Substituted in (5.11) this gives

$$\exp(-At)X - X_0 = \int_0^t \exp(-As)H_s ds + \int_0^t \exp(-As)KdB_s$$

or

$$X = \exp(At)[X_0 + \int_0^t \exp(-As)[H_s - AKB_s]ds],$$

by integration by parts (Theorem 4.5).

EXAMPLE 5.4. Choose $X = B$, 1-dimensional Brownian motion, and

$$g(t, x) = e^{ix} = (\cos x, \sin x) \in \mathbf{R}^2 \text{ for } x \in \mathbf{R}.$$

Then

$$Y = g(t, X) = e^{iB} = (\cos B, \sin B)$$

is by Ito's formula again an Ito process.

Its coordinates Y_1, Y_2 satisfy

$$\begin{cases} dY_1(t) = -\sin(B)dB - \frac{1}{2}\cos(B)dt \\ dY_2(t) = \cos(B)dB - \frac{1}{2}\sin(B)dt \end{cases}$$

Thus the process $Y = (Y_1, Y_2)$, which we could call *Brownian motion on the unit circle*, is the solution of the stochastic differential equations

$$\begin{cases} dY_1 = -\frac{1}{2}Y_1dt - Y_2dB \\ dY_2 = -\frac{1}{2}Y_2dt + Y_1dB \end{cases} \quad (5.13)$$

Or, in matrix notation,

$$dY = -\frac{1}{2}Ydt + KYdB, \text{ where } K = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

We now turn to the existence and uniqueness question (A) above.

THEOREM 5.5 (Existence and uniqueness theorem for stochastic differential equations)

Let $T > 0$ and $b(\cdot, \cdot) : [0, T] \times \mathbf{R}^n \rightarrow \mathbf{R}^n, \sigma(\cdot, \cdot) : [0, T] \times \mathbf{R}^n \rightarrow \mathbf{R}^{n \times m}$ be measurable functions satisfying

$$|b(t, x)| + |\sigma(t, x)| \leq C(1 + |x|); \quad x \in \mathbf{R}^n, \quad t \in [0, T] \quad (5.14)$$

for some constant C , (where $|\sigma|^2 = \sum |\sigma_{ij}|^2$) and such that

$$|b(t, x) - b(t, y)| + |\sigma(t, x) - \sigma(t, y)| \leq D|x - y|; \quad x, y \in \mathbf{R}^n, \quad t \in [0, T] \quad (5.15)$$

for some constant D . Let Z be a random variable which is independent of the σ -algebra \mathcal{F}_∞ generated by $B_s(\cdot), s \geq 0$ and such that

$$E[|Z|^2] < \infty. \quad (5.16)$$

Then the stochastic differential equation

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t, \quad 0 \leq t \leq T, \quad X_0 = Z \quad (5.17)$$

has a unique t -continuous solution $X_t(\omega)$, each component of which belongs to $\mathcal{V}[0, T]$.

REMARKS. Conditions (5.14) and (5.15) are natural in view of the following two simple examples from deterministic differential equations (i.e. $\sigma = 0$):

a) The equation

$$\frac{dX_t}{dt} = X_t^2, \quad X_0 = 1 \quad (5.18)$$

corresponding to $b(x) = x^2$ (which does not satisfy (5.14)) has the (unique) solution

$$X_t = \frac{1}{1-t}; \quad 0 \leq t < 1$$

Thus it is impossible to find a global solution (defined for all t) in this case.

More generally, condition (5.14) ensures that the solution $X_t(\omega)$ of (5.17) does not *explode*, i.e. that $|X_t(\omega)|$ does not tend to ∞ in a finite time.

b) The equation

$$\frac{dX_t}{dt} = 3X_t^{2/3}; \quad X(0) = 0 \quad (5.19)$$

has more than one solution. In fact, for any $a > 0$ the function

$$X_t = \begin{cases} 0 & \text{for } t \leq a \\ (t-a)^3 & \text{for } t > a \end{cases}$$

solves (5.19). In this case $b(x) = 3x^{2/3}$ does not satisfy the Lipschitz condition (5.15) at $x = 0$.

Thus condition (5.15) guarantees that equation (5.17) has a unique solution. Here uniqueness means that if $X_1(t, \omega)$ and $X_2(t, \omega)$ are two t-continuous processes in $\mathcal{V}[0, T]$ satisfying (5.17) then

$$X_1(t, \omega) = X_2(t, \omega) \quad \text{for all } t \leq T, \text{ a.s.} \quad (5.20)$$

Proof of Theorem 5.5. The uniqueness follows from the Ito isometry (Corollary 3.7) and the Lipschitz property (5.15): Let $X_1(t, \omega) = X_t(\omega)$ and $X_2(t, \omega) = \hat{X}_t(\omega)$ be solutions with initial values Z, \hat{Z} respectively, i.e. $X_1(0, \omega) = Z(\omega), X_2(0, \omega) = \hat{Z}(\omega), \omega \in \Omega$. For our purposes here we only need the case $Z = \hat{Z}$, but the following more general estimate will be useful for us later, in connection with Feller continuity (Chapter VIII).

Put $a(s, \omega) = b(s, X_s) - b(s, \hat{X}_s)$ and $\gamma(s, \omega) = \sigma(s, X_s) - \sigma(s, \hat{X}_s)$. Then

$$\begin{aligned} E[|X_t - \hat{X}_t|^2] &= E[(Z - \hat{Z} + \int_0^t a ds + \int_0^t \gamma dB_s)^2] \\ &\leq 3 \cdot E[|Z - \hat{Z}|^2] + 3E[(\int_0^t a ds)^2] + 3E[(\int_0^t \gamma dB_s)^2] \\ &\leq 3 \cdot E[|Z - \hat{Z}|^2] + 3tE[\int_0^t a^2 ds] + 3E[\int_0^t \gamma^2 ds] \\ &\leq 3 \cdot E[|Z - \hat{Z}|^2] + 3(1+t)D^2 \int_0^t E[|X_s - \hat{X}_s|^2] ds. \end{aligned}$$

So the function

$$v(t) = E[|X_t - \hat{X}_t|^2]; \quad 0 \leq t \leq T$$

satisfies

$$v(t) \leq F + A \int_0^t v(s) ds, \quad \text{where } F = 3E[|Z - \hat{Z}|^2] \text{ and } A = 3(1+T)D^2. \quad (5.21)$$

Let $w(t) = \int_0^t v(s) ds$. Then $w'(t) \leq F + Aw(t)$. Hence, since $w(0) = 0$, $w(t) \leq \frac{F}{A}(\exp(At) - 1)$. (Consider $f(t) = w(t) \exp(-At)$.) Therefore

$$v(t) \leq F \exp(At). \quad (5.22)$$

Now assume that $Z = \hat{Z}$. Then $F = 0$ and so $v(t) = 0$ for all $t \geq 0$. Hence

$$P[|X_t - \hat{X}_t| = 0 \quad \text{for all } t \in \mathbf{Q} \cap [0, T]] = 1,$$

where \mathbf{Q} denotes the rational numbers.

By continuity of $t \rightarrow |X_t - \hat{X}_t|$ it follows that

$$P[|X_1(t, \omega) - X_2(t, \omega)| = 0 \text{ for all } t \in [0, T]] = 1, \quad (5.23)$$

and the uniqueness is proved.

The proof of the existence is similar to the familiar existence proof for ordinary differential equations:

Define $Y_t^{(0)} = X_0$ and $Y_t^{(k)} = Y_t^{(k)}(\omega)$ inductively as follows

$$Y_t^{(k+1)} = X_0 + \int_0^t b(s, Y_s^{(k)}) ds + \int_0^t \sigma(s, Y_s^{(k)}) dB_s. \quad (5.24)$$

Then, similar computation as for the uniqueness above gives

$$E[|Y_t^{(k+1)} - Y_t^{(k)}|^2] \leq (1 + T)3D^2 \int_0^t E[|Y_s^{(k)} - Y_s^{(k-1)}|^2] ds,$$

for $k \geq 1, t \leq T$ and

$$\begin{aligned} E[|Y_t^{(1)} - Y_t^{(0)}|^2] &\leq 2C^2 t^2 (1 + E[|X_0|^2]) \\ &+ 2C^2 t (1 + E[|X_0|^2]) \leq A_1 t \end{aligned}$$

where the constant A_1 only depends on C, T and $E[|X_0|^2]$. So by induction on k we obtain

$$E[|Y_t^{(k+1)} - Y_t^{(k)}|^2] \leq \frac{A_2^{k+1} t^{k+1}}{(k+1)!}; \quad k \geq 0, t \in [0, T] \quad (5.25)$$

for some suitable constant A_2 depending only on C, D, T and $E[|X_0|^2]$. Now

$$\begin{aligned} \sup_{0 \leq t \leq T} |Y_t^{(k+1)} - Y_t^{(k)}| &\leq \int_0^T |b(s, Y_s^{(k)}) - b(s, Y_s^{(k-1)})| ds \\ &+ \sup_{0 \leq t \leq T} \left| \int_0^t (\sigma(s, Y_s^{(k)}) - \sigma(s, Y_s^{(k-1)})) dB_s \right|. \end{aligned}$$

By the martingale inequality (Theorem 3.12) we obtain

$$\begin{aligned}
P[\sup_{0 \leq t \leq T} |Y_t^{(k+1)} - Y_t^{(k)}| > 2^{-k}] &\leq P\left[\left(\int_0^T |b(s, Y_s^{(k)}) - b(s, Y_s^{(k-1)})| ds\right)^2 > 2^{-2k-2}\right] \\
&+ P\left[\sup_{0 \leq t \leq T} \left|\int_0^t (\sigma(s, Y_s^{(k)}) - \sigma(s, Y_s^{(k-1)})) dB_s\right| > 2^{-k-1}\right] \\
&\leq 2^{2k+2} T \int_0^T E(|b(s, Y_s^{(k)}) - b(s, Y_s^{(k-1)})|^2) dt \\
&+ 2^{2k+2} \int_0^T E[|\sigma(s, Y_s^{(k)}) - \sigma(s, Y_s^{(k-1)})|^2] ds \\
&\leq 2^{2k+2} D^2(T+1) \int_0^T \frac{A_2^k t^k}{k!} dt \leq \frac{(4A_2 T)^{k+1}}{(k+1)!}, \text{ if } A_2 \geq 4D^2(T+1).
\end{aligned}$$

Therefore, by the Borel-Cantelli lemma,

$$P[\sup_{0 \leq t \leq T} |Y_t^{(k+1)} - Y_t^{(k)}| > 2^{-k} \text{ for infinitely many } k] = 0.$$

Thus, for a.a. ω there exists $k_0 = k_0(\omega)$ such that

$$\sup_{0 \leq t \leq T} |Y_t^{(k+1)} - Y_t^{(k)}| < 2^{-k} \text{ for } k \geq k_0.$$

Therefore the sequence

$$Y_t^{(n)}(\omega) = Y_t^{(0)}(\omega) + \sum_{k=0}^{n-1} (Y_t^{(k+1)}(\omega) - Y_t^{(k)}(\omega))$$

is uniformly convergent in $[0, T]$, for a.a. ω .

Denote the limit by $X_t = X_t(\omega)$. Then X_t is t-continuous for a.a. ω since $Y_t^{(n)}$ is t-continuous for all n . Moreover, $X_t(\cdot)$ is \mathcal{F}_t -measurable for all t , since $Y_t^{(n)}(\cdot)$ has this property for all n .

Next, note that for $m > n \geq 0$ we have by (5.25)

$$\begin{aligned}
E[|Y_t^{(m)} - Y_t^{(n)}|^2]^{1/2} &= \|Y_t^{(m)} - Y_t^{(n)}\|_{L^2(P)} = \left\| \sum_{k=n}^{m-1} (Y_t^{(k+1)} - Y_t^{(k)}) \right\|_{L^2(P)} \\
&\leq \sum_{k=n}^{m-1} \|Y_t^{(k+1)} - Y_t^{(k)}\|_{L^2(P)} \leq \sum_{k=n}^{\infty} \left[\frac{(A_2 t)^{k+1}}{(k+1)!} \right]^{1/2} \rightarrow 0 \text{ as } n \rightarrow \infty. \quad (5.26)
\end{aligned}$$

So $\{Y_t^{(n)}\}$ converges in $L^2(P)$ to a limit Y_t , say. A subsequence of $Y_t^{(n)}(\omega)$ will then converge ω -pointwise to $Y_t(\omega)$ and therefore we must have $Y_t = X_t$ a.s. In particular, $E[|X_t|^2] < \infty$ so $X \in N[0, T]$.

It remains to show that X_t satisfies (5.17). For all n we have

$$Y_t^{(n+1)} = X_0 + \int_0^t b(s, Y_s^{(n)}) ds + \int_0^t \sigma(s, Y_s^{(n)}) dB_s \quad (5.27)$$

Now $Y_t^{(n+1)} \rightarrow X_t$ as $n \rightarrow \infty$, uniformly in $t \in [0, T]$ for a.a. ω . By (5.26) and the Fatou lemma we have

$$E\left[\int_0^T |X_t - Y_t^{(n)}|^2 dt\right] \leq \limsup_{m \rightarrow \infty} E\left[\int_0^T |Y_t^{(m)} - Y_t^{(n)}|^2 dt\right] \rightarrow 0$$

as $n \rightarrow \infty$. It follows by the Ito isometry that

$$\int_0^t \sigma(s, Y_s^{(n)}) dB_s \rightarrow \int_0^t \sigma(s, X_s) dB_s$$

and by the Hölder inequality that

$$\int_0^t b(s, Y_s^{(n)}) ds \rightarrow \int_0^t b(s, X_s) ds$$

in $L^2(P^0)$. Therefore, taking the limit of (5.27) as $n \rightarrow \infty$ we obtain (5.17) for X_t .

REMARKS.

- 1) The solution X_t found above is called a *strong* solution, because the version B_t of Brownian motion is given in advance and the solution X_t constructed from it is \mathcal{F}_t -adapted. If we are only given the functions $b(t, x)$ and $\sigma(t, x)$ and ask for a pair of processes $((\tilde{X}_t, \tilde{B}_t), \mathcal{H}_t)$ on a probability space $(\Omega, \mathcal{H}, P^0)$ such that (5.17) holds, then the solution \tilde{X}_t (or more precisely $(\tilde{X}_t, \tilde{B}_t)$) is called a *weak* solution. Here \mathcal{H}_t is an increasing family of σ -algebras such that \tilde{X}_t is \mathcal{H}_t -adapted and \tilde{B}_t is an \mathcal{H}_t -Brownian motion, i.e. \tilde{B}_t is a Brownian motion, and \tilde{B}_t is a martingale w.r.t. \mathcal{H}_t (and so $E[\tilde{B}_{t+h} - \tilde{B}_t | \mathcal{H}_t] = 0$ for all $t, h \geq 0$). Recall from Chapter III that this allows us to define the Ito integral on the right hand side of (5.17) exactly as before, even though \tilde{X}_t need not be \mathcal{F}_t -adapted.

A strong solution is of course also a weak solution, but the converse is not true in general. See the example of Tanaka, explained in Rogers and Williams (1987), Sec.V.16.

2) The uniqueness (5.20) that we obtain above is called *strong* or *pathwise* uniqueness, while *weak* uniqueness simply means that any two solutions (weak or strong) are identical in law, i.e. have the same finite-dimensional distributions. See Stroock and Varadhan (1979) for results about existence and uniqueness of weak solutions. A general discussion about strong and weak solutions can be found in Krylov and Zvonkin (1981).

3) If b and σ satisfy the conditions of Theorem 5.5 then we have

$$\text{A solution (weak or strong) of (5.17) is weakly unique.} \quad (5.28)$$

To see this we argue as follows:

Let $((\tilde{X}_t, \tilde{B}_t), \tilde{\mathcal{H}}_t)$ and $((\hat{X}_t, \hat{B}_t), \hat{\mathcal{H}}_t)$ be two weak solutions. Let X_t and Y_t be the *strong* solutions constructed from \tilde{B}_t and \hat{B}_t , respectively, as above. Then the same uniqueness argument as above applies to show that $X_t = \tilde{X}_t$ and $Y_t = \hat{X}_t$ for all t , a.s. Therefore it suffices to show that X_t and Y_t must be identical in law. We show this by proving by induction that if $X_t^{(k)}, Y_t^{(k)}$ are the processes in the Picard iteration defined by (5.24) with Brownian motions \tilde{B}_t and \hat{B}_t , then

$$(X_t^{(k)}, \tilde{B}_t) \text{ and } (Y_t^{(k)}, \hat{B}_t)$$

have the same law for all k .

This observation will be useful for us in Chapter VII and later, where we will investigate further the properties of processes which are solutions of stochastic differential equations (Ito diffusions).

EXERCISES

5.1.

Verify that the given processes solve the given corresponding stochastic differential equations: (B_t denotes 1-dimensional Brownian motion)

- (i) $X_t = e^{B_t}$ solves $dX_t = \frac{1}{2}X_t dt + X_t dB_t$
- (ii) $X_t = \frac{B_t}{1+t}; B_0 = 0$ solves

$$dX_t = -\frac{1}{1+t}X_t dt + \frac{1}{1+t}dB_t; X_0 = 0$$

- (iii) $X_t = \sin B_t$ with $B_0 = a \in (-\frac{\pi}{2}, \frac{\pi}{2})$ solves

$$dX_t = -\frac{1}{2}X_t dt + \sqrt{1-X_t^2} dB_t \text{ for } t < T(\omega) = \inf\{s > 0; B_s \notin [-\frac{\pi}{2}, \frac{\pi}{2}]\}$$

- (iv) $(X_1(t), X_2(t)) = (t, e^t B_t)$ solves

$$\begin{bmatrix} dX_1 \\ dX_2 \end{bmatrix} = \begin{bmatrix} 1 \\ X_2 \end{bmatrix} dt + \begin{bmatrix} 0 \\ e^{X_1} \end{bmatrix} dB_t$$

(v) $(X_1(t), X_2(t)) = (\cosh(B_t), \sinh(B_t))$ solves

$$\begin{bmatrix} dX_1 \\ dX_2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} dt + \begin{bmatrix} X_2 \\ X_1 \end{bmatrix} dB_t$$

5.2.

A natural candidate for what we could call *Brownian motion on the ellipse*

$$\{(x, y); \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1\} \quad \text{where } a > 0, b > 0$$

is the process $X_t = (X_1(t), X_2(t))$ defined by

$$X_1(t) = a \cos B_t, X_2(t) = b \sin B_t$$

where B_t is 1-dimensional Brownian motion. Show that X_t is a solution of the stochastic differential equation

$$dX_t = -\frac{1}{2} X_t dt + M X_t dB_t$$

where $M = \begin{bmatrix} 0 & -\frac{a}{b} \\ \frac{b}{a} & 0 \end{bmatrix}$.

5.3.

Let (B_1, \dots, B_n) be Brownian motion in \mathbf{R}^n , $\alpha_1, \dots, \alpha_n$ constants. Solve the stochastic differential equation

$$dX_t = r X_t dt + X_t \left(\sum_{k=1}^n \alpha_k dB_k(t) \right); X_0 > 0.$$

(This is a model for exponential growth with several independent white noise sources in the relative growth rate).

5.4.

Solve the following stochastic differential equations:

(i) $\begin{bmatrix} dX_1 \\ dX_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} dt + \begin{bmatrix} 1 & 0 \\ 0 & X_1 \end{bmatrix} \begin{bmatrix} dB_1 \\ dB_2 \end{bmatrix}$

(ii) $dX_t = X_t dt + dB_t$

(Hint: Multiply both sides with “the integrating factor” e^{-t} and compare with $d(e^{-t} X_t)$)

(iii) $dX_t = -X_t dt + e^{-t} dB_t$

5.5.

- a) Solve the *Ornstein-Uhlenbeck equation* (or *Langevin equation*)

$$dX_t = \mu X_t dt + \sigma dB_t$$

where μ, σ are real constants, $B_t \in \mathbf{R}$.

The solution is called the *Ornstein-Uhlenbeck process*. (Hint: See Exercise 5.4 (ii).)

- b) Find $E[X_t]$ and $\text{Var}[X_t] := E[(X_t - E[X_t])^2]$.

5.6.

Solve the stochastic differential equation

$$dY_t = r dt + \alpha Y_t dB_t$$

where r, α are real constants, $B_t \in \mathbf{R}$.

(Hint: Multiply the equation by the 'integrating factor'

$$F_t = \exp(-\alpha B_t + \frac{1}{2}\alpha^2 t).$$

5.7.

The *mean-reverting Ornstein-Uhlenbeck process* is the solution X_t of the stochastic differential equation

$$dX_t = (m - X_t)dt + \sigma dB_t$$

where m, σ are real constants, $B_t \in \mathbf{R}$.

- a) Solve this equation by processing as in Exercise 5.5 a).
 b) Find $E[X_t]$ and $\text{Var}[X_t] := E[(X_t - E[X_t])^2]$.

5.8.

Solve the (2-dimensional) stochastic differential equation

$$\begin{aligned} dX_1(t) &= X_2(t)dt + \alpha dB_1(t) \\ dX_2(t) &= X_1(t)dt + \beta dB_2(t) \end{aligned}$$

where $(B_1(t), B_2(t))$ is 2-dimensional Brownian motion and α, β are constants.

This is a model for a vibrating string subject to a stochastic force. See Example 5.3.

5.9.

Show that there is a unique strong solution X_t of the 1-dimensional stochastic differential equation

$$dX_t = \ln(1 + X_t^2)dt + \chi_{\{X_t > 0\}} X_t dB_t, X_0 = a \in \mathbf{R}.$$

5.10.

Let b, σ satisfy (5.14), (5.15) and let X_t be the unique strong solution of (5.17). Show that

$$E[|X_t|^2] \leq K_1 \cdot \exp(K_2 t) \text{ for } t \leq T \quad (5.29)$$

where $K_1 = 3E[|Z|^2] + 6C^2T(T+1)$ and $K_2 = 6(1+T)C^2$.

(Hint: Use the argument in the proof of (5.22)).

Remark. With global estimates of the growth of b and σ in (5.17) it is possible to improve (5.29) to a global estimate of $E[|X_t|^2]$. See Exercise 7.5.

5.11.

(The Brownian bridge). For fixed $a, b \in \mathbf{R}$ consider the following 1-dimensional equation

$$dY_t = \frac{b - Y_t}{1 - t} dt + dB_t; \quad 0 \leq t < 1, \quad Y_0 = a. \quad (5.30)$$

Verify that

$$Y_t = a(1-t) + bt + (1-t) \int_0^t \frac{dB_s}{1-s}; \quad 0 \leq t < 1 \quad (5.31)$$

solves the equation and prove that $\lim_{t \rightarrow 1^-} Y_t = b$ a.s. The process Y_t is called *the Brownian bridge* (from a to b). For other characterizations of Y_t see Rogers and Williams (1987, pp. 86–89).

5.12.

To describe the motion of a pendulum with small, random perturbations in its environment we try an equation of the form

$$y''(t) + (1 + \epsilon W_t)y = 0; \quad y(0), y'(0) \text{ given,}$$

where $W_t = \frac{dB_t}{dt}$ is 1-dimensional white noise, $\epsilon > 0$ is constant.

- a) Discuss this equation, for example by proceeding as in Example 5.3.
- b) Show that $y(t)$ solves a *stochastic Volterra equation* of the form

$$y(t) = y(0) + y'(0) \cdot t + \int_0^t a(t, r)y(r)dr + \int_0^t \gamma(t, r)y(r)dB_r$$

where $a(t, r) = r - t$, $\gamma(t, r) = \epsilon(r - t)$.

5.13.

As a model for the horizontal slow drift motions of a moored floating platform or ship responding to incoming irregular waves John Grue (1989) introduced the equation

$$x_t'' + a_0 x_t' + w^2 x_t = (T_0 - \alpha_0 x_t') \eta W_t, \quad (5.32)$$

where W_t is 1-dimensional white noise, a_0, w, T_0, α_0 and η are constants.

(i) Put $X_t = \begin{bmatrix} x_t \\ x_t' \end{bmatrix}$ and rewrite the equation in the form

$$dX_t = AX_t dt + KX_t dB_t + MdB_t,$$

where

$$A = \begin{bmatrix} 0 & 1 \\ -w^2 & -a_0 \end{bmatrix}, \quad K = \alpha_0 \eta \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix} \text{ and } M = T_0 \eta \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

(ii) Show that X_t satisfies the integral equation

$$X_t = \int_0^t e^{A(t-s)} K X_s dB_s + \int_0^t e^{A(t-s)} M dB_s \text{ if } X_0 = 0$$

(iii) Verify that

$$e^{At} = \frac{e^{-\lambda t}}{\xi} \{ (\xi \cos \xi t + \lambda \sin \xi t) I + A \sin \xi t \}$$

where $\lambda = \frac{a_0}{2}$, $\xi = (w^2 - \frac{a_0^2}{4})^{\frac{1}{2}}$ and use this to prove that

$$x_t = \eta \int_0^t (T_0 - \alpha_0 y_s) g_{t-s} dB_s \quad (5.33)$$

and

$$y_t = \eta \int_0^t (T_0 - \alpha_0 y_s) h_{t-s} dB_s, \quad \text{with } y_t := x'_t, \quad (5.34)$$

where

$$g_t = \frac{1}{\xi} \operatorname{Im}(e^{\zeta t})$$

$$h_t = \frac{1}{\xi} \operatorname{Im}(\zeta e^{\zeta t}), \quad \zeta = -\lambda + i\xi \quad (i = \sqrt{-1}).$$

So we can solve for y_t first in (5.34) and then substitute in (5.33) to find x_t .

5.14.

If (B_1, B_2) denotes 2-dimensional Brownian motion we may introduce complex notation and put

$$\mathbf{B}(t) := B_1(t) + iB_2(t) \quad (i = \sqrt{-1}).$$

$\mathbf{B}(t)$ is called *complex Brownian motion*.

- (i) If $F(z) = u(z) + iv(z)$ is an *analytic* function i.e. F satisfies the Cauchy-Riemann equations

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}; \quad z = x + iy$$

and we define

$$Z_t = F(\mathbf{B}(t))$$

prove that

$$dZ_t = F'(\mathbf{B}(t))d\mathbf{B}(t), \quad (5.35)$$

where F' is the (complex) derivative of F . (Note that the usual second order terms in the (real) Ito formula are not present in (5.35)!!)

- (ii) Solve the complex stochastic differential equation

$$dZ_t = \alpha Z_t d\mathbf{B}(t) \quad (\alpha \text{ constant})$$

For more information about complex stochastic calculus involving analytic functions see e.g. Ubøe (1987).

VI. The Filtering Problem

Problem 3 in the introduction is a special case of the following general *filtering problem*:

Suppose the state $X_t \in \mathbf{R}^n$ at time t of a system is given by a stochastic differential equation

$$\frac{dX_t}{dt} = b(t, X_t) + \sigma(t, X_t)W_t, \quad t \geq 0, \quad (6.1)$$

where $b : \mathbf{R}^{n+1} \rightarrow \mathbf{R}^n, \sigma : \mathbf{R}^{n+1} \rightarrow \mathbf{R}^{n \times p}$ satisfy conditions (5.14), (5.15) and W_t is p -dimensional white noise. As discussed earlier the Ito interpretation of this equation is

$$(\text{system}) \quad dX_t = b(t, X_t)dt + \sigma(t, X_t)dU_t, \quad (6.2)$$

where U_t is p -dimensional Brownian motion. We also assume that the distribution of X_0 is known and independent of U_t . Similarly to the 1-dimensional situation (3.20) there is an explicit several-dimensional formula which expresses the *Stratonovich* interpretation of (6.1):

$$dX_t = b(t, X_t)dt + \sigma(t, X_t) \circ dB_t$$

in terms of Ito integrals as follows:

$$dX_t = \tilde{b}(t, X_t)dt + \sigma(t, X_t)dU_t, \quad \text{where} \\ \tilde{b}_i(t, x) = b_i(t, x) + \frac{1}{2} \sum_{j=1}^p \sum_{k=1}^n \frac{\partial \sigma_{ij}}{\partial x_k} \sigma_{kj}; \quad 1 \leq i \leq n. \quad (6.3)$$

(See Stratonovich (1966)). From now on we will use the Ito interpretation (6.2).

In the continuous version of the filtering problem we assume that the observations $H_t \in \mathbf{R}^m$ are performed continuously and are of the form

$$H_t = c(t, X_t) + \gamma(t, X_t) \cdot \tilde{W}_t, \quad (6.4)$$

where $c : \mathbf{R}^{n+1} \rightarrow \mathbf{R}^m, \gamma : \mathbf{R}^{n+1} \rightarrow \mathbf{R}^{m \times r}$ are functions satisfying (5.14) and \tilde{W}_t denotes r -dimensional white noise, independent of U_t and X_0 .

To obtain a tractable mathematical interpretation of (6.4) we introduce

$$Z_t = \int_0^t H_s ds \quad (6.5)$$

and thereby we obtain the stochastic integral representation

$$(observations) \quad dZ_t = c(t, X_t)dt + \gamma(t, X_t)dV_t, Z_0 = 0 \quad (6.6)$$

where V_t is r-dimensional Brownian motion, independent of U_t and X_0 .

Note that if H_s is known for $0 \leq s \leq t$, then Z_s is also known for $0 \leq s \leq t$ and vice versa. So no information is lost or gained by considering Z_t as our “observations” instead of H_t . But this allows us to obtain a well-defined mathematical model of the situation.

The filtering problem is the following:

Given the observations Z_s satisfying (6.6) for $0 \leq s \leq t$, what is the best estimate \hat{X}_t of the state X_t of the system (6.2) based on these observations?

As we have pointed out earlier, it is necessary to find a precise mathematical formulation of this problem:

By saying that the estimate \hat{X}_t is based on the observations $\{Z_s; s \leq t\}$ we mean that

$$\begin{aligned} \hat{X}_t(\cdot) &\text{ is } \mathcal{G}_t\text{-measurable,} \\ \text{where } \mathcal{G}_t &\text{ is the } \sigma\text{-algebra generated by } \{Z_s(\cdot), s \leq t\}. \end{aligned} \quad (6.7)$$

By saying that \hat{X}_t is the best such estimate we mean that

$$\int_{\Omega} |X_t - \hat{X}_t|^2 dP = E[|X_t - \hat{X}_t|^2] = \inf\{E[|X_t - Y|^2]; Y \in \mathcal{K}\} \quad (6.8)$$

Here – and in the rest of this chapter – (Ω, \mathcal{F}, P) is the probability space corresponding to the $(p+r)$ -dimensional Brownian motion (U_t, V_t) starting at 0, E denotes expectation w.r.t. P and

$$\mathcal{K} := \mathcal{K}_t := \mathcal{K}(Z, t) := \{Y: \Omega \rightarrow \mathbf{R}^n; Y \in L^2(P) \text{ and } Y \text{ is } \mathcal{G}_t\text{-measurable}\}, \quad (6.9)$$

where $L^2(P) = L^2(\Omega, P)$.

Having found the mathematical formulation of our problem, we now start to study the properties of the solution \hat{X}_t .

We first establish the following useful connection between conditional expectation and projection:

LEMMA 6.1. Let $\mathcal{H} \subset \mathcal{F}$ be a σ -algebra and let $X \in L^2(P)$ be \mathcal{F} -measurable. Put $\mathcal{N} = \{Y \in L^2(P); Y \text{ is } \mathcal{H}\text{-measurable}\}$ and let $\mathcal{P}_{\mathcal{N}}$ denote the (orthogonal) projection from the Hilbert space $L^2(P)$ into the subspace \mathcal{N} . Then

$$\mathcal{P}_{\mathcal{N}}(X) = E[X|\mathcal{H}].$$

Proof. Recall (see Appendix B) that $E[X|\mathcal{H}]$ is by definition the P -unique function from Ω to \mathbf{R} such that

- (i) $E[X|\mathcal{H}]$ is \mathcal{H} -measurable
- (ii) $\int_A E[X|\mathcal{H}] dP = \int_A X dP$ for all $A \in \mathcal{H}$.

Now $\mathcal{P}_{\mathcal{N}}(X)$ is \mathcal{H} -measurable and

$$\int_{\Omega} Y(X - \mathcal{P}_{\mathcal{N}}(X)) dP = 0 \text{ for all } Y \in \mathcal{N}.$$

In particular,

$$\int_A (X - \mathcal{P}_{\mathcal{N}}(X)) dP = 0 \text{ for all } A \in \mathcal{H}$$

i.e.

$$\int_A \mathcal{P}_{\mathcal{N}}(X) dP = \int_A X dP \text{ for all } A \in \mathcal{H}.$$

Hence, by uniqueness, $\mathcal{P}_{\mathcal{N}}(X) = E[X|\mathcal{H}]$.

From the general theory of Hilbert spaces we know that the solution \hat{X}_t of the problem (6.8) is given by the projection $\mathcal{P}_{\mathcal{K}_t}(X_t)$. Therefore Lemma 6.1 leads to the following useful result:

THEOREM 6.2.

$$\hat{X}_t = \mathcal{P}_{\mathcal{K}_t}(X_t) = E[X_t|\mathcal{G}_t]$$

This is the basis for the general Fujisaki-Kallianpur-Kunita equation of filtering theory. See for example Davis (1984) or Kallianpur (1980).

Here we will concentrate on the linear case, which allows an explicit solution in terms of a stochastic differential equation for \hat{X}_t (the Kalman-Bucy filter):

In the *linear* filtering problem the system and observation equations have the form:

$$(\text{linear system}) \quad dX_t = F(t)X_t dt + C(t)dU_t; \quad F(t) \in \mathbf{R}^{n \times n}, C(t) \in \mathbf{R}^{n \times p} \quad (6.10)$$

$$(\text{linear observations}) \quad dZ_t = G(t)X_t dt + D(t)dV_t; \quad G(t) \in \mathbf{R}^{m \times n}, D(t) \in \mathbf{R}^{m \times r} \quad (6.11)$$

To be able to focus on the main ideas in the solution of the filtering problem, we will consider only the 1-dimensional case:

$$\text{(linear system)} \quad dX_t = F(t)X_t dt + C(t)dU_t; \quad F(t), C(t) \in \mathbf{R} \quad (6.12)$$

$$\text{(linear observations)} \quad dZ_t = G(t)X_t dt + D(t)dV_t; \quad G(t), D(t) \in \mathbf{R} \quad (6.13)$$

We assume (see (5.14)) that F, G, C, D are bounded on bounded intervals. Based on our interpretation (6.5) of Z_t we assume $Z_0 = 0$. We also assume that X_0 is normally distributed (and independent of $\{U_t\}, \{V_t\}$). Finally we assume that $D(t)$ is bounded away from 0 on bounded intervals.

The (important) extension to the several-dimensional case (6.10), (6.11) is technical, but does not require any essentially new ideas. Therefore we shall only state the result for this case after we have discussed the 1-dimensional situation. The reader is encouraged to work out the necessary modifications for the general case for himself or consult Kallianpur (1980) or Davis (1977) for a full treatment.

From now on we let X_t, Z_t be processes satisfying (6.12), (6.13). Here is an outline of the solution of the filtering problem in this case.

Step 1. Let $\mathcal{L} = \mathcal{L}(Z, t)$ be the closure in $L^2(P)$ of functions which are linear combinations of the form

$$c_0 + c_1 Z_{s_1}(\omega) + \cdots + c_k Z_{s_k}(\omega), \text{ with } s_j \leq t, c_j \in \mathbf{R}.$$

Let

$$\mathcal{P}_{\mathcal{L}} \text{ denote the projection from } L^2(P) \text{ onto } \mathcal{L}.$$

Then, with \mathcal{K} as in (6.9),

$$\hat{X}_t = \mathcal{P}_{\mathcal{K}}(X_t) = E[X_t | \mathcal{G}_t] = \mathcal{P}_{\mathcal{L}}(X_t).$$

Thus, the best Z -measurable estimate of X_t coincides with the best Z -linear estimate of X_t .

Step 2. Replace Z_t by the innovation process N_t :

$$N_t = Z_t - \int_0^t (GX)_s^\wedge ds, \text{ where } (GX)_s^\wedge = \mathcal{P}_{\mathcal{L}(Z, s)}(G(s)X_s) = G(s)\hat{X}_s.$$

Then

- (i) N_t has orthogonal increments, i.e.
 $E[(N_{t_1} - N_{s_1})(N_{t_2} - N_{s_2})] = 0$ for non-overlapping intervals $[s_1, t_1], [s_2, t_2]$.
- (ii) $\mathcal{L}(N, t) = \mathcal{L}(Z, t)$, so $\hat{X}_t = \mathcal{P}_{\mathcal{L}(N, t)}(X_t)$.

Step 3. If we put

$$dR_t = \frac{1}{D(t)} dN_t,$$

then R_t is a 1-dimensional Brownian motion. Moreover,

$$\mathcal{L}(N, t) = \mathcal{L}(R, t) \text{ and}$$

$$\hat{X}_t = \mathcal{P}_{\mathcal{L}(N, t)}(X_t) = \mathcal{P}_{\mathcal{L}(R, t)}(X_t) = E[X_t] + \int_0^t \frac{\partial}{\partial s} E[X_t R_s] dR_s.$$

Step 4. Find an expression for X_t by solving the (linear) stochastic differential equation

$$dX_t = F(t)X_t dt + C(t)dU_t.$$

Step 5. Substitute the formula for X_t from Step 4 into $E[X_t R_s]$ and use Step 3 to obtain a stochastic differential equation for \hat{X}_t :

$$d\hat{X}_t = \frac{\partial}{\partial s} E[X_t R_s]_{s=t} dR_t + \left(\int_0^t \frac{\partial^2}{\partial t \partial s} E[X_t R_s] dR_s \right) dt \text{ etc.}$$

Before we proceed to establish Steps 1-5, let us consider a simple, but motivating example:

EXAMPLE 6.3. Suppose X, W_1, W_2, \dots are independent real random variables, $E[X] = E[W_j] = 0$ for all j , $E[X^2] = a^2$, $E[W_j^2] = m^2$ for all j . Put $Z_j = X + W_j$.

What is the best linear estimate \hat{X} of X based on $\{Z_j; j \leq k\}$? More precisely, let

$$\mathcal{L} = \mathcal{L}(Z, k) = \{c_1 Z_1 + \dots + c_k Z_k; c_1, \dots, c_k \in \mathbf{R}\}.$$

Then we want to find

$$\hat{X}_k = \mathcal{P}_k(X),$$

where \mathcal{P}_k denotes the projection into $\mathcal{L}(Z, k)$.

We use the Gram-Schmidt procedure to obtain random variables A_1, A_2, \dots such that

- (i) $E[A_i A_j] = 0$ for $i \neq j$
- (ii) $\mathcal{L}(A, k) = \mathcal{L}(Z, k)$ for all k .

Then

$$\hat{X}_k = \sum_{j=1}^k \frac{E[X A_j]}{E[A_j^2]} A_j \quad \text{for } k = 1, 2, \dots \quad (6.14)$$

We obtain a recursive relation between \hat{X}_k and \hat{X}_{k-1} from this by observing that

$$A_j = Z_j - \hat{X}_{j-1}, \quad (6.15)$$

which follows from

$$A_j = Z_j - \mathcal{P}_{j-1}(Z_j) = Z_j - \mathcal{P}_{j-1}(X), \text{ since } \mathcal{P}_{j-1}(W_j) = 0.$$

By (6.15)

$$E[XA_j] = E[X(Z_j - \hat{X}_{j-1})] = E[X(X - \hat{X}_{j-1})] = E[(X - \hat{X}_{j-1})^2]$$

and

$$E[A_j^2] = E[(X + W_j - \hat{X}_{j-1})^2] = E[(X - \hat{X}_{j-1})^2] + m^2.$$

Hence

$$\hat{X}_k = \hat{X}_{k-1} + \frac{E[(X - \hat{X}_{k-1})^2]}{E[(X - \hat{X}_{k-1})^2] + m^2}(Z_k - \hat{X}_{k-1}). \quad (6.16)$$

If we introduce

$$\bar{Z}_k = \frac{1}{k} \sum_{j=1}^k Z_j,$$

then this can be simplified to

$$\hat{X}_k = \frac{a^2}{a^2 + \frac{1}{k} \cdot m^2} \bar{Z}_k. \quad (6.17)$$

(This can be seen as follows:

Put

$$\alpha_k = \frac{a^2}{a^2 + \frac{1}{k} m^2}, U_k = \alpha_k \bar{Z}_k.$$

Then

- (i) $U_k \in \mathcal{L}(Z, k)$
- (ii) $X - U_k \perp \mathcal{L}(Z, k)$, since

$$\begin{aligned} E[(X - U_k)Z_i] &= E[XZ_i] - \alpha_k E[\bar{Z}_k Z_i] \\ &= E[X(X + W_i)] - \alpha_k \frac{1}{k} \sum_j E[Z_j Z_i] \\ &= a^2 - \frac{1}{k} \alpha_k \sum_j E[(X + W_j)(X + W_i)] = a^2 - \frac{1}{k} \alpha_k [ka^2 + m^2] = 0. \end{aligned}$$

The result can be interpreted as follows:

For large k we put $\hat{X}_k \approx \bar{Z}_k$, while for small k the relation between a^2 and m^2 becomes more important. If $m^2 \gg a^2$, the observations are to a large extent neglected (for small k) and \hat{X}_k is put equal to its mean value, 0.

This example gives the motivation for our approach:

We replace the process Z_t by an orthogonal increment process N_t (Step 2) in order to obtain a representation for \hat{X}_t analogous to (6.14). Such a representation is obtained in Step 3, after we have identified the best *linear* estimate with the best *measurable* estimate (Step 1) and established the connection between N_t and Brownian motion.

STEP 1. Z-LINEAR AND Z-MEASURABLE ESTIMATES

LEMMA 6.4. Let $X, Z_s; s \leq t$ be random variables in $L^2(P)$ and assume that

$$(X, Z_{s_1}, Z_{s_2}, \dots, Z_{s_n}) \in \mathbf{R}^{n+1}$$

has a normal distribution for all $s_1, s_2, \dots, s_n \leq t, n \geq 1$. Then

$$\mathcal{P}_{\mathcal{L}}(X) = E[X|\mathcal{G}] = \mathcal{P}_{\mathcal{K}}(X).$$

In other words, the best *Z-linear* estimate for X coincides with the best *Z-measurable* estimate in this case.

Proof. Put $\check{X} = \mathcal{P}_{\mathcal{L}}(X)$, $\tilde{X} = X - \check{X}$. Then we claim that \tilde{X} is independent of \mathcal{G} : Recall that a random variable $(Y_1, \dots, Y_k) \in \mathbf{R}^k$ is normal iff $c_1 Y_1 + \dots + c_k Y_k$ is normal, for all choices of $c_1, \dots, c_k \in \mathbf{R}$. And an L^2 -limit of normal variables is again normal (Appendix A). Therefore

$$(\tilde{X}, Z_{s_1}, \dots, Z_{s_n}) \text{ is normal for all } s_1, \dots, s_n \leq t.$$

Since $E[\tilde{X} Z_{s_j}] = 0$, \tilde{X} and Z_{s_j} are uncorrelated, for $1 \leq j \leq n$. It follows (Appendix A) that

$$\tilde{X} \text{ and } (Z_{s_1}, \dots, Z_{s_n}) \text{ are independent.}$$

So \tilde{X} is independent from \mathcal{G} as claimed. But then

$$E[\mathcal{X}_G(X - \tilde{X})] = E[\mathcal{X}_G \tilde{X}] = E[\mathcal{X}_G] \cdot E[\tilde{X}] = 0 \text{ for all } G \in \mathcal{G}$$

i.e. $\int_G X dP = \int_G \tilde{X} dP$. Since \tilde{X} is \mathcal{G} -measurable, we conclude that $\tilde{X} = E[X|\mathcal{G}]$.

There is a curious interpretation of this result: Suppose $X, \{Z_t\}_{t \in T}$ are $L^2(P)$ -functions with given covariances. Among all possible distributions of

$$(X, Z_{t_1}, \dots, Z_{t_n})$$

with these covariances, the *normal* distribution will be the “worst” w.r.t. estimation, in the following sense: For any distribution we have

$$E[(X - E[X|\mathcal{G}])^2] \leq E[(X - \mathcal{P}_{\mathcal{L}}(X))^2],$$

with equality for the normal distribution, by Lemma 6.4. (Note that the quantity on the right hand side only depends on the covariances, not on the distribution we might choose to obtain these covariances). For a broad discussion of similar conclusions, based on an information theoretical game between nature and the observer, see Tops  e (1978).

Finally, to be able to apply Lemma 6.4 to our filtering problem, we need the following result:

LEMMA 6.5.

$$M_t = \begin{bmatrix} X_t \\ Z_t \end{bmatrix} \in \mathbf{R}^2 \text{ is a Gaussian process.}$$

Proof. We may regard M_t as the solution of a 2-dimensional linear stochastic differential equation of the form

$$dM_t = H(t)M_t dt + K(t)dB_t, M_0 = \begin{bmatrix} X_0 \\ 0 \end{bmatrix}; \quad (6.18)$$

where $H(t) \in \mathbf{R}^{2 \times 2}$, $K(t) \in \mathbf{R}^{2 \times 2}$ and B_t is 2-dimensional Brownian motion.

Use Picard iteration to solve (6.18), i.e. put

$$M_t^{(n+1)} = M_0 + \int_0^t H(s)M_s^{(n)} ds + \int_0^t K(s)dB_s, n = 0, 1, 2, \dots \quad (6.19)$$

Then $M_t^{(n)}$ is Gaussian for all n and $M_t^{(n)} \rightarrow M_t$ in $L^2(P)$ (see the proof of Theorem 5.5) and therefore M_t is Gaussian (Theorem A.7).

STEP 2. THE INNOVATION PROCESS

Before we introduce the innovation process we will establish a useful representation of the functions in the space

$$\begin{aligned} \mathcal{L}(Z, T) &= \text{the closure in } L^2(P) \text{ of all linear combinations} \\ &c_0 + c_1 Z_{t_1} + \dots + c_k Z_{t_k}; 0 \leq t_i \leq T, c_j \in \mathbf{R}. \end{aligned}$$

If $f \in L^2[0, T]$, note that

$$\begin{aligned} E\left[\left(\int_0^T f(t)dZ_t\right)^2\right] &= E\left[\left(\int_0^T f(t)G(t)X_t dt\right)^2\right] + E\left[\left(\int_0^T f(t)D(t)dV_t\right)^2\right] \\ &+ 2E\left[\left(\int_0^T f(t)G(t)X_t dt\right)\left(\int_0^T f(t)D(t)dV_t\right)\right]. \end{aligned}$$

Since

$$\begin{aligned} E\left[\left(\int_0^T f(t)G(t)X_t dt\right)^2\right] &\leq A_1 \cdot \int_0^T f(t)^2 dt \quad \text{by the Cauchy-Schwartz inequality,} \\ E\left[\left(\int_0^T f(t)D(t)dV_t\right)^2\right] &= \int_0^T f(t)^2 D^2(t)dt \quad \text{by the Ito isometry} \end{aligned}$$

and $\{X_t\}, \{V_t\}$ are independent, we conclude that

$$A_0 \int_0^T f^2(t)dt \leq E\left[\left(\int_0^T f(t)dZ_t\right)^2\right] \leq A_1 \int_0^T f^2(t)dt, \quad (6.20)$$

for some constants A_0, A_1 not depending on f . We can now show

LEMMA 6.6. $\mathcal{L}(Z, T) = \{c_0 + \int_0^T f(t)dZ_t; f \in L^2[0, T], c_0 \in \mathbf{R}\}$.

Proof. Denote the right hand side by $\mathcal{N}(Z, T)$. It is enough to show that

- a) $\mathcal{N}(Z, T) \subset \mathcal{L}(Z, T)$
- b) $\mathcal{N}(Z, T)$ contains all linear combinations of the form

$$c_0 + c_1 Z_{t_1} + \cdots + c_k Z_{t_k}; 0 \leq t_i \leq T$$

- c) $\mathcal{N}(Z, T)$ is closed in $L^2(P)$

a): This follows from the fact that if f is continuous then

$$\int_0^T f(t)dZ_t = \lim_{n \rightarrow \infty} \sum_j f(j \cdot 2^{-n}) \cdot (Z_{(j+1)2^{-n}} - Z_{j \cdot 2^{-n}}).$$

b): Suppose $0 \leq t_1 < t_2 < \cdots < t_k \leq T$. We can write

$$\sum_{i=1}^k c_i Z_{t_i} = \sum_{j=0}^{k-1} c'_j \Delta Z_j = \sum_{j=0}^{k-1} \int_{t_j}^{t_{j+1}} c'_j dZ_t = \int_0^T \left(\sum_{j=0}^{k-1} c'_j \mathcal{X}_{[t_j, t_{j+1})}(t) \right) dZ_t,$$

where $\Delta Z_j = Z_{t_{j+1}} - Z_{t_j}$. c): This follows from (6.20) and the fact that $L^2[0, T]$ is complete.

Now we define the *innovation process* N_t as follows:

$$N_t = Z_t - \int_0^t (GX)_s^\wedge ds, \text{ where } (GX)_s^\wedge = \mathcal{P}_{\mathcal{L}(Z,s)}(G(s)X_s) = G(s)\hat{X}_s. \quad (6.21)$$

i.e.

$$dN_t = G(t)(X_t - \hat{X}_t)dt + D(t)dV_t. \quad (6.22)$$

LEMMA 6.7.

- (i) N_t has orthogonal increments
- (ii) $E[N_t^2] = \int_0^t D^2(s)ds$
- (iii) $\mathcal{L}(N, t) = \mathcal{L}(Z, t)$ for all $t \geq 0$
- (iv) N_t is a Gaussian process

Proof.

(i): If $s < t$ and $Y \in \mathcal{L}(Z, s)$ we have

$$\begin{aligned} E[(N_t - N_s)Y] &= E\left[\left(\int_s^t G(r)(X_r - \hat{X}_r)dr + \int_s^t D(r)dV_r\right)Y\right] \\ &= \int_s^t G(r)E[(X_r - \hat{X}_r)Y]dr + E\left[\left(\int_s^t D(r)dV_r\right)Y\right] = 0, \end{aligned}$$

since $X_r - \hat{X}_r \perp \mathcal{L}(Z, r) \supset \mathcal{L}(Z, s)$ for $r \geq s$ and V has independent increments.

(ii): By Ito's formula, with $g(t, x) = x^2$, we have

$$d(N_t^2) = 2N_t dN_t + \frac{1}{2}2(dN_t)^2 = 2N_t dN_t + D^2 dt.$$

So

$$E[N_t^2] = E\left[\int_0^t 2N_s dN_s\right] + \int_0^t D^2(s)ds.$$

Now

$$\int_0^t N_s dN_s = \lim_{\Delta t_j \rightarrow 0} \sum N_{t_j} [N_{t_{j+1}} - N_{t_j}],$$

so since N has orthogonal increments we have

$$E\left[\int_0^t N_s dN_s\right] = 0, \text{ and (ii) follows.}$$

(iii): It is clear that $\mathcal{L}(N, t) \subset \mathcal{L}(Z, t)$ for all $t \geq 0$. To establish the opposite inclusion we use Lemma 6.6. So choose $f \in L^2[0, t]$ and let us see what functions can be obtained in the form

$$\begin{aligned} \int_0^t f(s)dN_s &= \int_0^t f(s)dZ_s - \int_0^t f(r)G(r)\hat{X}_r dr \\ &= \int_0^t f(s)dZ_s - \int_0^t f(r)\left[\int_0^r g(r, s)dZ_s\right]dr - \int_0^t f(r)c(r)dr \\ &= \int_0^t [f(s) - \int_s^t f(r)g(r, s)dr]dZ_s - \int_0^t f(r)c(r)dr, \end{aligned}$$

where we have used Lemmas 6.4 and 6.6 to write, for each r ,

$$(GX)_r^\wedge = c(r) + \int_0^r g(r, s)dZ_s \text{ for some } g(r, \cdot) \in L^2[0, r], c(r) \in \mathbf{R}.$$

From the theory of Volterra integral equations (see e.g. Davis (1977), p. 125) there exists for all $h \in L^2[0, t]$ an $f \in L^2[0, t]$ such that

$$f(s) - \int_s^t f(r)g(r, s)dr = h(s).$$

So by choosing $h = \mathcal{X}_{[0, t_1]}$ where $0 \leq t_1 \leq t$, we obtain

$$\int_0^t f(r)c(r)dr + \int_0^t f(s)dN_s = \int_0^t \mathcal{X}_{[0, t_1]}(s)dZ_s = Z_{t_1},$$

which shows that $\mathcal{L}(N, t) \supset \mathcal{L}(Z, t)$.

(iv): \hat{X}_t is a limit (in $L^2(P)$) of linear combinations of the form

$$M = c_0 + c_1 Z_{s_1} + \cdots + c_k Z_{s_k}, \text{ where } s_k \leq t.$$

Therefore

$$(\hat{X}_{t_1}, \dots, \hat{X}_{t_m})$$

is a limit of m -dimensional random variables $(M^{(1)}, \dots, M^{(m)})$ whose components $M^{(j)}$ are linear combinations of this form. $(M^{(1)}, \dots, M^{(m)})$ has a normal distribution since $\{Z_t\}$ is Gaussian, and therefore the limit has. Hence $\{\hat{X}_t\}$ is Gaussian. It follows that

$$N_t = Z_t - \int_0^t G(s) \hat{X}_s ds$$

is Gaussian, by a similar argument.

STEP 3. THE INNOVATION PROCESS AND BROWNIAN MOTION

Let $N_t = Z_t - \int_0^t G(s) \hat{X}_s ds$ be the innovation process defined in Step 2. Recall that we have assumed that $D(t)$ is bounded away from 0 on bounded intervals. Define the process $R_t(\omega)$ by

$$dR_t = \frac{1}{D(t)} dN_t(\omega); \quad t \geq 0, \quad R_0 = 0. \quad (6.23)$$

LEMMA 6.8. R_t is a 1-dimensional Brownian motion.

Proof. Observe that

- (i) R_t has continuous paths
- (ii) R_t has orthogonal increments (since N_t has)
- (iii) R_t is Gaussian (since N_t is)
- (iv) $E[R_t] = 0$ and $E[R_t R_s] = \min(s, t)$

To prove the last assertion in (iv), note that by Ito's formula

$$d(R_t^2) = 2R_t dR_t + (dR_t)^2 = 2R_t dR_t + dt,$$

so, since R_t has orthogonal increments,

$$E[R_t^2] = E\left[\int_0^t ds\right] = t.$$

Therefore, if $s < t$,

$$E[R_t R_s] = E[(R_t - R_s) R_s] + E[R_s^2] = E[R_s^2] = s.$$

Properties (i), (iii) and (iv) constitute one of the many characterizations of a 1-dimensional Brownian motion (see Williams (1979), I. 15). (Alternatively, we could easily deduce that R_t has stationary, independent increments and therefore - by continuity - must be Brownian motion, by the result previously referred to in the beginning of Chapter III). For a general characterization of Brownian motion see Corollary 8.13.

Since

$$\mathcal{L}(N, t) = \mathcal{L}(R, t)$$

we conclude that

$$\hat{X}_t = \mathcal{P}_{\mathcal{L}(R,t)}(X_t).$$

It turns out that the projection down to the space $\mathcal{L}(R, t)$ can be described very nicely: (compare with formula (6.14) in Example 6.3)

LEMMA 6.9.

$$\hat{X}_t = E[X_t] + \int_0^t \frac{\partial}{\partial s} E[X_t R_s] dR_s. \quad (6.24)$$

Proof. From Lemma 6.6 we know that

$$\hat{X}_t = c_0(t) + \int_0^t g(s) dR_s \text{ for some } g \in L^2[0, t], c_0(t) \in \mathbf{R}.$$

Taking expectations we see that $c_0(t) = E[\hat{X}_t] = E[X_t]$. We have

$$(X_t - \hat{X}_t) \perp \int_0^t f(s) dR_s \text{ for all } f \in L^2[0, t].$$

Therefore

$$\begin{aligned} E[X_t \int_0^t f(s) dR_s] &= E[\hat{X}_t \int_0^t f(s) dR_s] = E[\int_0^t g(s) dR_s \int_0^t f(s) dR_s] \\ &= E[\int_0^t g(s) f(s) ds] = \int_0^t g(s) f(s) ds, \text{ for all } f \in L^2[0, 1], \end{aligned}$$

where we have used the Ito isometry. In particular, if we choose $f = \chi_{[0, r]}$ for some $r \leq t$, we obtain

$$E[X_t R_r] = \int_0^r g(s) ds$$

or

$$g(r) = \frac{\partial}{\partial r} E[X_t R_r], \text{ as asserted.}$$

This completes Step 3.

STEP 4. AN EXPLICIT FORMULA FOR X_t

This is easily obtained using Ito's formula, as in the examples in Chapter V. The result is

$$\begin{aligned} X_t &= \exp\left(\int_0^t F(s)ds\right)[X_0 + \int_0^t \exp\left(-\int_0^s F(u)du\right)C(s)dU_s] \\ &= \exp\left(\int_0^t F(s)ds\right)X_0 + \int_0^t \exp\left(\int_s^t F(u)du\right)C(s)dU_s. \end{aligned}$$

In particular, we note that $E[X_t] = E[X_0] \exp\left(\int_0^t F(s)ds\right)$. More generally, if $0 \leq r \leq t$,

$$X_t = \exp\left(\int_r^t F(s)ds\right)X_r + \int_r^t \exp\left(\int_s^t F(u)du\right)C(s)dU_s. \quad (6.25)$$

STEP 5.

THE STOCHASTIC DIFFERENTIAL EQUATION FOR \hat{X}_t

We now combine the previous steps to obtain the solution of the filtering problem, i.e. a stochastic differential equation for \hat{X}_t . Starting with the formula from Lemma 6.9 (6.24)

$$\hat{X}_t = E[X_t] + \int_0^t f(s, t)dR_s,$$

where

$$f(s, t) = \frac{\partial}{\partial s}E[X_t R_s], \quad (6.26)$$

we use that

$$R_s = \int_0^s \frac{G(r)}{D(r)}(X_r - \hat{X}_r)dr + V_s \text{ from (6.22) and (6.23)}$$

and obtain

$$E[X_t R_s] = \int_0^s \frac{G(r)}{D(r)}E[X_t \tilde{X}_r]dr,$$

where

$$\tilde{X}_r = X_r - \hat{X}_r. \quad (6.27)$$

Using formula (6.25) for X_t , we obtain

$$E[X_t \tilde{X}_r] = \exp\left(\int_r^t F(v)dv\right)E[X_r \tilde{X}_r] = \exp\left(\int_r^t F(v)dv\right)S(r),$$

where

$$S(r) = E[(\hat{X}_r)^2], \quad (6.28)$$

i.e. the mean square error of the estimate at time $r \geq 0$. Thus

$$E[X_t R_s] = \int_0^s \frac{G(r)}{D(r)} \exp\left(\int_r^t F(v)dv\right) S(r) dr$$

so that

$$f(s, t) = \frac{G(s)}{D(s)} \exp\left(\int_s^t F(v)dv\right) S(s). \quad (6.29)$$

We claim that $S(t)$ satisfies the (deterministic) differential equation

$$\frac{dS}{dt} = 2F(t)S(t) - \frac{G^2(t)}{D^2(t)}S^2(t) + C^2(t) \quad (\text{The Riccati equation}). \quad (6.30)$$

To prove (6.30) note that by the Pythagorean theorem, (6.24) and the Ito isometry

$$\begin{aligned} S(t) &= E[(X_t - \hat{X}_t)^2] = E[X_t^2] - 2E[X_t \hat{X}_t] + E[\hat{X}_t^2] = E[X_t^2] - E[\hat{X}_t^2] \\ &= T(t) - \int_0^t f(s, t)^2 ds - E[X_t]^2, \end{aligned} \quad (6.31)$$

where

$$T(t) = E[X_t^2]. \quad (6.32)$$

Now by (6.25) and the Ito isometry we have

$$T(t) = \exp\left(2 \int_0^t F(s)ds\right) E[X_0^2] + \int_0^t \exp\left(2 \int_s^t F(u)du\right) C^2(s) ds,$$

using that X_0 is independent of $\{U_s\}_{s \geq 0}$. So

$$\frac{dT}{dt} = 2F(t) \cdot \exp\left(2 \int_0^t F(s)ds\right) E[X_0^2] + C^2(t) + \int_0^t 2F(t) \exp\left(2 \int_s^t F(u)du\right) C^2(s) ds$$

i.e.

$$\frac{dT}{dt} = 2F(t)T(t) + C^2(t). \quad (6.33)$$

Substituting in (6.31) we obtain, using Step 4,

$$\begin{aligned}
\frac{dS}{dt} &= \frac{dT}{dt} - f(t, t)^2 - \int_0^t 2f(s, t) \cdot \frac{\partial}{\partial t} f(s, t) ds - 2F(t)E[X_t]^2 \\
&= 2F(t)T(t) + C^2(t) - \frac{G^2(t)S^2(t)}{D^2(t)} - 2 \int_0^t f^2(s, t)F(t)ds - 2F(t)E[X_t]^2 \\
&= 2F(t)S(t) + C^2(t) - \frac{G^2(t)S^2(t)}{D^2(t)}, \text{ which is (6.30).}
\end{aligned}$$

We are now ready for the stochastic differential equation for \hat{X}_t :

From the formula

$$\hat{X}_t = c_0(t) + \int_0^t f(s, t)dR_s \quad \text{where } c_0(t) = E[X_t]$$

it follows that

$$d\hat{X}_t = c'_0(t)dt + f(t, t)dR_t + \left(\int_0^t \frac{\partial}{\partial t} f(s, t)dR_s \right) dt, \quad (6.34)$$

since

$$\begin{aligned}
\int_0^u \left(\int_0^t \frac{\partial}{\partial t} f(s, t)dR_s \right) dt &= \int_0^u \left(\int_s^u \frac{\partial}{\partial t} f(s, t)dt \right) dR_s \\
&= \int_0^u (f(s, u) - f(s, s))dR_s = \hat{X}_u - c_0(u) - \int_0^u f(s, s)dR_s.
\end{aligned}$$

So

$$d\hat{X}_t = c'_0(t)dt + \frac{G(t)S(t)}{D(t)}dR_t + \left(\int_0^t f(s, t)dR_s \right) F(t)dt$$

or

$$\begin{aligned}
d\hat{X}_t &= c'_0(t)dt + F(t) \cdot (\hat{X}_t - c_0(t))dt + \frac{G(t)S(t)}{D(t)}dR_t \\
&= F(t)\hat{X}_t dt + \frac{G(t)S(t)}{D(t)}dR_t,
\end{aligned} \quad (6.35)$$

since $c'_0(t) = F(t)c_0(t)$ (Step 4).

If we substitute

$$dR_t = \frac{1}{D(t)}[dZ_t - G(t)\hat{X}_t dt]$$

we obtain

$$d\hat{X}_t = \left(F(t) - \frac{G^2(t)S(t)}{D^2(t)} \right) \hat{X}_t dt + \frac{G(t)S(t)}{D^2(t)} dZ_t. \quad (6.36)$$

So the conclusion is:

THEOREM 6.10.

(THE 1-DIMENSIONAL KALMAN – BUCY FILTER)

The solution $\hat{X}_t = E[X_t | \mathcal{G}_t]$ of the 1-dimensional linear filtering problem

$$\text{(linear system)} \quad dX_t = F(t)X_t dt + C(t)dU_t; \quad F(t), C(t) \in \mathbf{R} \quad (6.12)$$

$$\text{(linear observations)} \quad dZ_t = G(t)X_t dt + D(t)dV_t; \quad G(t), D(t) \in \mathbf{R} \quad (6.13)$$

(with conditions as stated earlier) satisfies the stochastic differential equation

$$d\hat{X}_t = \left(F(t) - \frac{G^2(t)S(t)}{D^2(t)} \right) \hat{X}_t dt + \frac{G(t)S(t)}{D^2(t)} dZ_t; \quad \hat{X}_0 = E[X_0] \quad (6.37)$$

where

$S(t) = E[(X_t - \hat{X}_t)^2]$ satisfies the (deterministic) Riccati equation

$$\frac{dS}{dt} = 2F(t)S(t) - \frac{G^2(t)}{D^2(t)} S^2(t) + C^2(t), \quad S(0) = E[(X_0 - E[X_0])^2]. \quad (6.38)$$

EXAMPLE 6.11. (Noisy observations of a constant process)

Consider the simple case

$$\begin{aligned} \text{(system)} \quad dX_t &= 0, \text{ i.e. } X_t = X_0; \quad E[X_0] = 0, E[X_0^2] = a^2 \\ \text{(observations)} \quad dZ_t &= X_t dt + m dV_t; \quad Z_0 = 0 \end{aligned}$$

(corresponding to

$$H_t = \frac{dZ}{dt} = X_t + mW_t, \quad W_t = \text{white noise}).$$

First we solve the corresponding Riccati equation for

$$\begin{aligned} S(t) &= E[(X_t - \hat{X}_t)^2] : \\ \frac{dS}{dt} &= -\frac{1}{m^2} S^2, \quad S(0) = a^2 \end{aligned}$$

i.e.

$$S(t) = \frac{a^2 m^2}{m^2 + a^2 t}; \quad t \geq 0.$$

This gives the following equation for \hat{X}_t :

$$d\hat{X}_t = -\frac{a^2}{m^2 + a^2 t} \hat{X}_t dt + \frac{a^2}{m^2 + a^2 t} dZ_t; \hat{X}_0 = E[X_0] = 0$$

or

$$d(\hat{X}_t \exp(\int \frac{a^2}{m^2 + a^2 t} dt)) = \exp(\int \frac{a^2}{m^2 + a^2 t} dt) \frac{a^2}{m^2 + a^2 t} dZ_t$$

which gives

$$\hat{X}_t = \frac{a^2}{m^2 + a^2 t} Z_t; \quad t \geq 0. \quad (6.39)$$

This is the continuous analogue of Example 6.3.

EXAMPLE 6.12. (Noisy observations of a Brownian motion)

If we modify the preceding example slightly, so that

$$\begin{array}{ll} \text{(system)} & dX_t = c dU_t; \quad E[X_0] = 0, E[X_0^2] = a^2, \quad c \text{ constant} \\ \text{(observations)} & dZ_t = X_t dt + m dV_t, \end{array}$$

the Riccati equation becomes

$$\frac{dS}{dt} = -\frac{1}{m^2} S^2 + c^2, \quad S(0) = a^2$$

or

$$\frac{m^2 dS}{m^2 c^2 - S^2} = dt, \quad (S \neq mc).$$

This gives

$$\left| \frac{mc + s}{mc - s} \right| = K \exp\left(\frac{2ct}{m}\right); \quad K = \left| \frac{mc + a^2}{mc - a^2} \right|.$$

Or

$$S(t) = \begin{cases} mc \frac{K \cdot \exp(\frac{2ct}{m}) - 1}{K \cdot \exp(\frac{2ct}{m}) + 1}; & \text{if } S(0) < mc \\ mc \text{ (constant)} & \text{if } S(0) = mc \\ mc \frac{K \cdot \exp(\frac{2ct}{m}) + 1}{K \cdot \exp(\frac{2ct}{m}) - 1} & \text{if } S(0) > mc \end{cases}$$

Thus in all cases the mean square error $S(t)$ tends to mc as $t \rightarrow \infty$.

For simplicity let us put $a = 0, m = c = 1$. Then

$$S(t) = \frac{\exp(2t) - 1}{\exp(2t) + 1} = \tanh(t).$$

The equation for \hat{X}_t is

$$d\hat{X}_t = -\tanh(t) \hat{X}_t dt + \tanh(t) dZ_t, \quad \hat{X}_0 = 0$$

or

$$d(\cosh(t)\hat{X}_t) = \sinh(t)dZ_t.$$

So

$$\hat{X}_t = \frac{1}{\cosh(t)} \int_0^t \sinh(s)dZ_s.$$

If we return to the interpretation of Z_t :

$$Z_t = \int_0^t H_s ds,$$

where H_s are the “original” observations (see (6.3)), we can write

$$\hat{X}_t = \frac{1}{\cosh(t)} \int_0^t \sinh(s)H_s ds, \quad (6.40)$$

so \hat{X}_t is approximately (for large t) a weighted average of the observations H_s , with increasing emphasis on observations as time increases.

REMARK. It is interesting to compare formula (6.40) with established formulas in forecasting. For example, the *exponentially weighted moving average* \tilde{X}_n suggested by C.C. Holt in 1958 is given by

$$\tilde{X}_n = (1 - \alpha)^n Z_0 + \alpha \sum_{k=1}^n (1 - \alpha)^{n-k} Z_k,$$

where α is some constant; $0 \leq \alpha \leq 1$. See The Open University (1981), p. 16.

This may be written

$$\tilde{X}_n = \beta^{-n} Z_0 + (\beta - 1)\beta^{-n} \sum_{k=1}^n \beta^k Z_k,$$

where $\beta = \frac{1}{1-\alpha}$ (assuming $\alpha < 1$), which is a discrete version of (6.40), or - more precisely - of the formula corresponding to (6.40) in the general case when $a \neq 0$ and m, c are not necessarily equal to 1.

EXAMPLE 6.13. (Estimation of a parameter)

Suppose we want to estimate the value of a (constant) parameter θ , based on observations Z_t satisfying the model

$$dZ_t = \theta M(t)dt + N(t)dB_t,$$

where $M(t), N(t)$ are known functions. In this case the stochastic differential equation for θ is of course

$$d\theta = 0,$$

so the Riccati equation for $S(t) = E[(\theta - \hat{\theta}_t)^2]$ is

$$\frac{dS}{dt} = -\left(\frac{M(t)S(t)}{N(t)}\right)^2$$

which gives

$$S(t) = (S_0^{-1} + \int_0^t M(s)^2 N(s)^{-2} ds)^{-1}$$

and the Kalman-Bucy filter is

$$d\hat{\theta}_t = \frac{M(t)S(t)}{N(t)^2}(dZ_t - M(t)\hat{\theta}_t dt).$$

This can be written

$$(S_0^{-1} + \int_0^t M(s)^2 N(s)^{-2} ds)d\hat{\theta}_t + M(t)^2 N(t)^{-2}\hat{\theta}_t dt = M(t)N(t)^{-2}dZ_t.$$

We recognize the left hand side as

$$d((S_0^{-1} + \int_0^t M(s)^2 N(s)^{-2} ds)\hat{\theta}_t)$$

so we obtain

$$\hat{\theta}_t = \frac{\hat{\theta}_0 S_0^{-1} + \int_0^t M(s)N(s)^{-2} dZ_s}{S_0^{-1} + \int_0^t M(s)^2 N(s)^{-2} ds}.$$

This estimate coincides with the maximum likelihood estimate in classical estimation theory if $S_0^{-1} = 0$. See Liptser and Shirayev (1978).

For more information about estimates of drift parameters in diffusions and generalizations, see for example Aase (1982), Brown and Hewitt (1975) and Taraskin (1974).

EXAMPLE 6.14. (Noisy observations of a population growth)

Consider a simple growth model (r constant)

$$dX_t = rX_t dt, E[X_0] = b > 0, E[(X_0 - b)^2] = a^2,$$

with observations

$$dZ_t = X_t dt + m dV_t; m \text{ constant.}$$

The corresponding Riccati equation

$$\frac{dS}{dt} = 2rS - \frac{1}{m^2} S^2, S(0) = a^2,$$

gives the logistic curve

$$S(t) = \frac{2rm^2}{1 + Ke^{-2rt}}; \text{ where } K = \frac{2rm^2}{a^2} - 1.$$

So the equation for \hat{X}_t becomes

$$d\hat{X}_t = (r - \frac{S}{m^2})\hat{X}_t dt + \frac{S}{m^2} dZ_t; \hat{X}_0 = E[X_0] = b.$$

For simplicity let us assume that $a^2 = 2rm^2$, so that

$$S(t) = 2rm^2 \text{ for all } t.$$

(In the general case $S(t) \rightarrow 2rm^2$ as $t \rightarrow \infty$, so this is not an unreasonable approximation for large t). Then we get

$$d(\exp(rt)\hat{X}_t) = \exp(rt)2rdZ_t, \hat{X}_0 = b$$

or

$$\hat{X}_t = \exp(-rt) \left[\int_0^t 2r \exp(rs) dZ_s + b \right].$$

As in Example 6.12 this may be written

$$\hat{X}_t = \exp(-rt) \left[\int_0^t 2r \exp(rs) H_s ds + b \right], \text{ if } Z_t = \int_0^t H_s ds. \quad (6.41)$$

For example, assume that $H_s = \beta$ (constant) for $0 \leq s \leq t$, i.e. that our observations (for some reason) give the same value β for all times $s \leq t$. Then

$$\hat{X}_t = 2\beta - (2\beta - b) \exp(-rt) \rightarrow 2\beta \text{ as } t \rightarrow \infty.$$

If $H_s = \beta \cdot \exp(\alpha s)$, $s \geq 0$ (α constant), we get

$$\begin{aligned} \hat{X}_t &= \exp(-rt) \left[\frac{2r\beta}{r + \alpha} (\exp(r + \alpha)t - 1) + b \right] \\ &\approx \frac{2r\beta}{r + \alpha} \exp(\alpha t) \text{ for large } t. \end{aligned}$$

Thus, only if $\alpha = r$, i.e. $H_s = \beta \exp(rs)$; $s \geq 0$, does the filter “believe” the observations in the long run. And only if $\alpha = r$ and $\beta = b$, i.e. $H_s = b \exp(rs)$; $s \geq 0$, does the filter “believe” the observations at all times.

EXAMPLE 6.15. (Constant coefficients – general discussion)

Now consider the system

$$dX_t = FX_t dt + CdU_t; \quad F, C \text{ constants, } \neq 0$$

with observations

$$dZ_t = GX_t dt + DdV_t; \quad G, D \text{ constants, } \neq 0.$$

The corresponding Riccati equation

$$S' = 2FS - \frac{G^2}{D^2}S^2 + C^2, S(0) = a^2$$

has the solution

$$S(t) = \frac{\alpha_1 - K\alpha_2 \exp(\frac{(\alpha_2 - \alpha_1)G^2 t}{D^2})}{1 - K \exp(\frac{(\alpha_2 - \alpha_1)G^2 t}{D^2})},$$

where

$$\begin{aligned}\alpha_1 &= G^{-2}(FD^2 - D\sqrt{F^2D^2 + G^2C^2}) \\ \alpha_2 &= G^{-2}(FD^2 + D\sqrt{F^2D^2 + G^2C^2})\end{aligned}$$

and

$$K = \frac{a^2 - \alpha_1}{a^2 - \alpha_2}.$$

This gives the solution for \hat{X}_t of the form

$$\hat{X}_t = \exp\left(\int_0^t H(s)ds\right)\hat{X}_0 + \frac{G}{D^2} \int_0^t \exp\left(\int_s^t H(u)du\right)S(s)dZ_s,$$

where

$$H(s) = F - \frac{G^2}{D^2}S(s).$$

For large s we have $S(s) \approx \alpha_2$. This gives

$$\begin{aligned}\hat{X}_t &\approx \hat{X}_0 \exp\left((F - \frac{G^2\alpha_2}{D^2})t\right) + \frac{G\alpha_2}{D^2} \int_0^t \exp\left((F - \frac{G^2\alpha_2}{D^2})(t-s)\right)dZ_s \\ &= \hat{X}_0 \exp(-\beta t) + \frac{G\alpha_2}{D^2} \exp(-\beta t) \int_0^t \exp(\beta s)dZ_s\end{aligned}\tag{6.42}$$

where $\beta = D^{-1}\sqrt{F^2D^2 + G^2C^2}$. So we get approximately the same behaviour as in the previous example.

Finally we formulate the solution of the n-dimensional linear filtering problem (6.10), (6.11):

THEOREM 6.16.

(THE MULTI-DIMENSIONAL KALMAN-BUCY FILTER)

The solution $\hat{X}_t = E[X_t | \mathcal{G}_t]$ of the multi-dimensional linear filtering problem

$$\text{(linear system)} \quad dX_t = F(t)X_t dt + C(t)dU_t; \quad F(t) \in \mathbf{R}^{n \times n}, \quad C(t) \in \mathbf{R}^{n \times p} \quad (6.10)$$

$$\text{(linear observations)} \quad dZ_t = G(t)X_t dt + D(t)dV_t; \quad G(t) \in \mathbf{R}^{m \times n}, \quad D(t) \in \mathbf{R}^{m \times r} \quad (6.11)$$

satisfies the stochastic differential equation

$$d\hat{X}_t = (F - SG^T(DD^T)^{-1}G)\hat{X}_t dt + SG^T(DD^T)^{-1}dZ_t; \quad \hat{X}_0 = E[X_0] \quad (6.43)$$

where $S(t) := E[(X_t - \hat{X}_t)(X_t - \hat{X}_t)^T] \in \mathbf{R}^{n \times n}$ satisfies the matrix Riccati equation

$$\begin{aligned} \frac{dS}{dt} &= FS + SF^T - SG^T(DD^T)^{-1}GS + CC^T; \\ S(0) &= E[(X_0 - E[X_0])(X_0 - E[X_0])^T] \end{aligned} \quad (6.44)$$

The condition on $D(t) \in \mathbf{R}^{m \times r}$ is now that $D(t)D(t)^T$ is invertible for all t and that $(D(t)D(t)^T)^{-1}$ is bounded on every bounded t -interval.

A similar solution can be found for the more general situation

$$\text{(system)} \quad dX_t = [F_0(t) + F_1(t)X_t + F_2(t)Z_t]dt + C(t)dU_t \quad (6.45)$$

$$\text{(observations)} \quad dZ_t = [G_0(t) + G_1(t)X_t + G_2(t)Z_t]dt + D(t)dV_t, \quad (6.46)$$

where $X_t \in \mathbf{R}^n$, $Z_t \in \mathbf{R}^m$ and $B_t = (U_t, V_t)$ is $n + m$ -dimensional Brownian motion, with appropriate dimensions on the matrix coefficients. See Kallianpur (1980), who also treats the non-linear case. An account of non-linear filtering theory is also given in Pardoux (1979) and Davis (1984).

For the solution of linear filtering problems governed by more general processes than Brownian motion (processes with orthogonal increments) see Davis (1977).

For various applications of filtering theory see Bucy and Joseph (1968), Jazwinski (1970), Gelb (1974), Maybeck (1979) and the references in these books.

EXERCISES

6.1. (Time-varying observations of a constant)

Prove that if the (1-dimensional) system is

$$dX_t = 0, E[X_0] = 0, E[X_0^2] = a^2$$

and the observation process is

$$dZ_t = G(t)X_t dt + dV_t, Z_0 = 0$$

then $S(t) = E[(X_t - \hat{X}_t)^2]$ is given by

$$S(t) = \frac{1}{\frac{1}{S(0)} + \int_0^t G^2(s) ds} \quad (6.47)$$

We say that we have *exact asymptotic estimation* if $S(t) \rightarrow 0$ as $t \rightarrow \infty$, i.e. if

$$\int_0^\infty G^2(s) ds = \infty$$

Thus for

$$G(s) = \frac{1}{(1+s)^p} \quad (p > 0 \text{ constant})$$

we have exact asymptotic expansion iff $p \leq \frac{1}{2}$.

6.2.

Consider the linear 1-dimensional filtering problem with no noise in the system:

$$(\text{system}) \quad dX_t = F(t)X_t dt \quad (6.48)$$

$$(\text{observations}) \quad dZ_t = G(t)X_t dt + D(t)dV_t \quad (6.49)$$

Put $S(t) = E[(X_t - \hat{X}_t)^2]$ as usual and assume $S(0) > 0$.

a) Show that

$$R(t) := \frac{1}{S(t)}$$

satisfies the *linear* differential equation

$$R'(t) = -2F(t)R(t) + \frac{G^2(t)}{D^2(t)}; R(0) = \frac{1}{S(0)} \quad (6.50)$$

b) Use (6.50) to prove that for the filtering problem (6.48), (6.49) we have

$$\frac{1}{S(t)} = \frac{1}{S(0)} \exp\left(-2 \int_0^t F(s) ds\right) + \int_0^t \exp\left(-2 \int_s^t F(u) du\right) \frac{G^2(s)}{D^2(s)} ds \quad (6.51)$$

6.3.

In Example 6.14 we found that

$$S(t) \rightarrow 2rm^2 \text{ as } t \rightarrow \infty,$$

so exact asymptotic estimation (Exercise 6.1) of X_t is not possible. However, prove that we can obtain exact asymptotic estimation of X_0 , in the sense that

$$E[(X_0 - E[X_0 | \mathcal{G}_t])^2] \rightarrow 0 \text{ as } t \rightarrow \infty.$$

(Hint: Note that $X_0 = e^{-rt} X_t$ and therefore $E[X_0 | \mathcal{G}_t] = e^{-rt} \hat{X}_t$, so that

$$E[(X_0 - E[X_0 | \mathcal{G}_t])^2] = e^{-2rt} S(t))$$

6.4.

Consider the multi-dimensional linear filtering problem with no noise in the system:

$$(\text{system}) \quad dX_t = F(t)X_t dt; X_t \in \mathbf{R}^n, F(t) \in \mathbf{R}^{n \times n} \quad (6.52)$$

$$(\text{observations}) \quad dZ_t = G(t)X_t dt + D(t)dV_t; G(t) \in \mathbf{R}^{m \times n}, D(t) \in \mathbf{R}^{m \times r} \quad (6.53)$$

Assume that $S(t)$ is nonsingular and define $R(t) = S(t)^{-1}$. Prove that $R(t)$ satisfies the *Lyapunov equation* (compare with Exercise 6.2)

$$R'(t) = -R(t)F(t) - F(t)^T R(t) + G(t)^T (D(t)D(t)^T)^{-1} G(t). \quad (6.54)$$

(Hint: Note that since $S(t)S^{-1}(t) = I$ we have

$S'(t)S^{-1}(t) + S(t)(S^{-1})'(t) = 0$, which gives

$$(S^{-1})'(t) = -S^{-1}(t)S'(t)S^{-1}(t).$$

6.5. (Prediction)

In the *prediction problem* one seeks to estimate the value of the system X at a future time T based on the observations \mathcal{G}_t up to the present time $t < T$.

Prove that in the linear setup (6.12), (6.13) the predicted value

$$E[X_T | \mathcal{G}_t], \quad t > T$$

is given by

$$E[X_T | \mathcal{G}_t] = \exp\left(\int_t^T F(s) ds\right) \cdot \hat{X}_t. \quad (6.55)$$

(Hint: Use formula (6.25))

6.6. (Interpolation/smoothing)

The *interpolation* or *smoothing problem* consists of estimating the value of the system X at a time $s < t$, given the observations \mathcal{G}_r up to time t .

With notation as in (6.10), (6.11) one can show that $M_s := E[X_s | \mathcal{G}_t]$ satisfies the differential equation

$$\begin{cases} \frac{d}{ds} M_s = F(s)M_s + C(s)C^T(s)S^{-1}(s)(M_s - \hat{X}_t); s < t \\ M_t = \hat{X}_t \end{cases} \quad (6.56)$$

(See Davis (1977, Th. 4.4.4).)

Use this result to find $E[X_s | \mathcal{G}_t]$ in Example 6.11.

6.7. (Estimating 2-dimensional Brownian motion)

Consider the system

$$dX_t = \begin{bmatrix} dX_1(t) \\ dX_2(t) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, E[X_0] = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

with observations

$$\begin{bmatrix} dZ_1(t) \\ dZ_2(t) \end{bmatrix} = \begin{bmatrix} X_1 \\ X_1 + X_2 \end{bmatrix} dt + \begin{bmatrix} dV_1(t) \\ dV_2(t) \end{bmatrix}.$$

Apply (6.54) from Exercise 6.4 to prove that $S(t) := E[(X_t - \hat{X}_t)(X_t - \hat{X}_t)^T]$ is given by

$$S(t)^{-1} = S^{-1}(0) + \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} t$$

if $S(0)$ is invertible. Then show that

$$\hat{X}_t = S(t) \int_0^t \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} dZ_s.$$

6.8.

Transform the following Stratonovich equation

$$dX_t = b(t, X_t)dt + \sigma(t, X_t) \circ dB_t$$

into the corresponding Ito equation

$$dX_t = \tilde{b}(t, X_t)dt + \sigma(t, X_t)dB_t$$

using (6.3):

a)

$$\begin{bmatrix} dX_1 \\ dX_2 \end{bmatrix} = \begin{bmatrix} 1 \\ X_2 + e^{2X_1} \end{bmatrix} dt + \begin{bmatrix} 0 \\ e^{X_1} \end{bmatrix} \circ dB_t \quad (B_t \in \mathbf{R})$$

b)

$$\begin{bmatrix} dX_1 \\ dX_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} dt + \begin{bmatrix} X_2 \\ X_1 \end{bmatrix} \circ dB_t \quad (B_t \in \mathbf{R})$$

6.9.

Transform the following Ito equation

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t$$

into the corresponding Stratonovich equation

$$dX_t = \hat{b}(t, X_t)dt + \sigma(t, X_t) \circ dB_t,$$

using (the converse of) (6.3):

a) $dX_t = -\frac{1}{2}X_tdt + KX_tdB_t$, where

$$K = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, X_t = \begin{bmatrix} X_1(t) \\ X_2(t) \end{bmatrix} \in \mathbf{R}^2 \text{ and } B_t \in \mathbf{R}$$

(i.e. X_t is Brownian motion on the unit circle (Example 5.4)).

b) $\begin{bmatrix} dX_1 \\ dX_2 \end{bmatrix} = \begin{bmatrix} X_1 & -X_2 \\ X_2 & X_1 \end{bmatrix} \begin{bmatrix} dB_1 \\ dB_2 \end{bmatrix}.$

6.10. (On the support of an Ito diffusion)The *support* of an Ito diffusion X in \mathbf{R}^n starting at $x \in \mathbf{R}^n$ is the smallest closed set F with the property that

$$X_t(\omega) \in F \text{ for all } t \geq 0, \text{ for a.a.}\omega.$$

In Example 5.4 we found that Brownian motion on the unit circle, X_t , satisfies the (Ito) stochastic differential equation

$$\begin{bmatrix} dX_1(t) \\ dX_2(t) \end{bmatrix} = -\frac{1}{2} \begin{bmatrix} X_1(t) \\ X_2(t) \end{bmatrix} dt + \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} X_1(t) \\ X_2(t) \end{bmatrix} dB_t. \quad (6.57)$$

From this equation it is not at all apparent that its solution is situated on the same circle as the starting point. However, this can be detected by proceeding as follows: First transform the equation into its Stratonovich form, which in Exercise 6.9 is found to be

$$\begin{bmatrix} dX_1(t) \\ dX_2(t) \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} X_1(t) \\ X_2(t) \end{bmatrix} \circ dB_t. \quad (6.58)$$

Then (formally) replace $\circ dB_t$ by $\phi'(t)dt$, where ϕ is some smooth (deterministic) function, $\phi(0) = 0$. This gives the deterministic equation

$$\begin{bmatrix} dX_1^{(\phi)}(t) \\ dX_2^{(\phi)}(t) \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \phi'(t) dt. \quad (6.59)$$

If $(X_1^{(\phi)}(0), X_2^{(\phi)}(0)) = (1, 0)$ the solution of (6.59) is

$$\begin{bmatrix} X_1^{(\phi)}(t) \\ X_2^{(\phi)}(t) \end{bmatrix} = \begin{bmatrix} \cos \phi(t) \\ \sin \phi(t) \end{bmatrix}.$$

So for any smooth ϕ the corresponding solution $X^{(\phi)}(t)$ of (6.59) has its support on this unit circle. We can conclude that the original solution $X(t, \omega)$ is supported on the unit circle also, in virtue of the *Stroock-Varadhan support theorem*. This theorem says that, quite generally, the support of an Ito diffusion $X_t(\omega)$ coincides with the closure (in $C([0, \infty); \mathbf{R}^n)$) of $\{X^{(\phi)}(\cdot); \phi \text{ smooth}\}$, where $X^{(\phi)}(t)$ is obtained by replacing $\circ dB_t$ by $\phi'(t)dt$ in the same way as above. See e.g. Ikeda and Watanabe (1989, Th. VI. 8.1).

(In this special case above the support could also have been found directly from (6.58)).

Use the procedure above to find the support of the process $X_t \in \mathbf{R}^2$ given by

$$dX_t = \frac{1}{2} X_t dt + \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} X_t dB_t.$$

VII. Diffusions: Basic Properties

Suppose we want to describe the motion of a small particle suspended in a moving liquid, subject to random molecular bombardments. If $b(t, x) \in \mathbf{R}^3$ is the velocity of the fluid at the point x at time t , then a reasonable mathematical model for the position X_t of the particle at time t would be a stochastic differential equation of the form

$$\frac{dX_t}{dt} = b(t, X_t) + \sigma(t, X_t)W_t, \quad (7.1)$$

where $W_t \in \mathbf{R}^3$ denotes “white noise” and $\sigma(t, x) \in \mathbf{R}^{3 \times 3}$. The Ito interpretation of this equation is

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t, \quad (7.2)$$

where B_t is 3-dimensional Brownian motion, and similarly (with a correction term added to b) for the Stratonovich interpretation (see (6.3)).

In a stochastic differential equation of the form

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t, \quad (7.3)$$

where $X_t \in \mathbf{R}^n$, $b(t, x) \in \mathbf{R}^n$, $\sigma(t, x) \in \mathbf{R}^{n \times m}$ and B_t is m-dimensional Brownian motion, we will call b the *drift coefficient* and σ - or sometimes $\frac{1}{2}\sigma\sigma^T$ – the *diffusion coefficient* (see Theorem 7.8).

Thus the solution of a stochastic differential equation may be thought of as the mathematical description of the motion of a small particle in a moving fluid: Therefore such stochastic processes are called (*Ito*) *diffusions*.

In this chapter we establish some of the most basic properties and results about Ito diffusions:

- (A) The Markov property.
- (B) The strong Markov property.
- (C) The generator A of X_t expressed in terms of b and σ .
- (D) The Dynkin formula.
- (E) The characteristic operator.

This will give us the necessary background for the applications in the remaining chapters.

(A) The Markov property

DEFINITION 7.1. A (time-homogeneous) *Ito diffusion* is a stochastic process $X_t(\omega) = X(t, \omega) : [0, \infty) \times \Omega \rightarrow \mathbf{R}^n$ satisfying a stochastic differential equation of the form

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t, \quad t \geq s; \quad X_s = x \quad (7.4)$$

where B_t is m-dimensional Brownian motion and $b : \mathbf{R}^n \rightarrow \mathbf{R}^n, \sigma : \mathbf{R}^n \rightarrow \mathbf{R}^{n \times m}$ satisfy the conditions in Theorem 5.5, which in this case simplify to:

$$|b(x) - b(y)| + |\sigma(x) - \sigma(y)| \leq D|x - y|; \quad x, y \in \mathbf{R}^n, \quad (7.5)$$

where $|\sigma|^2 = \sum |\sigma_{ij}|^2$.

We will denote the (unique) solution of (7.4) by $X_t = X_t^{s,x}; t \geq s$. If $s = 0$ we write X_t^x for $X_t^{0,x}$. Note that we have assumed in (7.4) that b and σ do not depend on t but on x only. We shall see later (Chapters X, XI) that the general case can be reduced to this situation. The resulting process $X_t(\omega)$ will have the property of being *time-homogeneous*, in the following sense:

Note that

$$\begin{aligned} X_{s+h}^{s,x} &= x + \int_s^{s+h} b(X_u^{s,x})du + \int_s^{s+h} \sigma(X_u^{s,x})dB_u \\ &= x + \int_0^h b(X_{s+v}^{s,x})dv + \int_0^h \sigma(X_{s+v}^{s,x})d\tilde{B}_v, \quad (u = s + v) \end{aligned} \quad (7.6)$$

where $\tilde{B}_v = B_{s+v} - B_s; v \geq 0$. (See Exercise 2.11). On the other hand of course

$$X_h^{0,x} = x + \int_0^h b(X_v^{0,x})dv + \int_0^h \sigma(X_v^{0,x})dB_v.$$

Since $\{\tilde{B}_v\}_{v \geq 0}$ and $\{B_v\}_{v \geq 0}$ have the same P^0 -distributions, it follows by weak uniqueness (see (5.28)) of the solution of the stochastic differential equation

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t; \quad X_0 = x$$

that

$$\{X_{s+h}^{s,x}\}_{h \geq 0} \text{ and } \{X_h^{0,x}\}_{h \geq 0}$$

have the same P^0 -distributions, i.e. $\{X_t\}_{t \geq 0}$ is *time-homogeneous*.

We now introduce the probability laws Q^x of $\{X_t\}_{t \geq 0}$, for $x \in \mathbf{R}^n$. Intuitively, Q^x gives the distribution of $\{X_t\}_{t \geq 0}$ assuming that $X_0 = x$. To express this mathematically, we let \mathcal{M} be the σ -algebra (of subsets of Ω) generated by the random variables $\omega \rightarrow X_t(\omega) = X_t^y(\omega)$, where $t \geq 0, y \in \mathbf{R}^n$.

Define Q^x on the members of \mathcal{M} by

$$Q^x[X_{t_1} \in E_1, \dots, X_{t_k} \in E_k] = P^0[X_{t_1}^x \in E_1, \dots, X_{t_k}^x \in E_k] \quad (7.7)$$

where $E_i \subset \mathbf{R}^n$ are Borel sets; $1 \leq i \leq k$.

As before we let \mathcal{F}_t be the σ -algebra generated by $\{B_r; r \leq t\}$. Similarly we let \mathcal{M}_t be the σ -algebra generated by $\{X_r; r \leq t\}$. We have established earlier (see Theorem 5.5) that X_t is measurable with respect to \mathcal{F}_t , so $\mathcal{M}_t \subseteq \mathcal{F}_t$.

We now prove that X_t satisfies the important *Markov property*: The future behaviour of the process given what has happened up to time t is the same as the behaviour obtained when starting the process at X_t . The precise mathematical formulation of this is the following:

THEOREM 7.2. (The Markov property for Ito diffusions)

Let f be a bounded Borel function from \mathbf{R}^n to \mathbf{R} . Then, for $t, h \geq 0$

$$E^x[f(X_{t+h})|\mathcal{F}_t](\omega) = E^{X_t(\omega)}[f(X_h)]. \quad (7.8)$$

(See Appendix B for definition and basic properties of conditional expectation). Here and in the following E^x denotes the expectation w.r.t. the probability measure Q^x . Thus $E^y[f(X_h)]$ means $E[f(X_h^y)]$, where E denotes the expectation w.r.t. the measure P^0 . The right hand side means the function $E^y[f(X_h)]$ evaluated at $y = X_t(\omega)$.

Proof. Since, for $r \geq t$,

$$X_r(\omega) = X_t(\omega) + \int_t^r b(X_u)du + \int_t^r \sigma(X_u)dB_u,$$

we have by uniqueness

$$X_r(\omega) = X_r^{t, X_t}(\omega).$$

In other words, if we define

$$F(x, t, r, \omega) = X_r^{t, x}(\omega) \text{ for } r \geq t,$$

we have

$$X_r(\omega) = F(X_t, t, r, \omega); \quad r \geq t. \quad (7.9)$$

Note that $\omega \rightarrow F(x, t, r, \omega)$ is independent of \mathcal{F}_t . Using (7.9) we may rewrite (7.8) as

$$E[f(F(X_t, t, t+h, \omega))|\mathcal{F}_t] = E[f(F(x, 0, h, \omega))]_{x=X_t}. \quad (7.10)$$

Put $g(x, \omega) = f \circ F(x, t, t+h, \omega)$. Then $(x, \omega) \rightarrow g(x, \omega)$ is measurable. (See Exercise 7.6). Hence we can approximate g pointwise boundedly by functions on the form

$$\sum_{k=1}^m \phi_k(x)\psi_k(\omega).$$

Using the properties of conditional expectation (see Appendix B) we get

$$\begin{aligned} E[g(X_t, \omega)|\mathcal{F}_t] &= E[\lim \sum \phi_k(X_t)\psi_k(\omega)|\mathcal{F}_t] \\ &= \lim \sum \phi_k(X_t) \cdot E[\psi_k(\omega)|\mathcal{F}_t] \\ &= \lim \sum E[\phi_k(y)\psi_k(\omega)|\mathcal{F}_t]_{y=X_t} \\ &= E[g(y, \omega)|\mathcal{F}_t]_{y=X_t} = E[g(y, \omega)]_{y=X_t}. \end{aligned}$$

Therefore, since $\{X_t\}$ is time-homogeneous,

$$\begin{aligned} E[f(F(X_t, t, t+h, \omega))|\mathcal{F}_t] &= E[f(F(y, t, t+h, \omega))]_{y=X_t} \\ &= E[f(F(y, 0, h, \omega))]_{y=X_t} \end{aligned}$$

which is (7.10).

REMARK. Theorem 7.2 states that X_t is a Markov process w.r.t. the family of σ -algebras $\{\mathcal{F}_t\}_{t \geq 0}$. Note that since $\mathcal{M}_t \subseteq \mathcal{F}_t$ this implies that X_t is also a Markov process w.r.t. the σ -algebras $\{\mathcal{M}_t\}_{t \geq 0}$. This follows from Theorem B.3 and Theorem B.2 c)(Appendix B):

$$\begin{aligned} E^x[f(X_{t+h})|\mathcal{M}_t] &= E^x[E^x[f(X_{t+h})|\mathcal{F}_t]|\mathcal{M}_t] \\ &= E^x[E^{X_t}[f(X_h)]|\mathcal{M}_t] = E^{X_t}[f(X_h)] \end{aligned}$$

since $E^{X_t}[f(X_h)]$ is \mathcal{M}_t -measurable.

(B) The strong Markov property

Roughly, the strong Markov property states that a relation of the form (7.8) continues to hold if the time t is replaced by a random time $\tau(\omega)$ of a more general type called *stopping time*:

DEFINITION 7.3. Let $\{\mathcal{N}_t\}$ be an increasing family of σ -algebras (of subsets of Ω). A function $\tau : \Omega \rightarrow [0, \infty]$ is called a (strict) *stopping time* w.r.t. \mathcal{N}_t if

$$\{\omega; \tau(\omega) \leq t\} \in \mathcal{N}_t, \text{ for all } t \geq 0.$$

In other words, it should be possible to decide whether or not $\tau \leq t$ has occurred on the basis of the knowledge of \mathcal{N}_t .

EXAMPLE 7.4. Let $U \subset \mathbf{R}^n$ be open. Then the *first exit time*

$$\tau_U = \inf\{t > 0; X_t \notin U\}$$

is a stopping time w.r.t. \mathcal{M}_t , since

$$\{\omega; \tau_U \leq t\} = \bigcap_m \bigcup_{\substack{r \in \mathbf{Q} \\ r < t}} \{\omega; X_r \notin K_m\} \in \mathcal{M}_t$$

where $\{K_m\}$ is an increasing sequence of closed sets such that $U = \bigcup_m K_m$.

More generally, if $H \subset \mathbf{R}^n$ is any set we define the *first exit time from H* , τ_H , as follows

$$\tau_H = \inf\{t > 0; X_t \notin H\}.$$

If we include the sets of measure 0 in \mathcal{M}_t (which we do) then the family $\{\mathcal{M}_t\}$ is right-continuous i.e. $\mathcal{M}_t = \mathcal{M}_{t+}$, where $\mathcal{M}_{t+} = \bigcap_{s > t} \mathcal{M}_s$ (see Chung (1982,

Theorem 2.3.4., p.61)) and therefore τ_H is a stopping time for any Borel set H (see Dynkin (1965 II, 4.5.C.e.), p.111)).

DEFINITION 7.5. Let τ be a stopping time w.r.t. $\{\mathcal{N}_t\}$ and let \mathcal{N}_∞ be the smallest σ -algebra containing \mathcal{N}_t for all $t \geq 0$. Then the σ -algebra \mathcal{N}_τ consists of all sets $N \in \mathcal{N}_\infty$ such that

$$N \bigcap \{\tau \leq t\} \in \mathcal{N}_t \text{ for all } t \geq 0.$$

In the case when $\mathcal{N}_t = \mathcal{M}_t$, an alternative and more intuitive description is:

$$\mathcal{M}_\tau = \text{the } \sigma\text{-algebra generated by } \{X_{\min(s, \tau)}; s \geq 0\}. \quad (7.11)$$

(See Rao (1977 p. 2.15) or Stroock and Varadhan (1979, Lemma 1.3.3, p. 33).) Similarly, if $\mathcal{N}_t = \mathcal{F}_t$, we get

$$\mathcal{F}_\tau = \text{the } \sigma\text{-algebra generated by } \{B_{\min(s, \tau)}; s \geq 0\}.$$

THEOREM 7.6. (The strong Markov property for Ito diffusions)

Let f be a bounded Borel function on \mathbf{R}^n , τ a stopping time w.r.t. $\mathcal{F}_t, \tau < \infty$ a.s. Then

$$E^x[f(X_{\tau+h})|\mathcal{F}_\tau] = E^{X_\tau}[f(X_h)] \quad \text{for all } h \geq 0. \quad (7.12)$$

Proof. We try to imitate the proof of the Markov property (Theorem 7.2). For a.a. ω we have that $X_r^{\tau,x}(\omega)$ satisfies

$$X_{\tau+h}^{\tau,x} = x + \int_{\tau}^{\tau+h} b(X_u^{\tau,x})du + \int_{\tau}^{\tau+h} \sigma(X_u^{\tau,x})dB_u.$$

By the strong Markov property for Brownian motion (Gihman and Skorohod (1974a, p. 30)) the process

$$\tilde{B}_v = B_{\tau+v} - B_\tau; \quad v \geq 0$$

is again a Brownian motion and independent of \mathcal{F}_τ . Therefore

$$X_{\tau+h}^{\tau,x} = x + \int_0^h b(X_{\tau+v}^{\tau,x}) dv + \int_0^h \sigma(X_{\tau+v}^{\tau,x}) d\tilde{B}_v.$$

Hence $\{X_{\tau+h}^{\tau,x}\}_{h \geq 0}$ must coincide a.e. with the strongly unique (see (5.20)) solution Y_h of the equation

$$Y_h = x + \int_0^h b(Y_v) dv + \int_0^h \sigma(Y_v) d\tilde{B}_v.$$

Since $\{Y_h\}_{h \geq 0}$ is independent of \mathcal{F}_τ , $\{X_{\tau+h}^{\tau,x}\}$ must be independent also. Moreover, by weak uniqueness (see (5.28)) we conclude that

$$\{Y_h\}_{h \geq 0}, \text{ and hence } \{X_{\tau+h}^{\tau,x}\}_{h \geq 0}, \text{ has the same law as } \{X_h^{0,x}\}_{h \geq 0}. \quad (7.13)$$

Put

$$F(x, t, r, \omega) = X_r^{t,s}(\omega) \quad \text{for } r \geq t.$$

Then (7.12) can be written

$$E[f(F(x, 0, \tau + h, \omega)) | \mathcal{F}_\tau] = E[f(F(x, 0, h, \omega))]_{x=X_\tau^{0,x}}.$$

Now, with $X_t = X_t^{0,x}$,

$$\begin{aligned} F(x, 0, \tau + h, \omega) &= X_{\tau+h}(\omega) = x + \int_0^{\tau+h} b(X_s) ds + \int_0^{\tau+h} \sigma(X_s) dB_s \\ &= x + \int_0^\tau b(X_s) ds + \int_0^\tau \sigma(X_s) dB_s + \int_\tau^{\tau+h} b(X_s) ds + \int_\tau^{\tau+h} \sigma(X_s) dB_s \\ &= X_\tau + \int_\tau^{\tau+h} b(X_s) ds + \int_\tau^{\tau+h} \sigma(X_s) dB_s \\ &= F(X_\tau, \tau, \tau + h, \omega). \end{aligned}$$

Hence (7.12) gets the form

$$E[f(F(X_\tau, \tau, \tau + h, \omega)) | \mathcal{F}_\tau] = E[f(F(x, 0, h, \omega))]_{x=X_\tau}.$$

Put $g(x, t, r, \omega) = f(F(x, t, r, \omega))$. As in the proof of Theorem 7.2 we may assume that g has the form

$$g(x, t, r, \omega) = \sum_k \phi_k(x) \psi_k(t, r, \omega).$$

Then, since $X_{\tau+h}^{\tau,x}$ is independent of \mathcal{F}_τ we get, using (7.13)

$$\begin{aligned} E[g(X_\tau, \tau, \tau+h, \omega) | \mathcal{F}_\tau] &= \sum_k E[\phi_k(X_\tau) \psi_k(\tau, \tau+h, \omega) | \mathcal{F}_\tau] \\ &= \sum_k \phi_k(X_\tau) E[\psi_k(\tau, \tau+h, \omega) | \mathcal{F}_\tau] = \sum_k E[\phi_k(x) \psi_k(\tau, \tau+h, \omega) | \mathcal{F}_\tau]_{x=X_\tau} \\ &= E[g(x, \tau, \tau+h, \omega) | \mathcal{F}_\tau]_{x=X_\tau} = E[g(x, \tau, \tau+h, \omega)]_{x=X_\tau} \\ &= E[f(X_{\tau+h}^{\tau,x})]_{x=X_\tau} = E[f(X_h^{0,x})]_{x=X_\tau} = E[f(F(x, 0, h, \omega))]_{x=X_\tau}. \end{aligned}$$

□

We now extend (7.12) to the following:

If f_1, \dots, f_k are bounded Borel functions on \mathbf{R}^n , τ an \mathcal{F}_t -stopping time, $\tau < \infty$ a.s. then

$$E^x[f_1(X_{\tau+h_1}) f_2(X_{\tau+h_2}) \cdots f_k(X_{\tau+h_k}) | \mathcal{F}_\tau] = E^{X_\tau}[f_1(X_{h_1}) \cdots f_k(X_{h_k})] \quad (7.14)$$

for all $0 \leq h_1 \leq h_2 \leq \cdots \leq h_k$. This follows by induction: To illustrate the argument we prove it in the case $k = 2$:

$$\begin{aligned} E^x[f_1(X_{\tau+h_1}) f_2(X_{\tau+h_2}) | \mathcal{F}_\tau] &= E^x[E^x[f_1(X_{\tau+h_1}) f_2(X_{\tau+h_2}) | \mathcal{F}_{\tau+h_1}] | \mathcal{F}_\tau] \\ &= E^x[f_1(X_{\tau+h_1}) E^x[f_2(X_{\tau+h_2}) | \mathcal{F}_{\tau+h_1}] | \mathcal{F}_\tau] \\ &= E^x[f_1(X_{\tau+h_1}) E^{X_{\tau+h_1}}[f_2(X_{h_2-h_1})] | \mathcal{F}_\tau] \\ &= E^{X_\tau}[f_1(X_{h_1}) E^{X_{h_1}}[f_2(X_{h_2-h_1})]] \\ &= E^{X_\tau}[f_1(X_{h_1}) E^x[f_2(X_{h_2}) | \mathcal{F}_{h_1}]] = E^{X_\tau}[f_1(X_{h_1}) f_2(X_{h_2})], \text{ as claimed.} \end{aligned}$$

Next we proceed to formulate the general version we need: Let \mathcal{H} be the set of all real \mathcal{M} -measurable functions. For $t \geq 0$ we define the *shift operator*

$$\theta_t : \mathcal{H} \rightarrow \mathcal{H}$$

as follows:

If $\eta = g_1(X_{t_1}) \cdots g_k(X_{t_k})$ (g_i Borel measurable, $t_i \geq 0$) we put

$$\theta_t \eta = g_1(X_{t_1+t}) \cdots g_k(X_{t_k+t}).$$

Now extend in the natural way to all functions in \mathcal{H} by taking limits of sums of such functions. Then it follows from (7.14) that

$$E^x[\theta_\tau \eta | \mathcal{F}_\tau] = E^{X_\tau}[\eta] \quad (7.15)$$

for all stopping times τ and all bounded $\eta \in \mathcal{H}$, where

$$(\theta_\tau \eta)(\omega) = (\theta_t \eta)(\omega) \text{ if } \tau(\omega) = t.$$

Hitting distribution, harmonic measure and the mean value property

We will apply this to the following situation: Let $H \subset \mathbf{R}^n$ be measurable and let τ_H be the first exit time from H for an Ito diffusion X_t . Let α be another stopping time, g a bounded continuous function on \mathbf{R}^n and put

$$\eta = g(X_{\tau_H}) \mathcal{X}_{\{\tau_H < \infty\}}, \tau_H^\alpha = \inf\{t > \alpha; X_t \notin H\}.$$

Then we have

$$\theta_\alpha \eta \cdot \mathcal{X}_{\{\alpha < \infty\}} = g(X_{\tau_H^\alpha}) \mathcal{X}_{\{\tau_H^\alpha < \infty\}}. \quad (7.16)$$

To prove (7.16) we approximate η by functions $\eta^{(k)}$; $k = 1, 2, \dots$, of the form

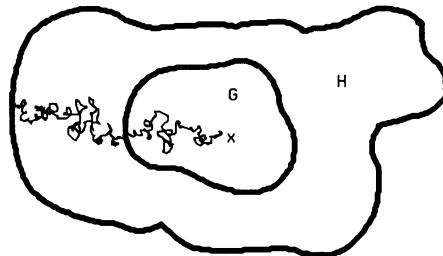
$$\eta^{(k)} = \sum_j g(X_{t_j}) \mathcal{X}_{[t_j, t_{j+1})}(\tau_H), \quad t_j = j \cdot 2^{-k}, \quad j = 0, 1, 2, \dots$$

Now

$$\begin{aligned} \theta_t \mathcal{X}_{[t_j, t_{j+1})}(\tau_H) &= \theta_t \mathcal{X}_{\{\forall r \in (0, t_j) X_r \in H \& \exists s \in [t_j, t_{j+1}) X_s \notin H\}} \\ &= \mathcal{X}_{\{\forall r \in (0, t_j) X_{r+t} \in H \& \exists s \in [t_j, t_{j+1}) X_{s+t} \notin H\}} \\ &= \mathcal{X}_{\{\forall u \in (t, t_j+t) X_u \in H \& \exists v \in [t_j+t, t_{j+1}+t) X_v \notin H\}} = \mathcal{X}_{[t_j+t, t_{j+1}+t)}(\tau_H^t). \end{aligned}$$

So we see that

$$\begin{aligned} \theta_t \eta &= \lim_k \theta_t \eta^{(k)} = \lim_k \sum_j g(X_{t_j+t}) \mathcal{X}_{[t_j+t, t_{j+1}+t)}(\tau_H^t) \\ &= g(X_{\tau_H^t}) \cdot \mathcal{X}_{\{\tau_H^t < \infty\}}, \text{ which is (7.16).} \end{aligned}$$



In particular, if $\alpha = \tau_G$ with $G \subset\subset H$ measurable, $\tau_H < \infty$ a.s. Q^x , then we have $\tau_H^\alpha = \tau_H$ and so

$$\theta_{\tau_G} g(X_{\tau_H}) = g(X_{\tau_H}). \quad (7.17)$$

So if f is any bounded measurable function we obtain from (7.15) and (7.17):

$$E^x[f(X_{\tau_H})] = E^x[E^{X_{\tau_G}}[f(X_{\tau_H})]] = \int_{\partial G} E^y[f(X_{\tau_H})] \cdot Q^x[X_{\tau_G} \in dy] \quad (7.18)$$

for $x \in G$.

(Define $\mu_H^x(F) = Q^x(X_{\tau_H} \in F)$ and approximate f in $L^1(\mu_H^x)$ by continuous functions g satisfying (7.17)). In other words, the expected value of f at X_{τ_H} when starting at $x \in G$ can be obtained by integrating the expected value when starting at $y \in \partial G$ with respect to the *hitting distribution* (“harmonic measure”) of X on ∂G . This can be restated as follows:

Define the harmonic measure of X on ∂G , μ_G^x , by

$$\mu_G^x(F) = Q^x[X_{\tau_G} \in F] \text{ for } F \subset \partial G, \quad x \in G.$$

Then the function

$$\phi(x) = E^x[f(X_{\tau_H})]$$

satisfies the mean value property:

$$\phi(x) = \int_{\partial G} \phi(y) d\mu_G^x(y), \quad \text{for all } x \in G \quad (7.19)$$

for all Borel sets $G \subset \subset H$.

This is an important ingredient in our solution of the generalized Dirichlet problem in Chapter IX.

(C) The generator of an Ito diffusion

It is fundamental for many applications that we can associate a second order partial differential operator A to an Ito diffusion X_t . The basic connection between A and X_t is that A is the generator of the process X_t :

DEFINITION 7.7. Let $\{X_t\}$ be a (time-homogeneous) Ito diffusion in \mathbf{R}^n . The (infinitesimal) generator A of X_t is defined by

$$Af(x) = \lim_{t \downarrow 0} \frac{E^x[f(X_t)] - f(x)}{t}; \quad x \in \mathbf{R}^n.$$

The set of functions $f : \mathbf{R}^n \rightarrow \mathbf{R}$ such that the limit exists at x is denoted by $\mathcal{D}_A(x)$, while \mathcal{D}_A denotes the set of functions for which the limit exists for all $x \in \mathbf{R}^n$.

To find the relation between A and the coefficients b, σ in the stochastic differential equation (7.4) defining X_t we need the following result, which is useful in many connections:

LEMMA 7.8. Let $Y_t = Y_t^x$ be an Ito process in \mathbf{R}^n of the form

$$Y_t^x(\omega) = x + \int_0^t u(s, \omega) ds + \int_0^t v(s, \omega) dB_s(\omega)$$

where u, v satisfy conditions (4.4), (4.5). Let $f \in C_0^2(\mathbf{R}^n)$ and let τ be a stopping time with respect to $\{\mathcal{F}_t\}$, and assume that $E^x[\tau] < \infty$. Assume that $u(t, \omega)$ and $v(t, \omega)$ are bounded on the set of (t, ω) such that $Y(t, \omega)$ belongs to the support of f . Then

$$\begin{aligned} E^x[f(Y_\tau)] &= f(x) + E^x\left[\int_0^\tau \left(\sum_i u_i(s, \omega) \frac{\partial f}{\partial x_i}(Y_s) \right.\right. \\ &\quad \left.\left. + \frac{1}{2} \sum_{i,j} (vv^T)_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(Y_s)\right) ds\right], \end{aligned}$$

where E^x is the expectation w.r.t. the natural probability law R^x for Y_t starting at x :

$$R^x[Y_{t_1} \in E_1, \dots, Y_{t_k} \in E_k] = P^0[Y_{t_1}^x \in E_1, \dots, Y_{t_k}^x \in E_k], E_i \text{ Borel sets.}$$

Proof. Put $Z = f(Y)$ and apply Ito's formula (To simplify the notation we suppress the index t and let Y_1, \dots, Y_n and B_1, \dots, B_m denote the coordinates of Y and B , respectively)

$$\begin{aligned} dZ &= \sum_i \frac{\partial f}{\partial x_i}(Y) dY_i + \frac{1}{2} \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(Y) dY_i dY_j \\ &= \sum_i u_i \frac{\partial f}{\partial x_i} dt + \frac{1}{2} \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j} (v dB)_i (v dB)_j + \sum_i \frac{\partial f}{\partial x_i} (v dB)_i. \end{aligned}$$

Since

$$\begin{aligned} (v dB)_i \cdot (v dB)_j &= (\sum_k v_{ik} dB_k)(\sum_n v_{jn} dB_n) \\ &= (\sum_k v_{ik} v_{jk}) dt = (vv^T)_{ij} dt, \end{aligned}$$

this gives

$$\begin{aligned} f(Y_t) &= f(Y_0) + \int_0^t \left(\sum_i u_i \frac{\partial f}{\partial x_i} + \frac{1}{2} \sum_{i,j} (vv^T)_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j} \right) ds \\ &\quad + \sum_{i,k} \int_0^t v_{ik} \frac{\partial f}{\partial x_i} dB_k. \end{aligned} \tag{7.20}$$

Hence

$$\begin{aligned} E^x[f(Y_\tau)] &= f(x) + E^x\left[\int_0^\tau \left(\sum_i u_i \frac{\partial f}{\partial x_i}(Y) + \frac{1}{2} \sum_{i,j} (vv^T)_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(Y)\right) ds\right] \\ &\quad + \sum_{i,k} \left(E^x \int_0^\tau v_{ik} \frac{\partial f}{\partial x_i}(Y) dB_k\right). \end{aligned} \quad (7.21)$$

If g is a bounded Borel function, $g \leq M$ say, then for all integers k we have

$$E^x\left[\int_0^{\tau \wedge k} g(Y_s) dB_s\right] = E^x\left[\int_0^k \chi_{\{s < \tau\}} g(Y_s) dB_s\right] = 0,$$

since $g(Y_s)$ and $\chi_{\{s < \tau\}}$ are both \mathcal{F}_s -measurable. Moreover

$$E^x\left[\left(\int_0^{\tau \wedge k} g(Y_s) dB_s\right)^2\right] = E^x\left[\int_0^{\tau \wedge k} g^2(Y_s) ds\right] \leq M^2 E^x[\tau] < \infty.$$

Therefore the family $\{\int_0^{\tau \wedge k} g(Y_s) dB_s\}_k$ is uniformly integrable w.r.t. the measure R^x (see Appendix C). It follows that

$$0 = \lim_{k \rightarrow \infty} E^x\left[\int_0^{\tau \wedge k} g(Y_s) dB_s\right] = E^x\left[\lim_{k \rightarrow \infty} \int_0^{\tau \wedge k} g(Y_s) dB_s\right] = E^x\left[\int_0^\tau g(Y_s) dB_s\right].$$

Combining this with (7.21) we get Lemma 7.8.

This gives immediately the formula for the generator A of an Ito diffusion:

THEOREM 7.9. Let X_t be the Ito diffusion

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

If $f \in C_0^2(\mathbf{R}^n)$, i.e. $f \in C^2(\mathbf{R}^n)$ and f has compact support, then $f \in \mathcal{D}_A$ and

$$Af(x) = \sum_i b_i(x) \frac{\partial f}{\partial x_i} + \frac{1}{2} \sum_{i,j} (\sigma \sigma^T)_{i,j}(x) \frac{\partial^2 f}{\partial x_i \partial x_j} \quad (7.22)$$

Proof. This follows from Lemma 7.8 (with $\tau = t$) and the definition of A .

EXAMPLE 7.10. The n-dimensional Brownian motion is of course the solution of the stochastic differential equation

$$dX_t = dB_t,$$

i.e. we have $b = 0$ and $\sigma = I_n$, the n-dimensional identity matrix. So the generator of B_t is

$$Af = \frac{1}{2} \sum \frac{\partial^2 f}{\partial x_i^2}; \quad f = f(x_1, \dots, x_n) \in C_0^2(\mathbf{R}^n)$$

i.e. $A = \frac{1}{2}\Delta$, where Δ is the Laplace operator.

EXAMPLE 7.11. (The graph of Brownian motion). Let B denote 1-dimensional Brownian motion and let $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ be the solution of the stochastic differential equation

$$\begin{cases} dX_1 = dt; & X_1(0) = t_0 \\ dX_2 = dB; & X_2(0) = x_0 \end{cases}$$

i.e.

$$dX = bdt + \sigma dB; \quad X(0) = \begin{pmatrix} t_0 \\ X_0 \end{pmatrix},$$

with $b = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\sigma = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. In other words, X may be regarded as the graph of Brownian motion. The generator A of X is given by

$$Af = \frac{\partial f}{\partial t} + \frac{1}{2} \frac{\partial^2 f}{\partial x^2}; \quad f = f(t, x) \in C_0^2(\mathbf{R}^n).$$

From now on we will, unless otherwise stated, let $A = A_X$ denote the generator of the Ito diffusion X_t . We let $L = L_X$ denote the differential operator given by the right hand side of (7.22). From Theorem 7.9 we know that A_X and L_X coincide on $C_0^2(\mathbf{R}^n)$.

(D) The Dynkin formula.

If we combine (7.21) and (7.22) we get:

THEOREM 7.12 (Dynkin's formula)

Let $f \in C_0^2(\mathbf{R}^n)$. Suppose τ is a stopping time, $E^x[\tau] < \infty$. Then

$$E^x[f(X_\tau)] = f(x) + E^x \left[\int_0^\tau Af(X_s) ds \right] \tag{7.23}$$

REMARKS.

- (i) Note that if τ is the first exit time of a bounded set, $E^x[\tau] < \infty$, then (7.23) holds for any function $f \in C^2$.
- (ii) It is possible to prove the theorem under the weaker assumption that f is a bounded function in \mathcal{D}_A . See Dynkin (1965 I), p. 133.

EXAMPLE 7.13. Consider n -dimensional Brownian motion $B = (B_1, \dots, B_n)$ starting at $a = (a_1, \dots, a_n) \in \mathbf{R}^n$ ($n \geq 2$) and assume $|a| < R$. What is the expected value of the first exit time τ_K of B from the ball

$$K = K_R = \{x \in \mathbf{R}^n; |x| < R\}?$$

Choose an integer k and apply Dynkin's formula with $X = B, \tau = \sigma_k = \min(k, \tau_K)$, and $f \in C_0^2$ such that $f(x) = |x|^2$ for $|x| \leq R$:

$$\begin{aligned} E^a[f(B_{\sigma_k})] &= f(a) + E^a\left[\int_0^{\sigma_k} \frac{1}{2} \Delta f(B_s) ds\right] \\ &= |a|^2 + E^a\left[\int_0^{\sigma_k} n \cdot ds\right] = |a|^2 + n \cdot E^a[\sigma_k]. \end{aligned}$$

Hence $E^a[\sigma_k] \leq \frac{1}{n}(R^2 - |a|^2)$ for all k . So letting $k \rightarrow \infty$ we conclude that $\tau_K = \lim \sigma_k < \infty$ a.s. and

$$E^a[\tau_K] = \frac{1}{n}(R^2 - |a|^2). \quad (7.24)$$

Next we assume that $|b| > R$. What is the probability that B starting at b ever hits K ?

Let α_k be the first exit time from the annulus

$$A_k = \{x; R < |x| < 2^k R\}; \quad k = 1, 2, \dots$$

Let $f = f_{n,k}$ be a C^2 function with compact support such that, if $R \leq |x| \leq 2^k R$,

$$f(x) = \begin{cases} -\log|x| & \text{when } n = 2 \\ |x|^{2-n} & \text{when } n > 2. \end{cases}$$

Then, since $\Delta f = 0$ in A_k , we have by Dynkin's formula

$$E^b[f(B_{\alpha_k})] = f(b) \quad \text{for all } k. \quad (7.25)$$

Put

$$p_k = P^b[|B_{\alpha_k}| = R], q_k = P^b[|B_{\alpha_k}| = 2^k R].$$

Let us now consider the two cases $n = 2$ and $n > 2$ separately:

$n = 2$. Then we get from (7.25)

$$-\log R \cdot p_k - (\log R + k \cdot \log 2)q_k = -\log|b| \quad \text{for all } k \quad (7.26)$$

This implies that $q_k \rightarrow 0$ as $k \rightarrow \infty$, so that

$$P^b[\tau_K < \infty] = 1, \quad (7.27)$$

i.e. Brownian motion is recurrent in \mathbf{R}^2 . (See Port and Stone (1979)).

n > 2. In this case (7.25) gives

$$p_k \cdot R^{2-n} + q_k \cdot (2^k R)^{2-n} = |b|^{2-n}$$

Since $0 \leq q_k \leq 1$ we get by letting $k \rightarrow \infty$

$$\lim_{k \rightarrow \infty} p_k = P^b[\tau_K < \infty] = \left(\frac{|b|}{R}\right)^{2-n},$$

i.e. Brownian motion is transient in \mathbf{R}^n for $n > 2$.

(E) The characteristic operator

We now introduce an operator which is closely related to the generator A , but is more suitable in many situations, for example in the solution of the Dirichlet problem.

DEFINITION 7.14. Let $\{X_t\}$ be an Ito diffusion. The *characteristic operator* $\mathcal{A} = \mathcal{A}_X$ of $\{X_t\}$ is defined by

$$\mathcal{A}f(x) = \lim_{U \downarrow x} \frac{E^x[f(X_{\tau_U})] - f(x)}{E^x[\tau_U]}, \quad (7.28)$$

where the U 's are open sets U_k decreasing to the point x , in the sense that $U_{k+1} \subset U_k$ and $\bigcap_k U_k = \{x\}$, and $\tau_U = \inf\{t > 0; X_t \notin U\}$ is the first exit time from U for X_t . The set of functions f such that the limit (7.28) exists for all $x \in \mathbf{R}^n$ (and all $\{U_k\}$) is denoted by $\mathcal{D}_{\mathcal{A}}$. If $E^x[\tau_U] = \infty$ for all open $U \ni x$, we define $\mathcal{A}f(x) = 0$.

It turns out that $\mathcal{D}_A \subseteq \mathcal{D}_{\mathcal{A}}$ always and that

$$Af = \mathcal{A}f \quad \text{for all } f \in \mathcal{D}_A.$$

(See Dynkin (1965 I, p. 143).)

We will only need that \mathcal{A}_X and L_X coincide on C^2 . To obtain this we first clarify a property of exit times.

DEFINITION 7.15. A point $x \in \mathbf{R}^n$ is called a *trap* for $\{X_t\}$ if

$$Q^x(\{X_t = x \text{ for all } t\}) = 1.$$

In other words, x is trap if and only if $\tau_{\{x\}} = \infty$ a.s. Q^x . For example, if $b(x_0) = \sigma(x_0) = 0$, then x_0 is a trap for X_t (by strong uniqueness of X_t).

LEMMA 7.16. If x is not a trap for X_t , then there exists an open set $U \ni x$ such that

$$E^x[\tau_U] < \infty.$$

Proof. See Lemma 5.5 p. 139 in Dynkin (1965 I).

THEOREM 7.17. Let $f \in C^2$. Then $f \in \mathcal{D}_A$ and

$$\mathcal{A}f = \sum_i b_i \frac{\partial f}{\partial x_i} + \frac{1}{2} \sum_{i,j} (\sigma \sigma^T)_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j} \quad (7.29)$$

Proof. As before we let L denote the operator defined by the right hand side of (7.29). If x is a trap for $\{X_t\}$ then $\mathcal{A}f(x) = 0$. Choose a bounded open set V such that $x \in V$. Modify f to f_0 outside V such that $f_0 \in C_0^2(\mathbf{R}^n)$. Then $f_0 \in \mathcal{D}_A(x)$ and $0 = A f_0(x) = L f_0(x) = L f(x)$. Hence $\mathcal{A}f(x) = L f(x) = 0$ in this case. If x is not a trap, choose a bounded open set $U \ni x$ such that $E^x[\tau_U] < \infty$. Then by Dynkin's formula (Theorem 7.12) (and the following Remark (i)), writing $\tau_U = \tau$

$$\begin{aligned} \left| \frac{E^x[f(X_\tau)] - f(x)}{E^x[\tau]} - L f(x) \right| &= \frac{|E^x[\int_0^\tau \{(L f)(X_s) - L f(x)\} ds]|}{E^x[\tau]} \\ &\leq \sup_{y \in U} |L f(x) - L f(y)| \rightarrow 0 \quad \text{as } U \downarrow x, \end{aligned}$$

since $L f$ is a continuous function.

REMARK. We have now obtained that an Ito diffusion is a continuous, strong Markov process such that the domain of definition of its characteristic operator includes C^2 . Thus an Ito diffusion is a *diffusion* in the sense of Dynkin (1965 I).

EXAMPLE 7.18. (Brownian motion on the unit circle). The characteristic operator of the process $Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$ from Example 5.4 satisfying the stochastic differential equations (5.13), i.e.

$$\begin{cases} dY_1 = -\frac{1}{2}Y_1 dt - Y_2 dB \\ dY_2 = -\frac{1}{2}Y_2 dt + Y_1 dB \end{cases}$$

is

$$\mathcal{A}f(y_1, y_2) = \frac{1}{2} [y_2^2 \frac{\partial^2 f}{\partial y_1^2} - 2y_1 y_2 \frac{\partial^2 f}{\partial y_1 \partial y_2} + y_1^2 \frac{\partial^2 f}{\partial y_2^2} - y_1 \frac{\partial f}{\partial y_1} - y_2 \frac{\partial f}{\partial y_2}].$$

This is because $dY = -\frac{1}{2}Y dt + KY dB$, where

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

so that

$$dY = b(Y)dt + \sigma(Y)dB$$

with

$$b(y_1, y_2) = \begin{pmatrix} -\frac{1}{2}y_1 \\ -\frac{1}{2}y_2 \end{pmatrix}, \quad \sigma(y_1, y_2) = \begin{pmatrix} -y_2 \\ y_1 \end{pmatrix}$$

and

$$a = \frac{1}{2}\sigma\sigma^T = \frac{1}{2} \begin{pmatrix} y_2^2 & -y_1y_2 \\ -y_1y_2 & y_1^2 \end{pmatrix}.$$

EXAMPLE 7.19. Let D be an open subset of \mathbf{R}^n such that $\tau_D < \infty$ a.s. Q^x for all x . Let ϕ be a bounded, measurable function on ∂D and define

$$\tilde{\phi}(x) = E^x[\phi(X_{\tau_D})]$$

($\tilde{\phi}$ is called the X -harmonic extension of ϕ). Then if U is open, $x \in U \subset\subset D$, we have by (7.18) that

$$E^x[\tilde{\phi}(X_{\tau_U})] = E^x[E^{X_{\tau_U}}[\phi(X_{\tau_D})]] = E^x[\phi(X_{\tau_D})] = \tilde{\phi}(x)$$

So $\tilde{\phi} \in \mathcal{D}_A$ and

$$\mathcal{A}\tilde{\phi} = 0 \text{ in } D,$$

in spite of the fact that in general $\tilde{\phi}$ need not even be continuous in D (See Example 9.7).

EXERCISES

7.1.

Find the generator of the following Ito diffusions:

- a) $dX_t = \mu X_t dt + \sigma dB_t$ (The Ornstein-Uhlenbeck process) ($B_t \in \mathbf{R}$; μ, σ constants).
- b) $dX_t = r X_t dt + \alpha X_t dB_t$ (The geometric Brownian motion) ($B_t \in \mathbf{R}$; r, α constants).
- c) $dY_t = r dt + \alpha Y_t dB_t$ ($B_t \in \mathbf{R}$; r, α constants)
- d) $dY_t = \begin{bmatrix} dt \\ dX_t \end{bmatrix}$ where X_t is as in a)
- e) $\begin{bmatrix} dX_1 \\ dX_2 \end{bmatrix} = \begin{bmatrix} 1 \\ X_2 \end{bmatrix} dt + \begin{bmatrix} 0 \\ e^{X_1} \end{bmatrix} dB_t$ ($B_t \in \mathbf{R}$)
- f) $\begin{bmatrix} dX_1 \\ dX_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} dt + \begin{bmatrix} 1 & 0 \\ 0 & X_1 \end{bmatrix} \begin{bmatrix} dB_1 \\ dB_2 \end{bmatrix}$

g) $X(t) = (X_1, X_2, \dots, X_n)$, where

$$dX_k(t) = r_k X_k dt + X_k \cdot \sum_{j=1}^n \alpha_{kj} dB_j; \quad 1 \leq k \leq n$$

$((B_1, \dots, B_n)$ is Brownian motion in \mathbf{R}^n , r_k and α_{kj} are constants).

7.2.

Find an Ito diffusion (i.e. write down the stochastic differential equation for it) whose generator is the following:

a) $Af(x) = f'(x) + f''(x); \quad f \in C_0^2(\mathbf{R})$

b) $Af(t, x) = \frac{\partial f}{\partial t} + cx \frac{\partial f}{\partial x} + \frac{1}{2}\alpha^2 x^2 \frac{\partial^2 f}{\partial x^2}; \quad f \in C_0^2(\mathbf{R}^2),$
where c, α are constants.

c) $Af(x_1, x_2) = 2x_2 \frac{\partial f}{\partial x_1} + \ln(1 + x_1^2 + x_2^2) \frac{\partial f}{\partial x_2}$
 $+ \frac{1}{2}(1 + x_1^2) \frac{\partial^2 f}{\partial x_1^2} + x_1 \frac{\partial^2 f}{\partial x_1 \partial x_2} + \frac{1}{2} \cdot \frac{\partial^2 f}{\partial x_2^2}; \quad f \in C_0^2(\mathbf{R}^2).$

7.3.

Let B_t be Brownian motion on \mathbf{R} , $B_0 = 0$ and define

$$X_t = X_t^x = x \cdot e^{ct + \alpha B_t},$$

where c, α are constants. Prove directly from the definition that X_t is a Markov process.

7.4.

Let B_t^x be 1-dimensional Brownian motion starting at $x \in \mathbf{R}^+$. Put

$$\tau = \inf\{t > 0; B_t^x = 0\}$$

- a) Prove that $\tau < \infty$ a.s. P^x for all $x > 0$. (Hint: See Example 7.13, second part).
- b) Prove that $E^x[\tau] = \infty$ for all $x > 0$. (Hint: See Example 7.13, first part).
- c) Show that Dynkin's formula does not hold if the condition $E^x[\tau] < \infty$ is dropped.

7.5.

Let the functions b, σ satisfy condition (5.14) of Theorem 5.5, with a constant C independent of t , i.e.

$$|b(t, x)| + |\sigma(t, x)| \leq C(1 + |x|) \quad \text{for all } x \in \mathbf{R}^n \text{ and all } t \geq 0$$

Let X_t be a solution of

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t.$$

Show that

$$E[|X_t|^2] \leq (1 + E[|X_0|^2])e^{Kt} - 1$$

for some constant K independent of t .

(Hint: Use Dynkin's formula with $f(x) = |x|^2$ and $\tau = t \wedge \tau_R$, where $\tau_R = \inf\{t > 0; |X_t| \geq R\}$, and let $R \rightarrow \infty$ to achieve the inequality

$$E[|X_t|^2] \leq E[|X_0|^2] + K \cdot \int_0^t (1 + E[|X_s|^2])ds,$$

which is of the form (5.21).)

7.6.

Let $g(x, \omega) = f \circ F(x, t, t+h, \omega)$ be as in the proof of Theorem 7.2. Assume that f is continuous.

- a) Prove that the map $x \rightarrow g(x, \cdot)$ is continuous from \mathbf{R}^n into $L^2(P)$ by using (5.22).

For simplicity assume that $n = 1$ in the following.

- b) Use a) to prove that $(x, \omega) \rightarrow g(x, \omega)$ is measurable. (Hint: For each $m = 1, 2, \dots$ put $\xi_k = \xi_k^{(m)} = k \cdot 2^{-m}$, $k = 1, 2, \dots$. Then

$$g^{(m)}(x, \cdot) := \sum_k g(\xi_k, \cdot) \cdot \chi_{\{\xi_k \leq x < \xi_{k+1}\}}$$

converges to $g(x, \cdot)$ in $L^2(P)$ for each x . Deduce that $g^{(m)} \rightarrow g$ in $L^2(dm_R \times dP)$ for all R , where dm_R is Lebesgue measure on $\{|x| \leq R\}$. So a subsequence of $g^{(m)}(x, \omega)$ converges to $g(x, \omega)$ for a.a. (x, ω) .)

7.7.

Let B_t be Brownian motion on \mathbf{R}^n starting at $x \in \mathbf{R}^n$ and let $D \subset \mathbf{R}^n$ be an open ball centered at x .

- a) Use Exercise 2.11 to prove that the harmonic measure μ_D^x of B_t is rotation invariant (about x) on the sphere ∂D . Conclude that μ_D^x coincides with normalized surface measure σ on ∂D .
- b) Let ϕ be a bounded measurable function on a bounded open set $W \subset \mathbf{R}^n$ and define

$$u(x) = E^x[\phi(B_{\tau_W})] \quad \text{for } x \in W.$$

Prove that u satisfies the classical mean value property:

$$u(x) = \int_{\partial D} u(y) d\sigma(y)$$

for all balls D centered at x with $\overline{D} \subset W$.

7.8.

Let $\{\mathcal{N}_t\}$ be a right-continuous family of σ -algebras of subsets of Ω , containing all sets of measure zero.

- a) Let τ_1, τ_2 be stopping times (w.r.t. \mathcal{N}_t). Prove that $\tau_1 \wedge \tau_2$ and $\tau_1 \vee \tau_2$ are stopping times.
- b) If $\{\tau_n\}$ is a decreasing family of stopping times prove that $\tau := \lim_n \tau_n$ is a stopping time.
- c) If X_t is an Ito diffusion in \mathbf{R}^n and $F \subset \mathbf{R}^n$ is closed, prove that τ_F is a stopping time w.r.t. \mathcal{M}_t . (Hint: Consider open sets decreasing to F).

7.9.

Let X_t be a geometric Brownian motion, i.e.

$$dX_t = rX_t dt + \alpha X_t dB_t, \quad X_0 = x > 0$$

where $B_t \in \mathbf{R}$; r, α are constants.

- a) Find the generator A of X_t and compute $Af(x)$ when $f(x) = x^\gamma$; $x > 0$, γ constant.
- b) If $r < \frac{1}{2}\alpha^2$ then $X_t \rightarrow 0$ as $t \rightarrow \infty$, a.s. Q^x (Example 5.1). But what is the probability p that X_t , when starting from $x < R$, ever hits the value R ? Use Dynkins formula with $f(x) = x^{\gamma_1}$, $\gamma_1 = 1 - \frac{2r}{\alpha^2}$, to prove that

$$p = \left(\frac{x}{R}\right)^{\gamma_1}$$

- c) If $r > \frac{1}{2}\alpha^2$ then $X_t \rightarrow \infty$ as $t \rightarrow \infty$, a.s. Q^x . Put

$$\tau = \inf\{t > 0; X_t \geq R\}.$$

Use Dynkin's formula with $f(x) = \ln x$, $x > 0$ to prove that

$$E^x[\tau] = \frac{\ln \frac{R}{x}}{r - \frac{1}{2}\alpha^2}$$

(Hint: First consider exit times from (ρ, R) , $\rho > 0$ and then let $\rho \rightarrow 0$. You need estimates for

$$(1 - p(\rho)) \ln \rho,$$

where

$$p(\rho) = Q^x[X_t \text{ reaches the value } R \text{ before } \rho],$$

which you can get from the calculations in a), b.).

7.10.

- a) Let $B_t \in \mathbf{R}^2$, $B_0 = x \neq 0$. Fix $0 < \epsilon < R < \infty$ and define

$$X_t = \ln |B_{t \wedge \tau}|; \quad t \geq 0$$

where

$$\tau = \inf\{t > 0; |B_t| \leq \epsilon \text{ or } |B_t| \geq R\}.$$

Prove that X_t is an $\mathcal{F}_{t \wedge \tau}$ -martingale. (Hint: Use Exercise 4.8.)

- b) Let $B_t \in \mathbf{R}^n$ for $n \geq 3$, $B_0 = x \neq 0$. Fix $\epsilon > 0$, $R < \infty$ and define

$$Y_t = [B_{t \wedge \tau}]^{2-n}; \quad t \geq 0$$

where

$$\tau = \inf\{t > 0; |B_t| \leq \epsilon \text{ or } |B_t| \geq R\}.$$

Prove that X_t is an $\mathcal{F}_{t \wedge \tau}$ -martingale.

7.11. (Doob's h-transform)

Let B_t be n-dimensional Brownian motion, $D \subset \mathbf{R}^n$ a bounded open set and $h > 0$ a harmonic function on D (i.e. $\Delta h = 0$ in D). Let X_t be the solution of the stochastic differential equation

$$dX_t = \nabla(\ln h)(X_t)dt + dB_t$$

More precisely, choose an increasing sequence $\{D_k\}$ of open subsets of D such that $\overline{D}_k \subset D$ and $\bigcup_{k=1}^{\infty} D_k = D$. Then for each k the equation above can be solved (strongly) for $t < \tau_{D_k}$. This gives in a natural way a solution for $t < \tau := \lim_{k \rightarrow \infty} \tau_{D_k}$.

- a) Show that the generator A of X_t satisfies

$$Af = \frac{\Delta(hf)}{2h} \quad \text{for } f \in C_0^2(D)$$

In particular, if $f = \frac{1}{h}$ then $Af = 0$.

- b) Use a) to show that if there exists $x_0 \in \partial D$ such that

$$\lim_{x \rightarrow y \in \partial D} h(x) = \begin{cases} 0 & \text{if } y \neq x_0 \\ \infty & \text{if } y = x_0 \end{cases}$$

(i.e. h is a *kernel function*), then

$$\lim_{t \rightarrow \tau} X_t = x_0 \text{ a.s.}$$

(Hint: Consider $E^x[f(X_T)]$ for suitable stopping times T and with $f = \frac{1}{h}$)

In other words, we have imposed a drift on B_t which causes the process to exit from D at the point x_0 only. This can also be formulated as follows: X_t is obtained by *conditioning B_t to exit from D at x_0* . See Doob (1984).

VIII. Other Topics in Diffusion Theory

In this chapter we study some other important topics in diffusion theory and related areas. While not strictly necessary for the remaining chapters, these topics are central in the theory of stochastic analysis and essential for further applications. The following topics will be treated:

- A) Kolmogorov's backward equation. The resolvent.
- B) The Feynman-Kac formula. Killing.
- C) The martingale problem.
- D) When is an Ito process a diffusion?
- E) Random time change.
- F) The Girsanov formula.

A) Kolmogorov's backward equation. The resolvent.

In the following we let X_t be an Ito diffusion in \mathbf{R}^n with generator A . If we choose $f \in C_0^2(\mathbf{R}^n)$ and $\tau = t$ in Dynkin's formula (7.23) we see that

$$u(t, x) = E^x[f(X_t)]$$

is differentiable with respect to t and

$$\frac{\partial u}{\partial t} = E^x[Af(X_t)]. \quad (8.1)$$

It turns out that the right hand side of (8.1) can be expressed in terms of u also:

THEOREM 8.1. (Kolmogorov's backward equation)

Let $f \in C_0^2(\mathbf{R}^n)$.

- a) Define

$$u(t, x) = E^x[f(X_t)]. \quad (8.2)$$

Then $u(t, \cdot) \in \mathcal{D}_A$ for each t and

$$\frac{\partial u}{\partial t} = Au, \quad t > 0, \quad x \in \mathbf{R}^n \quad (8.3)$$

$$u(0, x) = f(x); \quad x \in \mathbf{R}^n \quad (8.4)$$

where the right hand side is to be interpreted as A applied to the function $x \rightarrow u(t, x)$.

- b) Moreover, if $w(t, x) \in C^{1,2}(\mathbf{R} \times \mathbf{R}^n)$ is a bounded function satisfying (8.3), (8.4) then $w(t, x) = u(t, x)$, given by (8.2).

Proof. a) Let $g(x) = u(t, x)$. Then since $t \rightarrow u(t, x)$ is differentiable we have

$$\begin{aligned} \frac{E^x[g(X_r)] - g(x)}{r} &= \frac{1}{r} \cdot E^x[E^{X_r}[f(X_t)] - E^x[f(X_t)]] \\ &= \frac{1}{r} \cdot E^x[E^x[f(X_{t+r})|\mathcal{F}_r] - E^x[f(X_t)|\mathcal{F}_r]] \\ &= \frac{1}{r} \cdot E^x[f(X_{t+r}) - f(X_t)] \\ &= \frac{u(t+r, x) - u(t, x)}{r} \rightarrow \frac{\partial u}{\partial t} \text{ as } r \downarrow 0. \end{aligned}$$

Hence

$$Au = \lim_{r \downarrow 0} \frac{E^x[g(X_r)] - g(x)}{r} \quad \text{exists and } \frac{\partial u}{\partial t} = Au, \text{ as asserted.}$$

Conversely, to prove the uniqueness statement in b) assume that a function $w(t, x) \in C^{1,2}(\mathbf{R} \times \mathbf{R}^n)$ satisfies (8.3)–(8.4). Then

$$\tilde{A}w := -\frac{\partial w}{\partial t} + Aw = 0 \quad \text{for } t > 0, x \in \mathbf{R}^n \quad (8.5)$$

and

$$w(0, x) = f(x), \quad x \in \mathbf{R}^n. \quad (8.6)$$

Fix $(s, x) \in \mathbf{R} \times \mathbf{R}^n$. Define the process Y_t in \mathbf{R}^{n+1} by $Y_t = (s-t, X_t^{0,x})$, $t \geq 0$. Then Y_t has generator \tilde{A} and so by (8.5) and Dynkin's formula we have, for all $t \geq 0$ and all $T > 0$,

$$E^{s,x}[w(Y_{t \wedge \tau_R})] = w(s, x) + E^{s,x}\left[\int_0^{t \wedge \tau_R} \tilde{A}w(Y_r) dr\right] = w(s, x),$$

where $\tau_R = \inf\{t > 0; |X_t| \geq R\}$.

Letting $R \rightarrow \infty$ we get

$$w(s, x) = E^{s,x}[w(Y_t)]; \quad \forall t \geq 0.$$

In particular, choosing $t = s$ we get

$$w(s, x) = E^{s,x}[w(Y_s)] = E[w(0, X_s^{0,x})] = E[f(X_s^{0,x})] = E^x[f(X_s)].$$

□

REMARK. If we introduce the operator $Q_t : f \rightarrow E^\bullet[f(X_t)]$ then we have $u(t, x) = (Q_t f)(x)$ and we may rewrite (8.1) and (8.3) as follows:

$$\frac{d}{dt}(Q_t f) = Q_t(Af); \quad f \in C_0^2(\mathbf{R}^n) \quad (8.1)'$$

$$\frac{d}{dt}(Q_t f) = A(Q_t f); \quad f \in C_0^2(\mathbf{R}^n) \quad (8.3)'$$

Thus the equivalence of (8.1) and (8.3) amounts to saying that the operators Q_t and A commute, in some sense. Arguing formally, it is tempting to say that the solution of (8.1)' and (8.3)' is

$$Q_t = e^{tA}$$

and therefore $Q_t A = A Q_t$. However, this argument would require a further explanation, because in general A is an unbounded operator.

It is an important fact that the operator A always has an inverse, at least if a positive multiple of the identity is subtracted from A . This inverse can be expressed explicitly in terms of the diffusion X_t :

DEFINITION 8.2. For $\alpha > 0$ and $g \in C_b(\mathbf{R}^n)$ we define *the resolvent operator* R_α by

$$R_\alpha g(x) = E^x \left[\int_0^\infty e^{-\alpha t} g(X_t) dt \right]. \quad (8.7)$$

LEMMA 8.3. $R_\alpha g$ is a bounded continuous function.

Proof. Since $R_\alpha g(x) = \int_0^\infty e^{-\alpha t} E^x[g(X_t)] dt$, we see that Lemma 8.3 is a direct consequence of the next result:

LEMMA 8.4. Let g be a lower bounded, measurable function on \mathbf{R}^n and define, for fixed $t \geq 0$

$$u(x) = E^x[g(X_t)].$$

- a) If g is lower semicontinuous, then u is lower semicontinuous.
- b) If g is bounded and continuous, then u is continuous. In other words, any Ito diffusion X_t is *Feller-continuous*.

Proof. By (5.22) we have

$$E[|X_t^x - X_t^y|^2] \leq |y - x|^2 C(t),$$

where $C(t)$ does not depend on x and y . Let $\{y_n\}$ be a sequence of points converging to x . Then

$$X_t^{y_n} \rightarrow X_t^x \text{ in } L^2(\Omega, P^0) \text{ as } n \rightarrow \infty.$$

So, by taking a subsequence $\{z_n\}$ of $\{y_n\}$ we obtain that

$$X_t^{z_n}(\omega) \rightarrow X_t^x(\omega) \text{ for a.a. } \omega \in \Omega.$$

a) If g is lower bounded and lower semicontinuous, then by the Fatou lemma

$$u(x) = E[g(X_t^x)] \leq E[\liminf_{n \rightarrow \infty} g(X_t^{z_n})] \leq \liminf_{n \rightarrow \infty} E[g(X_t^{z_n})] = \liminf_{n \rightarrow \infty} u(z_n).$$

Therefore every sequence $\{y_n\}$ converging to x has a subsequence $\{z_n\}$ such that $u(x) \leq \liminf_{n \rightarrow \infty} u(z_n)$. That proves that u is lower semicontinuous.

b) If g is bounded and continuous, the result in a) can be applied both to g and $-g$. Hence both u and $-u$ are lower semicontinuous and we conclude that u is continuous.

We now prove that R_α and $\alpha - A$ are inverse operators:

THEOREM 8.5.

- a) If $f \in C_0^2(\mathbf{R}^n)$ then $R_\alpha(\alpha - A)f = f$ for all $\alpha > 0$.
- b) If $g \in C_b(\mathbf{R}^n)$ then $R_\alpha g \in \mathcal{D}_A$ and $(\alpha - A)R_\alpha g = g$ for all $\alpha > 0$.

Proof. a) If $f \in C_0^2(\mathbf{R}^n)$ then by Dynkin's formula

$$\begin{aligned} R_\alpha(\alpha - A)f(x) &= (\alpha R_\alpha f - R_\alpha Af)(x) \\ &= \alpha \int_0^\infty e^{-\alpha t} E^x[f(X_t)] dt - \int_0^\infty e^{-\alpha t} E^x[Af(X_t)] dt \\ &= \int_0^\infty -e^{-\alpha t} E^x[f(X_t)] + \int_0^\infty e^{-\alpha t} \frac{d}{dt} E^x[f(X_t)] dt - \int_0^\infty e^{-\alpha t} E^x[Af(X_t)] dt \\ &= E^x[f(X_0)] = f(x) \end{aligned}$$

b) If $g \in C_b(\mathbf{R}^n)$ then by the strong Markov property

$$\begin{aligned} E^x[R_\alpha g(X_t)] &= E^x[E^{X_t}[\int_0^\infty e^{-\alpha s} g(X_s) ds]] \\ &= E^x[E^x[\theta_t(\int_0^\infty e^{-\alpha s} g(X_s) ds) | \mathcal{F}_t]] = E^x[E^x[\int_0^\infty e^{-\alpha s} g(X_{t+s}) ds | \mathcal{F}_t]] \\ &= E^x[\int_0^\infty e^{-\alpha s} g(X_{t+s}) ds] = \int_0^\infty e^{-\alpha s} E^x[g(X_{t+s})] ds. \end{aligned}$$

Therefore, since $u(s, x) = E^x[g(X_s)]$ is s-differentiable,

$$\begin{aligned} \frac{1}{t}(E^x[R_\alpha g(X_t)] - R_\alpha g(x)) &= \int_0^\infty e^{-\alpha s} \frac{u(t+s, x) - u(s, x)}{t} ds \\ &\rightarrow \int_0^\infty e^{-\alpha s} \frac{\partial}{\partial s} u(s, x) ds \quad \text{as } t \rightarrow 0. \end{aligned}$$

Hence $R_\alpha g \in \mathcal{D}_A$ and

$$\begin{aligned} (\alpha - A)R_\alpha g(x) &= \int_0^\infty \alpha e^{-\alpha s} u(s, x) ds - \int_0^\infty e^{-\alpha s} \frac{\partial}{\partial s} u(s, x) ds \\ &= \left| -e^{-\alpha s} u(s, x) \right|_0^\infty = u(0, x) = g(x). \end{aligned}$$

B) The Feynman-Kac formula. Killing

With a little harder work we can obtain the following useful generalization of Kolmogorov's backward equation:

THEOREM 8.6 (The Feynman-Kac formula)

Let $f \in C_0^2(\mathbf{R}^n)$ and $q \in C(\mathbf{R}^n)$. Assume that q is lower bounded.

a) Put

$$v(t, x) = E^x[\exp(-\int_0^t q(X_s) ds) f(X_t)]. \quad (8.8)$$

Then

$$\frac{\partial v}{\partial t} = Av - qv; \quad t > 0, x \in \mathbf{R}^n \quad (8.9)$$

$$v(0, x) = f(x); \quad x \in \mathbf{R}^n \quad (8.10)$$

- b) Moreover, if $w(t, x) \in C^{1,2}(\mathbf{R} \times \mathbf{R}^n)$ is bounded on $K \times \mathbf{R}^n$ for each compact $K \subset \mathbf{R}$ and w solves (8.9), (8.10), then $w(t, x) = v(t, x)$, given by (8.8).

Proof. a) Let $Y_t = f(X_t)$, $Z_t = \exp(-\int_0^t q(X_s) ds)$. Then dY_t is given by (7.20) and

$$dZ_t = -Z_t q(X_t) dt$$

So

$$d(Y_t Z_t) = Y_t dZ_t + Z_t dY_t, \text{ since } dZ_t \cdot dY_t = 0.$$

Note that since $Y_t Z_t$ is an Ito process it follows from Lemma 7.8 that $v(t, x) = E^x[Y_t Z_t]$ is differentiable w.r.t. t .

Therefore, with $v(t, x)$ as in (8.8) we get

$$\begin{aligned} \frac{1}{r}(E^x[v(t, X_r)] - v(t, x)) &= \frac{1}{r}E^x[E^{X_r}[Z_t f(X_t)] - E^x[Z_t f(X_t)]] \\ &= \frac{1}{r}E^x[E^x[f(X_{t+r}) \exp(-\int_0^t q(X_{s+r})ds)|\mathcal{F}_r] - E^x[Z_t f(X_t)|\mathcal{F}_r]] \\ &= \frac{1}{r}E^x[Z_{t+r} \cdot \exp(\int_0^r q(X_s)ds)f(X_{t+r}) - Z_t f(X_t)] \\ &= \frac{1}{r}E^x[f(X_{t+r})Z_{t+r} - f(X_t)Z_t] + \frac{1}{r}E^x[f(X_{t+r})Z_{t+r} \cdot (\exp(\int_0^r q(X_s)ds) - 1)] \\ &\rightarrow \frac{\partial}{\partial t}v(t, x) + q(x)v(t, x) \quad \text{as } r \rightarrow 0, \end{aligned}$$

because

$$\frac{1}{r}f(X_{t+r})Z_{t+r}(\exp(\int_0^r q(X_s)ds) - 1) \rightarrow f(X_t)Z_t q(X_0)$$

pointwise boundedly. That completes the proof of a).

b) Assume that $w(t, x) \in C^{1,2}(\mathbf{R} \times \mathbf{R}^n)$ satisfies (8.9) and (8.10) and that $w(t, x)$ is bounded on $K \times \mathbf{R}^n$ for each compact $K \subset \mathbf{R}$. Then

$$\hat{A}w(t, x) := -\frac{\partial w}{\partial t} + Aw - qw = 0 \quad \text{for } t > 0, x \in \mathbf{R}^n \quad (8.11)$$

and

$$w(0, x) = f(x); \quad x \in \mathbf{R}^n. \quad (8.12)$$

Fix $(s, x, z) \in \mathbf{R} \times \mathbf{R}^n \times \mathbf{R}^n$ and define $Z_t = z + \int_0^t q(X_s)ds$ and

$H_t = (s - t, X_t^{0,x}, Z_t)$. Then H_t is an Ito diffusion with generator

$$A_H \phi(s, x, z) = -\frac{\partial \phi}{\partial s} + A\phi + q(x)\frac{\partial \phi}{\partial z}; \quad \phi \in C_0^2(\mathbf{R} \times \mathbf{R}^n \times \mathbf{R}^n).$$

Hence by (8.11) and Dynkin's formula we have, for all $t \geq 0$, $R > 0$ and with $\phi(s, x, z) = \exp(-z)w(s, x)$:

$$E^{s,x,z}[\phi(H_{t \wedge \tau_R})] = \phi(s, x, z) + E^{s,x,z}\left[\int_0^{t \wedge \tau_R} A_H \phi(H_r)dr\right],$$

where $\tau_R = \inf\{t > 0; |H_t| \geq R\}$.

Note that with this choice of ϕ we have by (8.11)

$$A_H \phi(s, x, z) = \exp(-z) \left[-\frac{\partial w}{\partial s} + Aw - q(x)w \right] = 0.$$

Hence

$$\begin{aligned} w(s, x) &= \phi(s, x, 0) = E^{s, x, 0}[\phi(H_{t \wedge \tau_R})] \\ &= E^x \left[\exp \left(- \int_0^{t \wedge \tau_R} q(X_r) dr \right) w(s - t \wedge \tau_R, X_{t \wedge \tau_R}) \right] \\ &\rightarrow E^x \left[\exp \left(- \int_0^t q(X_r) dr \right) w(s - t, X_t) \right] \quad \text{as } R \rightarrow \infty, \end{aligned}$$

since $w(r, x)$ is bounded for $(r, x) \in K \times \mathbf{R}^n$. In particular, choosing $t = s$ we get

$$w(s, x) = E^x \left[\exp \left(- \int_0^s q(X_r) dr \right) w(0, X_s^{0, x}) \right] = v(s, x), \quad \text{as claimed.}$$

□

REMARK. (About killing a diffusion)

In Theorem 7.9 we have seen that the generator of an Ito diffusion X_t given by

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t \tag{8.13}$$

is a partial differential operator L of the form

$$Lf = \sum a_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j} + \sum b_i \frac{\partial f}{\partial x_i} \tag{8.14}$$

where $[a_{ij}] = \frac{1}{2}\sigma\sigma^T$, $b = [b_i]$. It is natural to ask if one can also find processes whose generator has the form

$$Lf = \sum a_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j} + \sum b_i \frac{\partial f}{\partial x_i} - cf, \tag{8.15}$$

where $c(x)$ is a bounded and continuous function.

If $c(x) \geq 0$ the answer is yes and a process \tilde{X}_t with generator (8.15) is obtained by *killing* X_t at a certain (*killing*) time ζ . By this we mean that there exists a random time ζ such that if we put

$$\tilde{X}_t = X_t \quad \text{if } t < \zeta \tag{8.16}$$

and leave \tilde{X}_t undefined if $t \geq \zeta$ (alternatively, put $\tilde{X}_t = \partial$ if $t \geq \zeta$, where $\partial \notin \mathbf{R}^n$ is some “coffin” state), then \tilde{X}_t is also a strong Markov process and

$$E^x[f(\tilde{X}_t)] = E^x[f(X_t), t < \zeta] = E^x[f(X_t) \cdot e^{-\int_0^t c(X_s)ds}] \quad (8.17)$$

for all bounded continuous functions f on \mathbf{R}^n .

Let $v(t, x)$ denote the right hand side of (8.17) with $f \in C_0^2(\mathbf{R}^n)$. Then

$$\lim_{t \rightarrow 0} \frac{E^x[f(\tilde{X}_t)] - f(x)}{t} = \frac{\partial}{\partial t} v(t, x)_{t=0} = (Av - cv)_{t=0} = Af(x) - c(x)f(x),$$

by the Feynman-Kac formula.

So the generator of \tilde{X}_t is (8.15), as required. The function $c(x)$ can be interpreted as the *killing rate*:

$$c(x) = \lim_{t \downarrow 0} \frac{1}{t} Q^x[X_0 \text{ is killed in the time interval } (0, t)].$$

Thus by applying such a killing procedure we can come from the special case $c = 0$ in (8.14) to the general case (8.15) with $c(x) \geq 0$. Therefore, for many purposes it is enough to consider the equation (8.14).

If the function $c(x) \geq 0$ is given, an explicit construction of the killing time ζ such that (8.17) holds can be found in Karlin and Taylor (1975), p. 314. For a more general discussion see Blumenthal and Getoor (1968), Ch. III.

C) The martingale problem

If $dX_t = b(X_t)dt + \sigma(X_t)dB_t$ is an Ito diffusion in \mathbf{R}^n with generator A and if $f \in C_0^2(\mathbf{R}^n)$ then by (7.20)

$$f(X_t) = f(x) + \int_0^t Af(X_s)ds + \int_0^t \nabla f^T(X_s)\sigma(X_s)dB_s \quad (8.18)$$

Define

$$M_t = f(X_t) - \int_0^t Af(X_r)dr (= f(x) + \int_0^t \nabla f^T(X_r)\sigma(X_r)dB_r). \quad (8.19)$$

Then, since Ito integrals are martingales (w.r.t. the σ -algebras $\{\mathcal{F}_t\}$) we have for $s > t$

$$E^x[M_s | \mathcal{F}_t] = M_t.$$

It follows that

$$E^x[M_s | \mathcal{M}_t] = E^x[E^x[M_s | \mathcal{F}_t] | \mathcal{M}_t] = E^x[M_t | \mathcal{M}_t] = M_t,$$

since M_t is \mathcal{M}_t -measurable. We have proved:

THEOREM 8.7. If X_t is an Ito diffusion in \mathbf{R}^n with generator A , then for all $f \in C_0^2(\mathbf{R}^n)$ the process

$$M_t = f(X_t) - \int_0^t Af(X_r)dr$$

is a martingale w.r.t. $\{\mathcal{M}_t\}$.

If we identify each $\omega \in \Omega$ with the function

$$\omega_t = \omega(t) = X_t^x(\omega)$$

we see that the probability space $(\Omega, \mathcal{M}, Q^x)$ is identified with

$$((\mathbf{R}^n)^{[0,\infty)}, \mathcal{B}, \tilde{Q}^x)$$

where \mathcal{B} is the Borel σ -algebra on $(\mathbf{R}^n)^{[0,\infty)}$ (see Ch. II). Thus, regarding the law of X_t^x as a probability measure \tilde{Q}^x on \mathcal{B} we can formulate Theorem 8.7 as follows:

THEOREM 8.7'. If \tilde{Q}^x is the probability measure on \mathcal{B} induced by the law Q^x of an Ito diffusion X_t , then for all $f \in C_0^2(\mathbf{R}^n)$ the process

$$M_t = f(X_t) - \int_0^t Af(X_r)dr (= f(\omega_t) - \int_0^t Af(\omega_r)dr); \quad \omega \in (\mathbf{R}^n)^{[0,\infty)} \quad (8.20)$$

is a \tilde{Q}^x -martingale w.r.t. the Borel σ -algebras \mathcal{B}_t of $(\mathbf{R}^n)^{[0,t]}$, $t \geq 0$. In other words, the measure \tilde{Q}^x solves the *martingale problem* for the differential operator A , in the following sense:

DEFINITION 8.8. Let L be a semi-elliptic differential operator of the form

$$L = \sum b_i \frac{\partial}{\partial x_i} + \sum_{i,j} a_{ij} \frac{\partial^2}{\partial x_i \partial x_j}$$

where the coefficients b_i, a_{ij} are locally bounded Borel measurable functions on \mathbf{R}^n . Then we say that a probability measure \tilde{P}^x on $((\mathbf{R}^n)^{[0,\infty)}, \mathcal{B})$ solves the *martingale problem for L* (starting at x) if the process

$$M_t = f(\omega_t) - \int_0^t Lf(\omega_r)dr, M_0 = f(x) \text{ a.s. } \tilde{P}^x$$

is a \tilde{P}^x martingale w.r.t. \mathcal{B}_t , for all $f \in C_0^2(\mathbf{R}^n)$. The martingale problem is called *well posed* if there is a unique measure \tilde{P}^x solving the martingale problem.

The argument of Theorem 8.7 actually proves that \tilde{Q}^x solves the martingale problem for A whenever X_t is a weak solution of the stochastic differential equation

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t \quad (8.21)$$

Conversely, it can be proved that if \tilde{P}^x solves the martingale problem for

$$L = \sum b_i \frac{\partial}{\partial x_i} + \frac{1}{2} \sum (\sigma \sigma^T)_{ij} \frac{\partial^2}{\partial x_i \partial x_j} \quad (8.22)$$

starting at x , for all $x \in \mathbf{R}^n$, then there exists a weak solution X_t of the stochastic differential equation (8.21). Moreover, this weak solution X_t is a Markov process if and only if the martingale problem for L is well posed. (See Stroock and Varadhan (1979) or Rogers and Williams (1987)). Therefore, if the coefficients b, σ of (8.21) satisfy the conditions (5.14), (5.15) of Theorem 5.5, we conclude that

$$\begin{aligned} \tilde{Q}^x \text{ is the } & \text{unique solution of the martingale problem} \\ \text{for the operator } L \text{ given by (8.22).} & \end{aligned} \quad (8.23)$$

Lipschitz-continuity of the coefficients of L is not necessary for the uniqueness of the martingale problem. For example, one of the spectacular results of Stroock and Varadhan (1979) is that

$$L = \sum b_i \frac{\partial}{\partial x_i} + \sum a_{ij} \frac{\partial^2}{\partial x_i \partial x_j}$$

has a unique solution of the martingale problem if $[a_{ij}]$ is everywhere positive definite, $a_{ij}(x)$ is continuous, $b(x)$ is measurable and there exists a constant D such that

$$|b(x)| + |a(x)|^{\frac{1}{2}} \leq D(1 + |x|) \quad \text{for all } x \in \mathbf{R}^n.$$

D) When is an Ito process a diffusion?

The Ito formula gives that if we apply a C^2 function $\phi : U \subset \mathbf{R}^n \rightarrow \mathbf{R}^n$ to an Ito process X_t the result $\phi(X_t)$ is another Ito process. A natural question is: If X_t is an Ito diffusion will $\phi(X_t)$ be an Ito diffusion too? The answer is no in general, but it may be yes in some cases:

EXAMPLE 8.9. (The Bessel process). Let $n \geq 2$. In Example 4.7 we found that the process

$$R_t(\omega) = |B(t, \omega)| = (B_1(t, \omega)^2 + \cdots + B_n(t, \omega)^2)^{\frac{1}{2}}$$

satisfied the equation

$$dR_t = \sum_{i=1}^n \frac{B_i dB_i}{R_t} + \frac{n-1}{2R_t} dt. \quad (8.24)$$

However, as it stands this is *not* a stochastic differential equation of the form (5.17), so it is not apparent from (8.24) that R is an Ito diffusion. But this will follow if we can show that

$$Y_t = \int_0^t \sum_{i=1}^n B_i |B|^{-1} dB_i$$

coincides in law with (i.e. has the same finite-dimensional distributions as) 1-dimensional Brownian motion \tilde{B}_t . For then (8.24) can be written

$$dR_t = \frac{n-1}{2R_t} dt + d\tilde{B}$$

which is of the form (5.17), thus showing by weak uniqueness (see (5.28)) that R_t is an Ito diffusion with generator

$$Af(x) = \frac{1}{2} f''(x) + \frac{n-1}{2x} f'(x)$$

as claimed in Example 4.7. One way of seeing that the process Y_t coincides in law with 1-dimensional Brownian motion \tilde{B}_t is to apply the following result:

THEOREM 8.10. An Ito process

$$dY_t = v dB_t; \quad Y_0 = 0 \quad \text{with } v(t, \omega) \in \mathbf{R}^{n \times m}$$

coincides (in law) with n -dimensional Brownian motion if and only if

$$vv^T(t, \omega) = I_n \quad \text{for a.a. } (t, \omega) \text{ w.r.t. } dt \times dP^0, \quad (8.25)$$

where I_n is the n -dimensional identity matrix and P^0 is the probability law of m -dimensional Brownian motion starting at 0.

Note that in the example above we have

$$Y_t = \int_0^t v dB$$

with

$$v = [B_1 |B|^{-1}, \dots, B_n |B|^{-1}], \quad B = \begin{pmatrix} B_1 \\ \vdots \\ B_n \end{pmatrix}$$

and since $vv^T = 1$, we get that Y_t is a 1-dimensional Brownian motion, as required.

Theorem 8.10 is a special case of the following result, which gives a necessary and sufficient condition for an Ito process to coincide in law with a given diffusion: (We use the symbol \sim for “coincides in law with”).

THEOREM 8.11. Let X_t be an Ito diffusion given by

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t, \quad b \in \mathbf{R}^n, \quad \sigma \in \mathbf{R}^{n \times m}, \quad X_0 = x,$$

and let Y_t be an Ito process given by

$$dY_t = u(t, \omega)dt + v(t, \omega)dB_t, \quad u \in \mathbf{R}^n, \quad v \in \mathbf{R}^{n \times m}, \quad Y_0 = x.$$

Then $X_t \sim Y_t$ if and only if

$$E^x[u(t, \cdot)|\mathcal{N}_t] = b(Y_t^x) \quad \text{and} \quad vv^T(t, \omega) = \sigma\sigma^T(Y_t^x) \quad (8.26)$$

for a.a. (t, ω) w.r.t. $dt \times dP^0$, where \mathcal{N}_t is the σ -algebra generated by $Y_s; s \leq t$.

Proof. Assume that (8.26) holds. Let

$$A = \sum b_i \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j} (\sigma\sigma^T)_{ij} \frac{\partial^2}{\partial x_i \partial x_j}$$

be the generator of X_t and define, for $f \in C_0^2(\mathbf{R}^n)$,

$$Hf(t, \omega) = \sum_i u_i(t, \omega) \frac{\partial f}{\partial x_i}(Y_t) + \frac{1}{2} \sum_{i,j} (vv^T)_{ij}(t, \omega) \frac{\partial^2 f}{\partial x_i \partial x_j}(Y_t)$$

Then by Ito's formula (see (7.20)) we have, for $s > t$,

$$\begin{aligned} E^x[f(Y_s)|\mathcal{N}_t] &= f(Y_t) + E^x \left[\int_t^s Hf(r, \omega)dr | \mathcal{N}_t \right] + E^x \left[\int_t^s \nabla f^T v dB_r | \mathcal{N}_t \right] \\ &= f(Y_t) + E^x \left[\int_t^s E^x[Hf(r, \omega)|\mathcal{N}_r]dr | \mathcal{N}_t \right] \\ &= f(Y_t) + E^x \left[\int_t^s Af(Y_r)dr | \mathcal{N}_t \right] \quad \text{by (8.26)}, \end{aligned} \quad (8.27)$$

where E^x denotes expectation w.r.t. the law R^x of Y_t (see Lemma 7.8). Therefore, if we define

$$M_t = f(Y_t) - \int_0^t Af(Y_r)dr \quad (8.28)$$

then, for $s > t$,

$$\begin{aligned} E^x[M_s | \mathcal{N}_t] &= f(Y_t) + E^x\left[\int_t^s Af(Y_r)dr | \mathcal{N}_t\right] - E^x\left[\int_0^s Af(Y_r)dr | \mathcal{N}_t\right] \\ &= f(Y_t) - E^x\left[\int_0^t Af(Y_r)dr | \mathcal{N}_t\right] = M_t \end{aligned}$$

Hence M_t is a martingale w.r.t. the σ -algebras \mathcal{N}_t and the law R^x . By uniqueness of the solution of the martingale problem (see (8.16)) we conclude that $X_t \sim Y_t$.

Conversely, assume that $X_t \sim Y_t$. Choose $f \in C_0^2$. By Ito's formula (7.20) we have, for a.a. (t, ω) w.r.t. $dt \times dP^0$,

$$\begin{aligned} &\lim_{h \downarrow 0} \frac{1}{h} (E^x[f(Y_{t+h}) | \mathcal{N}_t] - f(Y_t)) \\ &= \lim_{h \downarrow 0} \frac{1}{h} \left(\int_t^{t+h} E^x\left[\sum_i u_i(s, \omega) \frac{\partial f}{\partial x_i}(Y_s)\right. \right. \\ &\quad \left. \left. + \frac{1}{2} \sum_{i,j} (vv^T)_{ij}(s, \omega) \frac{\partial^2 f}{\partial x_i \partial x_j}(Y_s) | \mathcal{N}_t\right] ds \right) \end{aligned} \tag{8.29}$$

$$= \sum_i E^x[u_i(t, \omega) | \mathcal{N}_t] \frac{\partial f}{\partial x_i}(Y_t) + \frac{1}{2} \sum_{i,j} E^x[(vv^T)_{ij}(t, \omega) | \mathcal{N}_t] \frac{\partial^2 f}{\partial x_i \partial x_j}(Y_t) \tag{8.30}$$

On the other hand, since $X_t \sim Y_t$ we know that Y_t is a Markov process. Therefore (8.29) coincides with

$$\begin{aligned} &\lim_{h \downarrow 0} \frac{1}{h} (E^{Y_t}[f(Y_h)] - E^{Y_t}[f(Y_0)]) \\ &= \sum_i E^{Y_t}[u_i(0, \omega) \frac{\partial f}{\partial x_i}(Y_0)] + \frac{1}{2} \sum_{i,j} E^{Y_t}[(vv^T)_{ij}(0, \omega) \frac{\partial^2 f}{\partial x_i \partial x_j}(Y_0)] \\ &= \sum_i E^{Y_t}[u_i(0, \omega) \frac{\partial f}{\partial x_i}(Y_t) + \frac{1}{2} \sum_{i,j} E^{Y_t}[(vv^T)_{ij}(0, \omega)] \frac{\partial^2 f}{\partial x_i \partial x_j}(Y_t)]. \end{aligned} \tag{8.31}$$

Comparing (8.30) and (8.31) we conclude that

$$E^x[u(t, \omega) | \mathcal{N}_t] = E^{Y_t}[u(0, \omega)] \text{ and } E^x[vv^T(t, \omega) | \mathcal{N}_t] = E^{Y_t}[vv^T(0, \omega)] \tag{8.32}$$

for a.a. (t, ω) .

On the other hand, since the generator of Y_t coincides with the generator A of X_t we get from (8.31) that

$$E^{Y_t}[u(0, \omega)] = b(Y_t) \text{ and } E^{Y_t}[vv^T(0, \omega)] = \sigma\sigma^T(Y_t) \text{ for a.a. } (t, \omega). \tag{8.33}$$

Combining (8.32) and (8.33) we conclude that

$$E^x[u|\mathcal{N}_t] = b(Y_t) \text{ and } E^x[vv^T|\mathcal{N}_t] = \sigma\sigma^T(Y_t) \text{ for a.a. } (t, \omega). \quad (8.34)$$

From this we obtain (8.26) by using that in fact $vv^T(t, \cdot)$ is always \mathcal{N}_t -measurable, in the following sense:

LEMMA 8.12. Let $dY_t = u(t, \omega)dt + v(t, \omega)dB_t$, $Y_0 = x$ be as in Theorem 8.11. Then there exists an \mathcal{N}_t -adapted process $W(t, \omega)$ such that

$$vv^T(t, \omega) = W(t, \omega) \text{ for a.a. } (t, \omega).$$

Proof. By Ito's formula we have (if $Y_i(t, \omega)$ denotes component number i of $Y(t, \omega)$)

$$Y_i Y_j(t, \omega) = x_i x_j + \int_0^t Y_i dY_j(s) + \int_0^t Y_j dY_i(s) + \int_0^t (vv^T)_{ij}(s, \omega) ds.$$

Therefore, if we put

$$H_{ij}(t, \omega) = Y_i Y_j(t, \omega) - x_i x_j - \int_0^t Y_i dY_j - \int_0^t Y_j dY_i, \quad 1 \leq i, j \leq n$$

then H_{ij} is \mathcal{N}_t -adapted and

$$H_{ij}(t, \omega) = \int_0^t (vv^T)_{ij}(s, \omega) ds.$$

Therefore

$$(vv^T)_{ij}(t, \omega) = \lim_{r \downarrow 0} \frac{H(t, \omega) - H(t - r, \omega)}{r}$$

for a.a. t . This shows Lemma 8.12 and the proof of Theorem 8.11 is complete.

REMARKS. 1) One may ask if it also true that $u(t, \cdot)$ must be \mathcal{N}_t -measurable. However, the following example shows that this fails even in the case when $v = n = 1$:

Let B_1, B_2 be two independent 1-dimensional Brownian motions and define

$$dY_t = B_1(t)dt + dB_2(t)$$

Then we may regard Y_t as noisy observations of the process $B_1(t)$. So by Example 6.12 we have that

$$E^0[(B_1(t, \omega) - \hat{B}_1(t, \omega))^2] = \tanh(t),$$

where $\hat{B}_1(t, \omega) = E^0[B_1(t)|\mathcal{N}_t]$ is the Kalman-Bucy filter. In particular, $B_1(t, \omega)$ cannot be \mathcal{N}_t -measurable.

2) The process $v(t, \omega)$ need not be \mathcal{N}_t -adapted either: Let B_t be 1-dimensional Brownian motion and define

$$dY_t = \text{sign}(B_t)dB_t \quad (8.35)$$

where

$$\text{sign}(z) = \begin{cases} 1 & \text{if } z > 0 \\ -1 & \text{if } z \leq 0. \end{cases}$$

Tanaka's formula says that

$$|B_t| = |B_0| + \int_0^t \text{sign}(B_s)dB_s + L_t \quad (8.36)$$

where $L_t = L_t(\omega)$ is *local time* of B_t at 0, a non-decreasing process which only increases when $B_t = 0$ (see Exercise 4.11). Therefore the σ -algebra \mathcal{N}_t generated by $\{Y_s; s \leq t\}$ is contained in the σ -algebra \mathcal{H}_t generated by $\{|B_s|; s \leq t\}$. It follows that $v(t, \omega) = \text{sign}(B_t)$ cannot be \mathcal{N}_t -adapted.

COROLLARY 8.13. (How to recognize a Brownian motion).

Let

$$dY_t = u(t, \omega)dt + v(t, \omega)dB_t$$

be an Ito process in \mathbf{R}^n . Then Y_t is a Brownian motion if and only if

$$E^x[u(t, \cdot)|\mathcal{N}_t] = 0 \quad \text{and} \quad vv^T(t, \omega) = I_n \quad (8.37)$$

for a.a. (t, ω) .

REMARK. Using Theorem 8.11 one may now proceed to investigate when the image $Y_t = \phi(X_t)$ of an Ito diffusion X_t by a C^2 -function ϕ coincide in law with an Ito diffusion Z_t . Applying the criterion (8.26) one obtains the following result:

$$\begin{aligned} \phi(X_t) &\sim Z_t && \text{if and only if} \\ A[f \circ \phi] &= \hat{A}[f] \circ \phi \end{aligned} \quad (8.38)$$

for all second order polynomials $f(x_1, \dots, x_n) = \sum a_i x_i + \sum c_{ij} x_i x_j$ (and hence for all $f \in C_0^2$) where A and \hat{A} are the generators of X_t and Z_t respectively. For generalizations of this result, see Csink and Øksendal (1983) and Øksendal (1990).

E) Random time change

Let $c(t, \omega) \geq 0$ be an \mathcal{F}_t -adapted process. Define

$$\beta_t = \beta(t, \omega) = \int_0^t c(s, \omega) ds. \quad (8.39)$$

We will say that β_t is a (random) time change with time change rate $c(t, \omega)$.

Note that $\beta(t, \omega)$ is also \mathcal{F}_t -adapted and for each ω the map $t \rightarrow \beta_t(\omega)$ is non-decreasing. Define $\alpha_t = \alpha(t, \omega)$ by

$$\alpha_t = \inf\{s; \beta_s > t\}. \quad (8.40)$$

Then α_t is a right-inverse of β_t , for each ω :

$$\beta(\alpha(t, \omega), \omega) = t \quad \text{for all } t \geq 0. \quad (8.41)$$

Moreover, $t \rightarrow \alpha_t(\omega)$ is right-continuous.

If $c(s, \omega) > 0$ for a.a. (s, ω) then $t \rightarrow \beta_t(\omega)$ is strictly increasing, $t \rightarrow \alpha_t(\omega)$ is continuous and α_t is also a left-inverse of β_t :

$$\alpha(\beta(t, \omega), \omega) = t \quad \text{for all } t \geq 0. \quad (8.42)$$

In general $\omega \rightarrow \alpha(t, \omega)$ is an $\{\mathcal{F}_s\}$ -stopping time for each t , since

$$\{\omega; \alpha(t, \omega) < s\} = \{\omega; t < \beta(s, \omega)\} \in \mathcal{F}_s. \quad (8.43)$$

We now ask the question: Suppose X_t is an Ito diffusion and Y_t an Ito process as in Theorem 8.11. When does there exist a time change β_t such that $Y_{\alpha_t} \sim X_t$? (Note that α_t is only defined up to time β_∞ . If $\beta_\infty < \infty$ we interpret $Y_{\alpha_t} \sim X_t$ to mean that Y_{α_t} has the same law as X_t up to time β_∞).

Here is a partial answer (see Øksendal (1990)):

THEOREM 8.14. Let X_t, Y_t be as in Theorem 8.11 and let β_t be a time change with right inverse α_t as in (8.39), (8.40) above. Assume that

$$u(t, \omega) = c(t, \omega)b(Y_t) \quad \text{and} \quad vv^T(t, \omega) = c(t, \omega) \cdot \sigma\sigma^T(Y_t) \quad (8.44)$$

for a.a. t, ω . Then

$$Y_{\alpha_t} \sim X_t.$$

This result allows us to recognize time changes of Brownian motion:

THEOREM 8.15. Let $dY_t = v(t, \omega)dB_t, v \in \mathbf{R}^{n \times m}, B_t \in \mathbf{R}^m$ be an Ito integral in \mathbf{R}^n , $Y_0 = 0$ and assume that

$$vv^T(t, \omega) = c(t, \omega)I_n \quad (8.45)$$

for some process $c(t, \omega) \geq 0$. Let α_t, β_t be as in (8.39), (8.40). Then

Y_{α_t} is an n-dimensional Brownian motion.

COROLLARY 8.16. Let $dY_t = \sum_{i=1}^n v_i(t, \omega) dB_i(t, \omega)$, $Y_0 = 0$, where $B = (B_1, \dots, B_n)$ is a Brownian motion in \mathbf{R}^n . Then

$$\hat{B}_t := Y_{\alpha_t} \text{ is a 1-dimensional Brownian motion,}$$

where α_t is defined by (8.40) and

$$\beta_s = \int_0^s \left\{ \sum_{i=1}^n v_i^2(r, \omega) \right\} dr. \quad (8.46)$$

COROLLARY 8.17. Let Y_t, β_s be as in Corollary 8.16. Assume that

$$\sum_{i=1}^n v_i^2(r, \omega) > 0 \quad \text{for a.a. } (r, \omega). \quad (8.47)$$

Then there exists a Brownian motion \hat{B}_t such that

$$Y_t = \hat{B}_{\beta_t}. \quad (8.49)$$

Proof. Let

$$\hat{B}_t = Y_{\alpha_t} \quad (8.50)$$

be the Brownian motion from Corollary 8.16. By (8.47) β_t is strictly increasing and hence (8.42) holds. So choosing $t = \beta_s$ in (8.50) we get (8.49). \square

COROLLARY 8.18. Let $c(t, \omega) \geq 0$ be given and define

$$dY_t = \int_0^t \sqrt{c(s, \omega)} dB_s,$$

where B_s is an n-dimensional Brownian motion. Then

$$Y_{\alpha_t} \text{ is also an n-dimensional Brownian motion.}$$

We now use this to prove that a time change of an Ito integral is again an Ito integral, but driven by a different Brownian motion \tilde{B}_t . First we construct \tilde{B}_t :

LEMMA 8.19. Suppose $s \rightarrow \alpha(s, \omega)$ is continuous, $\alpha(0, \omega) = 0$ for a.a. ω . Fix $t > 0$ such that $\beta_t < \infty$ a.s. and assume that $E[\alpha_t] < \infty$ where E denotes expectation w.r.t. P^0 . For $k = 1, 2, \dots$ put

$$t_j = \begin{cases} j \cdot 2^{-k} & \text{if } j \cdot 2^{-k} \leq \alpha_t \\ t & \text{if } j \cdot 2^{-k} > \alpha_t \end{cases}$$

and choose r_j such that $\alpha_{r_j} = t_j$. Suppose $f(s, \omega) \geq 0$ is \mathcal{F}_s -adapted, bounded and s-continuous for a.a. ω . Then

$$\lim_{k \rightarrow \infty} \sum_j f(\alpha_j, \omega) \Delta B_{\alpha_j} = \int_0^{\alpha_t} f(s, \omega) dB_s \text{ a.s.}, \quad (8.51)$$

where $\alpha_j = \alpha_{r_j}$, $\Delta B_{\alpha_j} = B_{\alpha_{j+1}} - B_{\alpha_j}$ and the limit is in $L^2(\Omega, P^0)$.

Proof. For all k we have

$$\begin{aligned} & E[(\sum_j f(\alpha_j, \omega) \Delta B_{\alpha_j} - \int_0^{\alpha_t} f(s, \omega) dB_s)^2] \\ &= E[(\sum_j \int_{\alpha_j}^{\alpha_{j+1}} (f(\alpha_j, \omega) - f(s, \omega) dB_s)^2)] = \sum_j E[(\int_{\alpha_j}^{\alpha_{j+1}} (f(\alpha_j, \omega) - f(s, \omega)) dB_s)^2] \\ &= \sum_j E[\int_{\alpha_j}^{\alpha_{j+1}} (f(\alpha_j, \omega) - f(s, \omega))^2 ds] = E[\int_0^{\alpha_t} (f - f_k)^2 ds], \end{aligned}$$

where $f_k(s, \omega) = \sum_j f(t_j, \omega) \chi_{[t_j, t_{j+1})}(s)$ is the elementary approximation to f . (See Ch. III). This implies (8.37).

We now use this to establish a general time change formula for Ito integrals. An alternative proof in the case $n = m = 1$ can be found in McKean (1969, §2.8).

THEOREM 8.20. (Time change formula for Ito integrals).

Suppose $c(s, \omega)$ and $\alpha(s, \omega)$ are s-continuous, $\alpha(0, \omega) = 0$ for a.a. ω and that $E[\alpha_t] < \infty$. Let (B_s, \mathcal{F}_s) be an m-dimensional Brownian motion and let $v(s, \omega) \in \mathbf{R}^{n \times m}$ be bounded and s-continuous. Define

$$\tilde{B}_t = \lim_{k \rightarrow \infty} \sum_j \sqrt{c(\alpha_j, \omega)} \Delta B_{\alpha_j} = \int_0^{\alpha_t} \sqrt{c(s, \omega)} dB_s \quad (8.52)$$

Then \tilde{B}_t is an (m-dimensional) \mathcal{F}_{α_t} -Brownian motion (i.e. \tilde{B}_t is a Brownian motion and \tilde{B}_t is a martingale w.r.t. \mathcal{F}_{α_t}) and

$$\int_0^{\alpha_t} v(s, \omega) dB_s = \int_0^t v(\alpha_r, \omega) \sqrt{\alpha'_r(\omega)} d\tilde{B}_r \text{ a.s. } P^0, \quad (8.53)$$

where $\alpha'_r(\omega)$ is the derivative of $\alpha(r, \omega)$ w.r.t. r , so that

$$\alpha'_r(\omega) = \frac{1}{c(\alpha_r, \omega)} \text{ for a.a. } r \geq 0, \text{ a.a. } \omega \in \Omega. \quad (8.54)$$

Proof. The existence of the limit in (8.52) and the second identity in (8.52) follow by applying Lemma 8.19 to the function

$$f(s, \omega) = \sqrt{c(s, \omega)}.$$

Then by Corollary 8.18 we have that \tilde{B}_t is an \mathcal{F}_{α_t} -Brownian motion. It remains to prove (8.53):

$$\begin{aligned} \int_0^{\alpha_t} v(s, \omega) dB_s &= \lim_{k \rightarrow \infty} \sum_j v(\alpha_j, \omega) \Delta B_{\alpha_j} \\ &= \lim_{k \rightarrow \infty} \sum_j v(\alpha_j, \omega) \sqrt{\frac{1}{c(\alpha_j, \omega)}} \sqrt{c(\alpha_j, \omega)} \Delta B_{\alpha_j} \\ &= \lim_{k \rightarrow \infty} \sum_j v(\alpha_j, \omega) \sqrt{\frac{1}{c(\alpha_j, \omega)}} \Delta \tilde{B}_j \\ &= \int_0^t v(\alpha_r, \omega) \sqrt{\frac{1}{c(\alpha_r, \omega)}} d\tilde{B}_r \end{aligned}$$

and the proof is complete.

EXAMPLE 8.21. (Brownian motion on the unit sphere in \mathbf{R}^n ; $n > 2$).

In Examples 5.4 and 7.18 we constructed Brownian motion on the unit circle. It is not obvious how to extend the method used there to obtain Brownian motion on the unit sphere S of \mathbf{R}^n ; $n \geq 3$. However, we may proceed as follows: Apply the function $\phi : \mathbf{R}^n \setminus \{0\} \rightarrow S$ defined by

$$\phi(x) = x \cdot |x|^{-1}; \quad x \in \mathbf{R}^n \setminus \{0\}$$

to n-dimensional Brownian motion $B = (B_1, \dots, B_n)$. The result is a stochastic integral $Y = (Y_1, \dots, Y_n) = \phi(B)$ which by Ito's formula is given by

$$dY_i = \frac{|B|^2 - B_i^2}{|B|^3} dB_i - \sum_{j \neq i} \frac{B_j B_i}{|B|^3} dB_j - \frac{n-1}{2} \cdot \frac{B_i}{|B|^3} dt; \quad i = 1, 2, \dots, n \quad (8.55)$$

Hence

$$dY = \frac{1}{|B|} \cdot \sigma(Y) dB + \frac{1}{|B|^2} b(Y) dt,$$

where

$$\sigma = [\sigma_{ij}] \in \mathbf{R}^{n \times n}, \quad \text{with } \sigma_{ij}(Y) = \delta_{ij} - Y_i Y_j; \quad 1 \leq i, j \leq n$$

and

$$b(y) = -\frac{n-1}{2} \cdot \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathbf{R}^n, \quad (y_1, \dots, y_n \text{ are the coordinates of } y \in \mathbf{R}^n).$$

Now perform the following time change: Define

$$Z_t(\omega) = Y_{\alpha(t,\omega)}(\omega)$$

where

$$\alpha_t = \beta_t^{-1}, \beta(t, \omega) = \int_0^t \frac{1}{|B|^2} ds.$$

Then Z is again an Ito process and by Theorem 8.20

$$dZ = \sigma(Z)dB + b(Z)dt.$$

Hence Z is a diffusion with characteristic operator

$$\mathcal{A}f(y) = \frac{1}{2}(\Delta f(y) - \sum_{i,j} y_i y_j \frac{\partial^2 f}{\partial y_i \partial y_j}) - \frac{n-1}{2} \cdot \sum_i y_i \frac{\partial f}{\partial y_i}; \quad |y| = 1 \quad (8.56)$$

Thus, $\phi(B) = \frac{B}{|B|}$ is – after a suitable change of time scale – equal to a diffusion Z living on the unit sphere S of \mathbf{R}^n . Note that Z is invariant under orthogonal transformations in \mathbf{R}^n (since B is). It is reasonable to call Z *Brownian motion on the unit sphere S* . For other constructions see Ito and McKean (1965, p. 269 (§7.15)) and Stroock (1971).

More generally, given a Riemannian manifold M with metric tensor $g = [g_{ij}]$ one may define a *Brownian motion on M* as a diffusion on M whose characteristic operator \mathcal{A} in local coordinates x_i is given by $\frac{1}{2}$ times the Laplace-Beltrami operator (here $[g^{ij}] = [g_{ij}]^{-1}$)

$$\Delta_M = \frac{1}{\sqrt{\det(g)}} \cdot \sum_i \frac{\partial}{\partial x_i} (\sqrt{\det(g)} \sum_j g^{ij} \frac{\partial}{\partial x_j}) \quad (8.57)$$

See for example Meyer (1966, p. 256–270), McKean (1969, §4.3). The subject of stochastic differential equations on manifolds is also treated in Ikeda and Watanabe (1989), Emery (1989) and Elworthy (1982).

EXAMPLE 8.22. (Harmonic and analytic functions)

Let $B = (B_1, B_2)$ be 2-dimensional Brownian motion. Let us investigate what happens if we apply a C^2 function

$$\phi(x_1, x_2) = (u(x_1, x_2), v(x_1, x_2))$$

to B :

Put $Y = (Y_1, Y_2) = \phi(B_1, B_2)$ and apply Ito's formula:

$$dY_1 = u'_1(B_1, B_2)dB_1 + u'_2(B_1, B_2)dB_2 + \frac{1}{2}[u''_{11}(B_1, B_2) + u''_{22}(B_1, B_2)]dt$$

and

$$dY_2 = v'_1(B_1, B_2)dB_1 + v'_2(B_1, B_2)dB_2 + \frac{1}{2}[v''_{11}(B_1, B_2) + v''_{22}(B_1, B_2)]dt,$$

where $u'_1 = \frac{\partial u}{\partial x_1}$ etc. So

$$dY = b(B_1, B_2)dt + \sigma(B_1, B_2)dB,$$

with $b = \frac{1}{2} \begin{pmatrix} \Delta u \\ \Delta v \end{pmatrix}$, $\sigma = \begin{pmatrix} u'_1 & u'_2 \\ v'_1 & v'_2 \end{pmatrix} = D_\phi$ (the derivative of ϕ).

So $Y = \phi(B_1, B_2)$ is a martingale if (and, in fact, only if) ϕ is harmonic, i.e. $\Delta\phi = 0$. If ϕ is harmonic, we get by Corollary 8.16 that

$$\phi(B_1, B_2) = (\tilde{B}_{\beta_1}^{(1)}, \tilde{B}_{\beta_2}^{(2)})$$

where $\tilde{B}^{(1)}$ and $\tilde{B}^{(2)}$ are two (not necessarily independent) versions of 1-dimensional Brownian motion, and

$$\beta_1(t, \omega) = \int_0^t |\nabla u|^2(B_1, B_2)ds, \quad \beta_2(t, \omega) = \int_0^t |\nabla v|^2(B_1, B_2)ds.$$

Since

$$\sigma\sigma^T = \begin{pmatrix} |\nabla u|^2 & \nabla u \cdot \nabla v \\ \nabla u \cdot \nabla v & |\nabla v|^2 \end{pmatrix}$$

we see that if (in addition to $\Delta u = \Delta v = 0$)

$$|\nabla u|^2 = |\nabla v|^2 \quad \text{and} \quad \nabla u \cdot \nabla v = 0 \tag{8.58}$$

then

$$Y_t = Y_0 + \int_0^t \sigma dB$$

with

$$\sigma\sigma^T = |\nabla u|^2(B_1, B_2)I_2, \quad Y_0 = \phi(B_1(0), B_2(0)).$$

Therefore, if we let

$$\beta_t = \beta(t, \omega) = \int_0^t |\nabla u|^2(B_1, B_2)ds, \quad \alpha_t = \beta_t^{-1}$$

we obtain by Theorem 8.15 that Y_{α_t} is a 2-dimensional Brownian motion. Conditions (8.58) – in addition to $\Delta u = \Delta v = 0$ – are easily seen to be equivalent to requiring that the function $\phi(x + iy) = \phi(x, y)$ regarded as a complex function is either *analytic* or *conjugate analytic*.

Thus we have proved a theorem of P. Lévy that $\phi(B_1, B_2)$ is – after a change of time scale – again Brownian motion in the plane if and only if ϕ is either analytic or conjugate analytic. For extensions of this result see Bernard, Campbell and Davie (1979), Csink and Øksendal (1983) and Csink, Fitzsimmons and Øksendal (1990).

F) The Girsanov theorem

We end this chapter by discussing a result, the Girsanov theorem, which is fundamental in the general theory of stochastic analysis. It is also very important in many applications, for example in economics (see e.g. Duffie (1992)).

Basically the Girsanov theorem says that if we change the *drift* coefficient of a given Ito process (with a nondegenerate diffusion coefficient), then the law of the process will not change dramatically. In fact, the law of the new process will be absolutely continuous w.r.t. the law of the original process and we can compute explicitly the Radon-Nikodym derivative.

We now make this precise. First we state (without proof) the useful Lévy characterization of Brownian motion. A proof can be found in e.g. Ikeda & Watanabe (1989), Th. II.6.1, or in Karatzas & Shreve (1991), Th. 3.3.16.

THEOREM 8.23. (The Lévy characterization of Brownian motion)

Let $X(t) = (X_1(t), \dots, X_n(t))$ be a continuous stochastic process on a probability space (Ω, \mathcal{H}, Q) with values in \mathbf{R}^n . Then the following, a) and b), are equivalent

- a) $X(t)$ is a Brownian motion w.r.t. Q , i.e. the law of $X(t)$ w.r.t. Q is the same as the law of an n -dimensional Brownian motion.
- b) (i) $X(t) = (X_1(t), \dots, X_n(t))$ is a martingale w.r.t. Q (and w.r.t. its own filtration) and
 - (ii) $X_i(t)X_j(t) - \delta_{ij}t$ is a martingale w.r.t. Q (and w.r.t. its own filtration) for all $i, j \in \{1, 2, \dots, n\}$.

REMARK. In this Theorem one may replace condition (ii) by the condition

- (ii)' The cross-variation processes $\langle X_i, X_j \rangle_t$ satisfy the identity

$$\langle X_i, X_j \rangle_t(\omega) = \delta_{ij}t \text{ a.s., } 1 \leq i, j \leq n \quad (8.59)$$

where

$$\langle X_i, X_j \rangle_t = \frac{1}{4} [\langle X_i + X_j, X_i + X_j \rangle_t - \langle X_i - X_j, X_i - X_j \rangle_t], \quad (8.60)$$

$\langle Y, Y \rangle_t$ being the quadratic variation process. (See Exercise 4.7.)

Next we need an auxiliary result about conditional expectation:

LEMMA 8.24. Let μ and ν be two probability measures on a measurable space (Ω, \mathcal{G}) such that $d\nu(\omega) = f(\omega)d\mu(\omega)$ for some $f \in L^1(\mu)$. Let X be a random variable on (Ω, \mathcal{G}) such that

$$E_\mu[|X|] = \int_{\Omega} |X(\omega)|d\mu(\omega) < \infty \text{ and } E_\nu[|X|] = \int_{\Omega} |X(\omega)|f(\omega)d\nu(\omega) < \infty.$$

Let \mathcal{H} be a σ -algebra, $\mathcal{H} \subset \mathcal{F}$. Then

$$E_\nu[X|\mathcal{H}] \cdot E_\mu[f|\mathcal{H}] = E_\mu[fX|\mathcal{H}] \text{ a.s.} \quad (8.61)$$

Proof. By the definition of conditional expectation (Appendix B) we have that if $H \in \mathcal{H}$ then

$$\begin{aligned} \int_H E_\nu[X|\mathcal{H}]fd\mu &= \int_H E_\nu[X|\mathcal{H}]d\nu = \int_H Xd\nu \\ &= \int_H Xfd\mu = \int_H E_\mu[X|\mathcal{H}]d\mu \end{aligned} \quad (8.62)$$

On the other hand, by Theorem B.3 (Appendix B) we have

$$\begin{aligned} \int_H E_\nu[X|\mathcal{H}]fd\mu &= E_\mu[E_\nu[X|\mathcal{H}]f \cdot \mathcal{X}_H] = E_\mu[E_\mu[E_\nu[X|\mathcal{H}]f \cdot \mathcal{X}_H|\mathcal{H}]] \\ &= E_\mu[\mathcal{X}_H E_\nu[X|\mathcal{H}] \cdot E_\mu[f|\mathcal{H}]] = \int_H E_\nu[X|\mathcal{H}] \cdot E_\mu[f|\mathcal{H}]d\mu \end{aligned} \quad (8.63)$$

Combining (8.62) and (8.63) we get

$$\int_N E_\nu[X|\mathcal{H}] \cdot E_\mu[f|\mathcal{H}]d\mu = \int_H E_\mu[fX|\mathcal{H}]d\mu$$

Since this holds for all $H \in \mathcal{H}$, (8.61) follows. \square

We can now prove the first version of the Girsanov formula:

THEOREM 8.25. (The Girsanov theorem I)

Let $Y(t) \in \mathbf{R}^n$ be an Ito process of the form

$$dY(t) = a(t, \omega)dt + dB(t); \quad t \leq T$$

where $T \leq \infty$ is a given constant and $B(t)$ is n-dimensional Brownian motion. Put

$$M_t = \exp\left(-\int_0^t a(s, \omega) dB_s - \frac{1}{2} \int_0^t a^2(s, \omega) ds\right); \quad t \leq T. \quad (8.64)$$

Assume that $a(s, \omega)$ satisfies Novikov's condition

$$E\left[\exp\left(\frac{1}{2} \int_0^T a^2(s, \omega) ds\right)\right] < \infty \quad (8.65)$$

where $E = E_{P^0}$ is the expectation w.r.t. P^0 . Define the measure Q on (Ω, \mathcal{F}_T) by

$$dQ(\omega) = M_T(\omega) dP^0(\omega) \quad (8.66)$$

Then $Y(t)$ is a 1-dimensional Brownian motion w.r.t. the probability law Q , for $t \leq T$.

REMARKS.

- 1) The transformation $P^0 \rightarrow Q$ given by (8.66) is called the *Girsanov transformation of measures*.
- 2) As pointed out in Exercise 4.4 the Novikov condition (8.65) is sufficient to guarantee that $\{M_t\}_{t \leq T}$ is a martingale (w.r.t. \mathcal{F}_t and P^0). This will be used in the proof below.
- 3) Note that since M_t is a martingale we actually have that

$$M_T dP^0 = M_t dP^0 \quad \text{on } \mathcal{F}_t; \quad t \leq T. \quad (8.67)$$

To see this, let f be a bounded \mathcal{F}_t -measurable function. Then by Theorem B.3 we have

$$\begin{aligned} \int_{\Omega} f(\omega) M_T(\omega) dP^0(\omega) &= E[f M_T] = E[E[f M_T | \mathcal{F}_t]] \\ &= E[f E[M_T | \mathcal{F}_t]] = E[f M_t] = \int_{\Omega} f(\omega) M_t(\omega) dP^0(\omega), \end{aligned}$$

where $E = E_{P^0}$ is the expectation w.r.t. P^0 .

Proof of Theorem 8.25. In view of Theorem 8.23 we have to verify that

$$(i) \quad Y(t) = (Y_1(t), \dots, Y_n(t)) \text{ is a martingale w.r.t. } Q \quad (8.68)$$

and

$$(ii) \quad Y_i(t)Y_j(t) - \delta_{ij}t \text{ is a martingale w.r.t. } Q, \quad \text{for all } i, j \in \{1, 2, \dots, n\}. \quad (8.69)$$

To verify (i) we put $K(t) = M_t Y(t)$ and use Ito's formula to get (see Exercise 4.4)

$$\begin{aligned} dK_i(t) &= M_t dY_i(t) + Y_i(t) dM_t + dY_i(t) dM_t \\ &= M_t(a_i(t)dt + dB_i(t)) + Y_i(t)M_t\left(\sum_{k=1}^n -a_k(t)dB_k(t)\right) \\ &\quad + (dB_i(t))(-M_t \sum_{k=1}^n a_k(t)dB_k(t)) \\ &= M_t(dB_i(t) - Y_i(t) \sum_{k=1}^n a_k(t)dB_k(t)) = M_t \gamma^{(i)}(t) dB(t) \end{aligned} \quad (8.70)$$

where $\gamma^{(i)}(t) = (\gamma_1^{(i)}(t), \dots, \gamma_n^{(i)}(t))$, with

$$\gamma_j^{(i)}(t) = \begin{cases} -Y_i(t)a_j(t) & \text{for } j \neq i \\ 1 - Y_i(t)a_i(t) & \text{for } j = i. \end{cases}$$

Hence $K_i(t)$ is a martingale w.r.t. P^0 , so by Lemma 8.24 we get, for $t > s$,

$$\begin{aligned} E_Q[Y_i(t)|\mathcal{F}_s] &= \frac{E[M_t Y_i(t)|\mathcal{F}_s]}{E[M_t|\mathcal{F}_s]} = \frac{E[K_i(t)|\mathcal{F}_s]}{M_s} \\ &= \frac{K_i(s)}{M_s} = Y_i(s), \end{aligned}$$

which shows that $Y_i(t)$ is a martingale w.r.t. Q . This proves (i). The proof of (ii) is similar and is left to the reader. \square

REMARK. Theorem 8.25 states that for all Borel sets $F_1, \dots, F_k \subset \mathbf{R}^n$ and all $t_1, t_2, \dots, t_k \leq T$, $k = 1, 2, \dots$ we have

$$Q[Y(t_1) \in F_1, \dots, Y(t_k) \in F_k] = P^0[B(t_1) \in F_1, \dots, B(t_k) \in F_k] \quad (8.71)$$

An equivalent way of expressing (8.66) is to say that $Q \ll P^0$ (Q is absolutely continuous w.r.t. P^0) with Radon-Nikodym derivative

$$\frac{dQ}{dP^0} = M_T \quad \text{on } \mathcal{F}_T. \quad (8.72)$$

Note that $M_T(\omega) > 0$ a.s., so we also have that $P^0 \ll Q$. Hence the two measures Q and P^0 are equivalent. Therefore we get from (8.71)

$$\begin{aligned} P^0[Y(t_1) \in F_1, \dots, Y(t_k) \in F_k] &> 0 \\ \iff Q[Y(t_1) \in F_1, \dots, Y(t_k) \in F_k] &> 0 \\ \iff P^0[B(t_1) \in F_1, \dots, B(t_k) \in F_k] &> 0; \quad t_1, \dots, t_k \in [0, T] \end{aligned} \quad (8.73)$$

THEOREM 8.26. (The Girsanov theorem II)

Let $Y(t) \in \mathbf{R}^n$ be an Ito process of the form

$$dY(t) = \beta(t, \omega) + \theta(t, \omega)dB(t); \quad t \leq T \quad (8.74)$$

where $B(t) \in \mathbf{R}^m$, $\beta(t, \omega) \in \mathbf{R}^n$ and $\theta(t, \omega) \in \mathbf{R}^{n \times m}$. Suppose there exist $\mathcal{V}[0, T]$ -processes $u(t, \omega) \in \mathbf{R}^m$ and $\alpha(t, \omega) \in \mathbf{R}^n$ such that

$$\theta(t, \omega)u(t, \omega) = \beta(t, \omega) - \alpha(t, \omega) \quad (8.75)$$

and assume that $u(t, \omega)$ satisfies Novikov's condition

$$E\left[\exp\left(\frac{1}{2} \int_0^T u^2(s, \omega)ds\right)\right] < \infty. \quad (8.76)$$

Put

$$M_t = \exp\left(-\int_0^t u(s, \omega)dB_s - \frac{1}{2} \int_0^t u^2(s, \omega)ds\right); \quad t \leq T \quad (8.77)$$

and

$$dQ(\omega) = M_T(\omega)dP^0(\omega) \quad \text{on } \mathcal{F}_T. \quad (8.78)$$

Then

$$\hat{B}(t) := \int_0^t u(s, \omega)ds + B(t); \quad t \leq T \quad (8.79)$$

is a Brownian motion w.r.t. Q and in terms of $\hat{B}(t)$ the process $Y(t)$ has the stochastic integral representation

$$dY(t) = \alpha(t, \omega) + \theta(t, \omega)d\hat{B}(t). \quad (8.80)$$

Proof. It follows from Theorem 8.25 that $\hat{B}(t)$ is a Brownian motion w.r.t. Q . So, substituting (8.79) in (8.74) we get, by (8.75),

$$\begin{aligned} dY(t) &= \beta(t, \omega) + \theta(t, \omega)(d\hat{B}(t) - u(t, \omega)dt) \\ &= [\beta(t, \omega) - \theta(t, \omega)u(t, \omega)]dt + \theta(t, \omega)d\hat{B}(t) \\ &= \alpha(t, \omega)dt + \theta(t, \omega)d\hat{B}(t). \end{aligned}$$

□

Note that if $n = m$ and $\theta \in \mathbf{R}^{n \times n}$ is invertible, then the process $u(t, \omega)$ satisfying (8.75) is given uniquely by

$$u(t, \omega) = \theta^{-1}(t, \omega)[\beta(t, \omega) - \alpha(t, \omega)]. \quad (8.81)$$

Finally we formulate a diffusion version:

THEOREM 8.27. (The Girsanov theorem III)

Let $X(t) = X^x(t) \in \mathbf{R}^n$ and $Y(t) \in \mathbf{R}^n$ be an Ito diffusion and an Ito process, respectively, of the forms

$$dX(t) = b(X(t))dt + \sigma(X(t))dB(t); \quad t \leq T, \quad X(0) = x \quad (8.82)$$

$$dY(t) = [\gamma(t, \omega) + b(Y(t))]dt + \sigma(Y(t))dB(t); \quad t \leq T, \quad Y(0) = x \quad (8.83)$$

where the functions $b : \mathbf{R}^n \rightarrow \mathbf{R}^n$ and $\sigma : \mathbf{R}^n \times \mathbf{R}^{n \times m}$ satisfy the conditions of Theorem 5.5 and $\gamma(t, \omega) \in \mathcal{V}[0, T]$, $x \in \mathbf{R}^n$. Suppose there exists a $\mathcal{V}[0, T]$ -process $u(t, \omega) \in \mathbf{R}^m$ such that

$$\sigma(Y(t))u(t, \omega) = \gamma(t, \omega) \quad (8.84)$$

and assume that $u(t, \omega)$ satisfies Novikov's condition

$$E\left[\exp\left(\frac{1}{2} \int_0^T u^2(s, \omega)ds\right)\right] < \infty \quad (8.85)$$

Define M_t , Q and $\hat{B}(t)$ as in (8.77), (8.78) and (8.79). Then

$$dY(t) = b(Y(t))dt + \sigma(Y(t))d\hat{B}(t). \quad (8.86)$$

Therefore,

$$\begin{aligned} &\text{the } Q\text{-law of } Y(t) \text{ is the same as} \\ &\text{the } P^0\text{-law of } X^x(t); \quad t \leq T. \end{aligned} \quad (8.87)$$

Proof. The representation (8.86) follows by applying Theorem 8.26 to the case $\theta(t, \omega) = \sigma(Y(t))$, $\beta(t, \omega) = \gamma(t, \omega) + b(Y(t))$, $\alpha(t, \omega) = b(Y(t))$. Then the conclusion (8.87) follows from the weak uniqueness of solutions of stochastic differential equations (see (5.28)). \square

The Girsanov theorem III can be used to produce weak solutions of stochastic differential equations. To illustrate this, suppose Y_t is a known weak or strong solution to the equation

$$dY_t = b(Y_t)dt + \sigma(Y_t)dB(t)$$

where $b : \mathbf{R}^n \rightarrow \mathbf{R}^n$, $\sigma : \mathbf{R}^n \rightarrow \mathbf{R}^{n \times m}$ and $B(t) \in \mathbf{R}^m$. We wish to find a weak solution $X(t)$ of a related equation

$$dX_t = a(X_t)dt + \sigma(X_t)dB(t) \quad (8.88)$$

where the drift function is changed to $a : \mathbf{R}^n \rightarrow \mathbf{R}^n$. Suppose we can find a function $u_0 : \mathbf{R}^n \rightarrow \mathbf{R}^m$ such that

$$\sigma(y)u_0(y) = b(y) - a(y); \quad y \in \mathbf{R}^n.$$

(If $n = m$ and σ is invertible we choose

$$u_0 = \sigma^{-1} \cdot (b - a).$$

Then if $u(t, \omega) = u_0(Y_t(\omega))$ satisfies Novikov's conditions, we have, with Q and $\hat{B}_t = \hat{B}(t)$ as in (8.78) and (8.79), that

$$dY_t = a(Y_t)dt + \sigma(Y_t)d\hat{B}_t. \quad (8.89)$$

Thus we have found a Brownian motion (\hat{B}_t, Q) such that Y_t satisfies (8.89). Therefore (Y_t, \hat{B}_t) is a weak solution of (8.88).

EXAMPLE 8.28. Let $a : \mathbf{R}^n \rightarrow \mathbf{R}^n$ be a bounded, measurable function. Then we can construct a weak solution $X_t = X_t^x$ of the stochastic differential equation

$$dX_t = a(X_t)dt + dB_t; \quad X_0 = x \in \mathbf{R}^n. \quad (8.90)$$

We proceed according to the procedure above, with $\sigma = I$, $b = 0$ and

$$dY_t = dB_t; \quad Y_0 = x.$$

Choose

$$u_0 = \sigma^{-1} \cdot (b - a) = -a$$

and define

$$M_t = \exp\left\{-\int_0^t u_0(Y_s)dB_s - \frac{1}{2} \int_0^t u_0^2(Y_s)ds\right\}$$

i.e.

$$M_t = \exp\left\{\int_0^t a(B_s)dB_s - \frac{1}{2} \int_0^t a^2(B_s)ds\right\}.$$

Fix $T < \infty$ and put

$$dQ = M_T dP \quad \text{on } \mathcal{F}_T.$$

Then

$$\hat{B}_t := - \int_0^t a(B_s)ds + B_t$$

is a Brownian motion w.r.t. Q for $t \leq T$ and

$$dB_t = dY_t = a(Y_t)dt + d\hat{B}_t$$

Hence if we set $Y_0 = x$ the pair (Y_t, \hat{B}_t) is a weak solution of (8.90) for $t \leq T$. By weak uniqueness (see the remarks following (8.23) above) the Q -law of $Y_t = B_t$ coincides with the P^0 -law of X_t^x , so that

$$\begin{aligned} E[f_1(X_{t_1}^x) \dots f_k(X_{t_k}^x)] &= E_Q[f_1(Y_{t_1}) \dots f_k(Y_{t_k})] \\ &= E[M_T f_1(B_{t_1}) \dots f_k(B_{t_k})] \end{aligned} \quad (8.91)$$

for all $f_1, \dots, f_k \in C_0(\mathbf{R}^n)$; $t_1, \dots, t_k \leq T$.

EXERCISES

8.1.

Let Δ denote the Laplace operator on \mathbf{R}^n .

- a) Write down (in terms of Brownian motion) a bounded solution g of the Cauchy problem

$$\begin{cases} \frac{\partial g(t, x)}{\partial t} - \frac{1}{2} \Delta_x g(t, x) = 0 & \text{for } t > 0, x \in \mathbf{R}^n \\ g(0, x) = \phi(x) \end{cases}$$

where $\phi \in C_0^2$ is given. (From general theory it is known that the solution is unique.)

- b) Let $\psi \in C_b(\mathbf{R}^n)$ and $\alpha > 0$. Find a bounded solution u of the equation

$$(\alpha - \frac{1}{2} \Delta) u = \psi \quad \text{in } \mathbf{R}^n.$$

Prove that the solution is unique.

8.2.

Show that the solution $u(t, x)$ of the initial value problem

$$\begin{aligned} \frac{\partial u}{\partial t} &= \frac{1}{2} \beta^2 x^2 \frac{\partial^2 u}{\partial x^2} + \alpha x \frac{\partial u}{\partial x}; \quad t > 0, x \in \mathbf{R} \\ u(0, x) &= f(x) \quad (f \in C_0^2(\mathbf{R}) \text{ given}) \end{aligned}$$

can be expressed as follows:

$$\begin{aligned} u(t, x) &= E[f(x \cdot \exp\{\beta B_t + (\alpha - \frac{1}{2} \beta^2)t\})] \\ &= \frac{1}{2\pi t} \int_{\mathbf{R}} f(x \cdot \exp\{\beta y + (\alpha - \frac{1}{2} \beta^2)t\} \exp(-\frac{y^2}{2t})) dy; \quad t > 0 \end{aligned}$$

8.3. (Kolmogorov's forward equation)

Let X_t be an Ito diffusion in \mathbf{R}^n with generator

$$Af(y) = \sum_{i,j} a_{ij}(y) \frac{\partial^2 f}{\partial y_i \partial y_j} + \sum_i b_i(y) \frac{\partial f}{\partial y_i}; \quad f \in C_0^2$$

and assume that the transition measure of X_t has a density $p_t(x, y)$, i.e. that

$$E^x[f(X_t)] = \int_{\mathbf{R}^n} f(y) p_t(x, y) dy; \quad f \in C_0^2 \quad (8.92)$$

Assume that $y \rightarrow p_t(x, y)$ is smooth for each t, x . Prove that $p_t(x, y)$ satisfies the *Kolmogorov forward equation*

$$\frac{d}{dt} p_t(x, y) = A_y^* p_t(x, y) \quad \text{for all } x, y, \quad (8.93)$$

where A_y^* operates on the variable y and is given by

$$A_y^* \phi(y) = \sum_{i,j} \frac{\partial^2}{\partial y_i \partial y_j} (a_{ij} \phi) - \sum_i \frac{\partial}{\partial y_i} (b_i \phi); \quad \phi \in C^2 \quad (8.94)$$

(i.e. A_y^* is the *adjoint* of A_y)

(Hint: By (8.92) and Dynkin's formula we have

$$\int_{\mathbf{R}^n} f(y) p_t(x, y) dy = f(x) + \int_0^t \int_{\mathbf{R}^n} A_y f(y) p_s(x, y) dy ds; \quad f \in C_0^2$$

Now differentiate w.r.t. t and use that

$$\langle A\phi, \psi \rangle = \langle \phi, A^* \psi \rangle \quad \text{for } \phi \in C_0^2, \psi \in C^2,$$

where $\langle \cdot, \cdot \rangle$ denotes inner product in $L^2(dy)$)

8.4.

Let B_t be n-dimensional Brownian motion ($n \geq 1$) and let F be a Borel set in \mathbf{R}^n . Prove that the expected total length of time t that B_t stays in F is zero if and only if the Lebesgue measure of F is zero.

Hint: Consider the resolvent R_α for $\alpha > 0$ and then let $\alpha \rightarrow 0$.

8.5.

Show that the solution $u(t, x)$ of the initial value problem

$$\begin{cases} \frac{\partial u}{\partial t} = \rho u + \frac{1}{2} \Delta u & t > 0; x \in \mathbf{R}^n \\ u(0, x) = f(x) & (f \in C_0^2(\mathbf{R}^n) \text{ given}) \end{cases}$$

(where $\rho \in \mathbf{R}$ is a constant) can be expressed by

$$u(t, x) = (2\pi)^{-n/2} \exp(\rho t) \int_{\mathbf{R}^n} f(y) \exp\left(-\frac{(x-y)^2}{2t}\right) dy.$$

8.6.

In connection with the deduction of the Black & Scholes formula for the price of an option (see e.g. Duffie (1992)) the following partial differential equation appears:

$$\begin{cases} \frac{\partial u}{\partial t} = -\rho u + \alpha x \frac{\partial u}{\partial x} + \frac{1}{2} \beta^2 x^2 \frac{\partial^2 u}{\partial x^2}; & t > 0, x \in \mathbf{R} \\ u(0, x) = (x - K)^+; & x \in \mathbf{R}, \end{cases}$$

where $\rho > 0$, α, β and $K > 0$ are constants and

$$(x - K)^+ = \max(x - K, 0).$$

Use the Feynman-Kac formula to prove that the solution u of this equation is given by

$$u(t, x) = \frac{e^{-\rho t}}{\sqrt{2\pi t}} \int_{\mathbf{R}} (x \cdot \exp\{(\alpha - \frac{1}{2}\beta^2)t + \beta y\} - K)^+ e^{-\frac{y^2}{2t}} dy; \quad t > 0.$$

(This expression can be simplified further. See e.g. Duffie (1992).)

8.7.

Let X_t be a sum of Ito integrals of the form

$$X_t = \sum_{k=1}^n \int_0^t v_k(s, \omega) dB_k(s),$$

where (B_1, \dots, B_n) is n-dimensional Brownian motion. Assume that

$$\beta_t := \int_0^t \sum_{k=1}^n v_k^2(s, \omega) ds \rightarrow \infty \text{ as } t \rightarrow \infty, \text{ a.s.}$$

Prove that

$$\limsup_{t \rightarrow \infty} \frac{X_t}{\sqrt{2\beta_t \log \log \beta_t}} = 1 \quad \text{a.s.}$$

(Hint: Use the law of iterated logarithm)

8.8.

Let Z_t be a 1-dimensional Ito process of the form

$$dZ_t = u(t, \omega) dt + dB_t$$

Let \mathcal{G}_t be the σ -algebra generated by $\{Z_s(\cdot); s \leq t\}$ and define

$$dN_t = (u(t, \omega) - E[u|\mathcal{G}_t])dt + dB_t$$

Use Corollary 8.13 to prove that N_t is a Brownian motion. (If we interpret Z_t as the *observation* process, then N_t is the *innovation* process. See Lemma 6.8)

8.9.

Define $\alpha(t) = \frac{1}{2} \ln(1 + \frac{2}{3}t^3)$. If B_t is a Brownian motion, prove that there exists another Brownian motion \tilde{B}_r such that

$$\int_0^{\alpha_t} e^s dB_s = \int_0^t r d\tilde{B}_r.$$

8.10.

Let B_t be a Brownian motion in \mathbf{R} . Show that

$$X_t := B_t^2$$

is a weak solution of the stochastic differential equation

$$dX_t = dt + 2\sqrt{|X_t|}d\tilde{B}_t. \quad (8.95)$$

(Hint: Use Ito's formula to express X_t as a stochastic integral and compare with (8.95) by using Corollary 8.13)

8.11.

There exist stochastic differential equations which have weak but no strong solutions. One such example is the 1-dimensional equation of Tanaka:

$$dX_t = \text{sign}(X_t)dB_t \quad (8.96)$$

where

$$\text{sign}(u) = \begin{cases} 1 & \text{if } u > 0 \\ -1 & \text{if } u \leq 0. \end{cases}$$

- a) Find a weak solution (X_t, \tilde{B}_t) of (8.96). (Hint: Choose X_t to be a Brownian motion. Then prove that

$$\tilde{B}_t := \int_0^t \text{sign}(X_s) dX_s$$

is also a Brownian motion, for example by using Theorem 8.10.)

- b) Prove that (8.96) has no strong solutions. (Hint: If X_t is a strong solution then by (8.96) we have

$$dB_t = \text{sign}(X_t) dX_t,$$

so from the discussion of Tanaka's formula (8.36) we conclude that the σ -algebra \mathcal{F}_t generated by $\{B_s; s \leq t\}$ is contained in the σ -algebra \mathcal{H}_t generated by $\{|X_s|; s \leq t\}$. Therefore $\text{sign}(X_t)$ - and hence X_t - cannot be \mathcal{F}_t -adapted).

8.12.

Let $b : \mathbf{R} \rightarrow \mathbf{R}$ be a Lipschitz-continuous function and define $X_t = X_t^x \in \mathbf{R}$ by

$$dX_t = b(X_t)dt + dB_t, X_0 = x \in \mathbf{R}$$

- a) Use the Girsanov theorem to prove that for all $M < \infty$, $x \in \mathbf{R}$ and $t > 0$ we have

$$P[X_t^x \geq M] > 0$$

- b) Choose $b(x) = -r$ where $r > 0$ is constant. Prove that for all x

$$X_t^x \rightarrow -\infty \text{ as } t \rightarrow \infty \text{ a.s.}$$

Compare this with the result in a).

8.13. (Polar sets for the graph of Brownian motion)

Let B_t be 1-dimensional Brownian motion starting at $x \in \mathbf{R}$.

- a) Prove that for every fixed time $t_0 > 0$ we have

$$P^x[B_{t_0} = 0] = 0$$

- b) Prove that for every (non-trivial) closed interval $J \subset \mathbf{R}^+$ we have

$$P^x[\exists t \in J \text{ such that } B_t = 0] > 0$$

(Hint: If $J = [t_1, t_2]$ consider $P^x[B_{t_1} < 0 \& B_{t_2} > 0]$ and then use the intermediate value theorem.)

- c) In view of a) and b) it is natural to ask what closed sets $F \subset \mathbf{R}^+$ have the property that

$$P^x[\exists t \in F \text{ such that } B_t = 0] = 0 \quad (8.97)$$

To investigate this question more closely we introduce the graph X_t of Brownian motion, given by

$$dX_t = \begin{bmatrix} 1 \\ 0 \end{bmatrix} dt + \begin{bmatrix} 0 \\ 1 \end{bmatrix} dB_t; \quad X_0 = \begin{bmatrix} t_0 \\ x_0 \end{bmatrix}$$

i.e.

$$X_t = X_t^{t_0, x_0} = \begin{bmatrix} t_0 + t \\ B_t^{x_0} \end{bmatrix} \text{ where } B_0^{x_0} = x_0 \text{ a.s.}$$

Then F satisfies (8.97) iff $K := F \times \{0\}$ polar for X_t , in the sense that

$$P^{t_0, x_0}[\exists t > 0; X_t \in K] = 0 \quad \text{for all } t_0, x_0 \quad (8.98)$$

The key to finding polar sets for a diffusion is to consider its *Green operator* R , which is simply the resolvent R_α with $\alpha = 0$:

$$Rf(t_0, x_0) = E^{t_0, x_0} \left[\int_{t_0}^{\infty} f(X_s) ds \right] \quad \text{for } f \in C_0(\mathbf{R}^2)$$

Show that

$$Rf(t_0, x_0) = E^{t_0, x_0} \left[\int_{\mathbf{R}^2} G(t_0, x_0; t, x) f(t, x) dt dx \right],$$

where

$$G(t_0, x_0; t, x) = \mathcal{X}_{t > t_0} \cdot (2\pi(t - t_0))^{-\frac{1}{2}} \exp\left(-\frac{|x - x_0|^2}{2(t - t_0)}\right) \quad (8.99)$$

(G is the *Green function* of X_t)

- d) The *capacity* of K , $C(K) = C_G(K)$, is defined by

$$C(K) = \sup\{\mu(K); \mu \in M_G(K)\},$$

where $M_G(K) = \{\mu; \mu \text{ measure on } K \text{ s.t. } \int_K G(t_0, x_0; t, x) d\mu(t, x) \leq 1 \text{ for all } t_0, x_0\}$

A general result from stochastic potential theory states that

$$P^{t_0, x_0}[X_t \text{ hits } K] = 0 \Leftrightarrow C(K) = 0 \quad (8.100)$$

See e.g. Blumenthal and Getoor (1968, Prop. VI.4.3). Use this to prove that

$$\Lambda_{\frac{1}{2}}(F) = 0 \Rightarrow P^{x_0}[\exists t \in F \text{ such that } B_t = 0] = 0,$$

where $\Lambda_{\frac{1}{2}}$ denotes 1/2-dimensional Hausdorff measure (Folland (1984, §10.2)).

8.14.

Let $f \in C_0^2(\mathbf{R}^n)$ and $\alpha(x) = (\alpha_1(x), \dots, \alpha_n(x))$ with $\alpha_i \in C_0^2(\mathbf{R}^n)$ be given functions and consider the partial differential equation

$$\begin{cases} \frac{\partial u}{\partial t} = \sum_{i=1}^n \alpha_i(x) \frac{\partial u}{\partial x_i} + \frac{1}{2} \sum_{i=1}^n \frac{\partial^2 u}{\partial x_i^2}; & t > 0, x \in \mathbf{R}^n \\ u(0, x) = f(x); & x \in \mathbf{R}^n \end{cases}$$

- a) Use the Girsanov theorem to show that the unique bounded solution $u(t, x)$ of this equation can be expressed by

$$u(t, x) = E^x \left[\exp \left(\int_0^t \alpha(B_s) dB_s - \frac{1}{2} \int_0^t \alpha^2(B_s) ds \right) f(B_t) \right],$$

where $E = E_{P^x}$ is the expectation w.r.t. P^x .

- b) Now assume that α is a gradient, i.e. that there exists $\gamma \in C^1(\mathbf{R}^n)$ such that

$$\nabla \gamma = \alpha.$$

Assume for simplicity that $\gamma \in C_0^2(\mathbf{R}^n)$. Use Ito's formula to prove that (see Exercise 4.8)

$$\begin{aligned} u(t, x) = \exp(-\gamma(x)) E^x & \left[\exp \left\{ -\frac{1}{2} \int_0^t (\nabla \gamma^2(B_s) \right. \right. \\ & \left. \left. + \Delta \gamma(B_s)) ds \right\} \exp(\gamma(B_t)) f(B_t) \right]. \end{aligned}$$

- c) Put $v(t, x) = \exp(\gamma(x)) u(t, x)$. Use the Feynmac-Kac formula to show that $v(t, x)$ satisfies the partial differential equation

$$\begin{cases} \frac{\partial v}{\partial t} = -\frac{1}{2} (\nabla \gamma^2 + \Delta \gamma) \cdot v + \frac{1}{2} \Delta v; & t > 0; x \in \mathbf{R}^n \\ v(0, x) = \exp(\gamma(x)) f(x); & x \in \mathbf{R}^n \end{cases}$$

(See also Exercise 8.15.)

8.15. (A connection between B.m. with drift and killed B.m.)

Let B_t denote Brownian motion in \mathbf{R}^n and consider the diffusion X_t in \mathbf{R}^n defined by

$$dX_t = \nabla h(X_t) dt + dB_t; \quad X_0 = x \in \mathbf{R}^n. \quad (8.101)$$

where $h \in C_0^1(\mathbf{R}^n)$.

- a) There is an important connection between this process and the process Y_t obtained by *killing* B_t at a certain rate V . More precisely, first prove that for $f \in C_0(\mathbf{R}^n)$ we have

$$E^x[f(X_t)] = E^x \left[\exp \left(- \int_0^t V(B_s) ds \right) \cdot \exp(h(B_t) - h(x)) \cdot f(B_t) \right], \quad (8.102)$$

where

$$V(x) = \frac{1}{2} |\nabla h(x)|^2 + \frac{1}{2} \Delta h(x). \quad (8.103)$$

(Hint: Use the Girsanov theorem to express the left hand side of (8.102) in terms of B_t . Then use the Ito formula on $Z_t = h(B_t)$ to achieve (8.102).)

- b) Then use the Feynman-Kac formula to restate (8.102) as follows (assuming $V \geq 0$):

$$T_t^X(f, x) = \exp(-h(x)) \cdot T_t^Y(f \cdot \exp h, x),$$

where T_t^X, T_t^Y denote the *transition operators* of the processes X and Y , respectively, i.e.

$$T_t^X(f, x) = E^x[f(X_t)] \quad \text{and similarly for } Y.$$

IX. Applications to Boundary Value Problems

A) The Dirichlet problem.

We now use the preceding results to solve the following generalization of the Dirichlet problem stated in the introduction:

Given a domain $D \subset \mathbf{R}^n$, a semi-elliptic partial differential operator L on $C^2(\mathbf{R}^n)$, of the form

$$L = \sum_{i,j=1}^n a_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^n b_i(x) \frac{\partial}{\partial x_i}$$

and a bounded continuous function ϕ on the boundary ∂D of D , find a function $\tilde{\phi}$ on D ("the L -harmonic extension of ϕ ") such that

- (i) $L\tilde{\phi} = 0$ in D
 - (ii) $\lim_{\substack{x \rightarrow y \\ x \in D}} \tilde{\phi}(x) = \phi(y)$ for all "reasonable" (i.e. regular) $y \in \partial D$.
- (9.1)

(L is called *semi-elliptic (elliptic)* when all the eigenvalues of the symmetric matrix $a_{ij}(x)$ are non-negative (positive), for all x). We will define regular points later. With this definition the Dirichlet problem turns out to have a (unique) solution under suitable conditions on L and D (See Theorems 9.11, 9.12 and 9.14).

The idea is the following: First we find an Ito diffusion $\{X_t\}$ whose generator A coincides with L on $C_0^2(\mathbf{R}^n)$. To obtain this we simply choose $\sigma \in \mathbf{R}^{n \times n}$ such that

$$\frac{1}{2}\sigma\sigma^T = [a_{ij}]. \quad (9.2)$$

We assume that σ and $b = [b_i]$ satisfy conditions (5.14), (5.15) of Theorem 5.5. (For example, if each $a_{ij} \in C^2$ is bounded and has bounded first and second partial derivatives, then such a square root σ can be found. See Fleming and Rishel (1975).) Next we let X_t be the solution of

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t, \quad (9.3)$$

where B_t is n-dimensional Brownian motion.

Then our candidate for the solution $\tilde{\phi}$ is

$$\tilde{\phi}(x) = E^x[\phi(X_{\tau_D})]. \quad (9.4)$$

The first problem we meet is that (9.4) only makes sense if

$$Q^x[\tau_D < \infty] = 1 \text{ for all } x \in D. \quad (9.5)$$

From now on we assume that (9.5) holds.

For example, if L is elliptic in \bar{D} and D is bounded, then this assumption is satisfied (Dynkin (1965 II) p. 43). We also see that it is satisfied for the process X_t associated to the parabolic L in Example 7.11. In Dynkin (1965 II) p. 34–35 it is proved that (9.5) is a consequence of the weaker condition

$$Q^x[\tau_D < \infty] > 0 \text{ for all } x \in D. \quad (9.6)$$

Next we define what we mean by *regular points*. First we will need the following useful results: (As before we let \mathcal{M}_t and \mathcal{M} denote the σ -algebras generated by X_s ; $s \leq t$ and by X_s ; $s \geq 0$ respectively).

LEMMA 9.1. (The 0-1 law) Let $H \in \bigcap_{t>0} \mathcal{M}_t$. Then either $Q^x(H) = 0$ or $Q^x(H) = 1$.

Proof. From the strong Markov property (7.15) we have

$$E^x[\theta_t \eta | \mathcal{M}_t] = E^{X_t}[\eta]$$

for all bounded, \mathcal{M} -measurable $\eta : \Omega \rightarrow \mathbf{R}$. This implies that

$$\int_H \theta_t \eta \cdot dQ^x = \int_H E^{X_t}[\eta] dQ^x, \text{ for all } t.$$

First assume that $\eta = \eta_k = g_1(X_{t_1}) \cdots g_k(X_{t_k})$, where each g_i is bounded and continuous. Then letting $t \rightarrow 0$ we obtain

$$\int_H \eta dQ^x = \lim_{t \rightarrow 0} \int_H \theta_t \eta dQ^x = \lim_{t \rightarrow 0} \int_H E^{X_t}[\eta] dQ^x = Q^x(H) E^x[\eta]$$

by Feller continuity (Lemma 8.4) and bounded convergence. Approximating the general η by functions η_k as above we conclude that

$$\int_H \eta dQ^x = Q^x(H) E^x[\eta]$$

for all bounded \mathcal{M} -measurable η . If we put $\eta = \chi_H$ we obtain $Q^x(H) = (Q^x(H))^2$, which completes the proof.

COROLLARY 9.2. Let $y \in \mathbf{R}^n$. Then

$$\text{either } Q^y[\tau_D = 0] = 0 \quad \text{or} \quad Q^y[\tau_D = 0] = 1.$$

Proof. $H = \{\omega; \tau_D = 0\} \in \bigcap_{t>0} \mathcal{M}_t$.

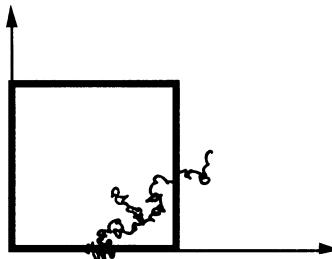
In other words, either a.a. paths X_t starting from y stay within D for a positive period of time or a.a. paths X_t starting from y leave D immediately. In the last case we call the point y *regular*, i.e.

DEFINITION 9.3. A point $y \in \partial D$ is called *regular* for D (w.r.t. X_t) if

$$Q^y[\tau_D = 0] = 1.$$

Otherwise the point y is called *irregular*.

EXAMPLE 9.4. Corollary 9.2 may seem hard to believe at first glance. For example, if X_t is a 2-dimensional Brownian motion B_t and \bar{D} is the square $[0, 1] \times [0, 1]$ one might think that, starting from $(\frac{1}{2}, 0)$, say, half of the paths will stay in the upper half plane and half in the lower, for a positive period of time. However, Corollary 9.2 says that this is not the case: Either they all stay in D initially or they all leave D immediately. Symmetry considerations imply that the first alternative is impossible. Thus $(\frac{1}{2}, 0)$, and similarly all the other points of ∂D , are regular for D w.r.t. B_t .



EXAMPLE 9.5. Let $D = [0, 1] \times [0, 1]$ and let L be the parabolic differential operator

$$Lf(t, x) = \frac{\partial f}{\partial t} + \frac{1}{2} \cdot \frac{\partial^2 f}{\partial x^2}; \quad (t, x) \in \mathbf{R}^2.$$

(See Example 7.11)

Here

$$b = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad a = [a_{ij}] = \frac{1}{2} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

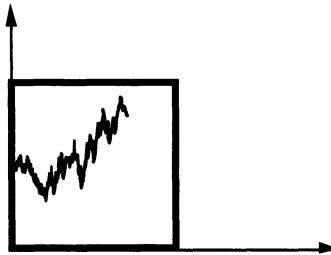
So, for example, if we choose $\sigma = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$, we have $\frac{1}{2}\sigma\sigma^T = a$. This gives the following stochastic differential equation for the Ito diffusion X_t associated with L :

$$dX_t = \begin{pmatrix} 1 \\ 0 \end{pmatrix} dt + \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} dB_t^{(1)} \\ dB_t^{(2)} \end{pmatrix}.$$

In other words,

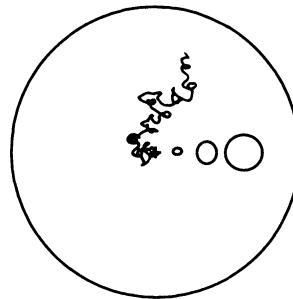
$$X_t = \begin{pmatrix} t + t_0 \\ B_t \end{pmatrix}, \quad X_0 = \begin{pmatrix} t_0 \\ a \end{pmatrix}$$

where B_t is 1-dimensional Brownian motion. So we end up with the graph of Brownian motion, which we started with in Example 7.11. In the case it is not hard to see that the irregular points of ∂D consist of the open line $\{0\} \times (0, 1)$, the rest of the boundary points being regular.



EXAMPLE 9.6. Let $\Delta = \{(x, y); x^2 + y^2 < 1\} \subset \mathbf{R}^2$ and let $\{\Delta_n\}$ be a sequence of disjoint open discs in Δ centered at $(2^{-n}, 0)$, respectively, $n = 1, 2, \dots$. Put

$$D = \Delta \setminus \overline{\left(\bigcup_{n=1}^{\infty} \Delta_n \right)}.$$



Then it is easy to see that all the points of $\partial\Delta \cup \bigcup_{n=1}^{\infty} \partial\Delta_n$ are regular for D w.r.t. 2-dimensional Brownian motion B_t , using a similar argument as in Example 9.4. But what about the point 0? The answer depends on the sizes of the discs

Δ_n . More precisely, if r_n is the radius of Δ_n then 0 is a regular point for D if and only if

$$\sum_{n=1}^{\infty} \frac{n}{\log \frac{1}{r_n}} = \infty.$$

(This is a consequence of the famous Wiener criterion. See Port and Stone (1979), p. 225). Having defined regular points we now return to the Dirichlet problem and would like to conclude:

- (i) $L\tilde{\phi} = 0$ in D
- (ii) $\lim_{\substack{x \rightarrow y \\ x \in D}} \tilde{\phi}(x) = \phi(y)$ for all regular $y \in \partial D$.

Unfortunately, this does not hold in general. Consider the following simple example:

EXAMPLE 9.7. Let $X(t) = (X_1(t), X_2(t))$ be the solution of the equations

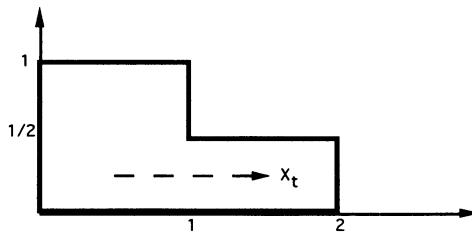
$$\begin{aligned} dX_1(t) &= dt \\ dX_2(t) &= 0 \end{aligned}$$

so that $X(t) = X(0) + t(1, 0) \in \mathbf{R}^2$; $t \geq 0$. Let

$$D = ((0, 1) \times (0, 1)) \cup ((0, 2) \times (0, \frac{1}{2}))$$

and let ϕ be a continuous function on ∂D such that

$$\begin{aligned} \phi &= 1 \text{ on } \{1\} \times [\frac{1}{2}, 1] \quad \text{and} \\ \phi &= 0 \text{ on } \{2\} \times [0, \frac{1}{2}]. \end{aligned}$$



Then

$$\tilde{\phi}(t, x) = E^{t,x}[\phi(X_{\tau_D})] = \begin{cases} 1 & \text{if } x \in (\frac{1}{2}, 1) \\ 0 & \text{if } x \in (0, \frac{1}{2}), \end{cases}$$

so $\tilde{\phi}$ is not even continuous.

The point $y = (1, \frac{1}{2})$ is regular for X_t , but

$$\lim_{x \uparrow \frac{1}{2}} \tilde{\phi}(1, x) = 0 \neq \phi(1, \frac{1}{2}).$$

However, it is possible to formulate a weak, stochastic version of this problem. This stochastic version will always have a solution, which coincides with the (unique) solution of the original problem (9.1) if case such a solution exists. In the stochastic version we will replace (i) and (ii) in (9.1) by weaker requirements $(i)_s$ and $(ii)_s$. A natural candidate for the requirement $(i)_s$ would be that $\mathcal{A}\tilde{\phi} = 0$ in D . We will prove a slightly stronger property, which also makes it easy to obtain uniqueness: We will prove that $\tilde{\phi}$ is X -harmonic, in the sense that $\tilde{\phi}$ satisfies the mean value property w.r.t. X (see (7.19)):

DEFINITION 9.8. Let f be a locally bounded, measurable function on D . Then f is called X -harmonic in D if

$$f(x) = E^x[f(X_{\tau_U})]$$

for all $x \in D$ and all bounded open sets U with $\overline{U} \subset D$. We make two important observations:

LEMMA 9.9.

- a) Let f be X -harmonic in D . Then $\mathcal{A}f = 0$ in D .
- b) Conversely, suppose $f \in C^2(D)$ and $\mathcal{A}f = 0$ in D . Then f is X -harmonic.

Proof.

- a) follows directly from the formula for \mathcal{A} .
- b) follows from the Dynkin formula: Choose U as Definition 9.8. Then

$$\begin{aligned} E^x[f(X_{\tau_U})] &= \lim_{k \rightarrow \infty} E^x[f(X_{\tau_U \wedge k})] \\ &= f(x) + \lim_{k \rightarrow \infty} E^x \left[\int_0^{\tau_U \wedge k} (Lf)(X_s) ds \right] = f(x), \end{aligned}$$

since $Lf = \mathcal{A}f = 0$ in U .

The most important examples of X -harmonic functions are given in the next result:

LEMMA 9.10. Let ϕ be a bounded measurable function on ∂D and put

$$\tilde{\phi}(x) = E^x[\phi(X_{\tau_D})]; x \in D.$$

Then $\tilde{\phi}$ is X -harmonic. Thus, in particular, $\mathcal{A}\tilde{\phi} = 0$.

Proof. From (7.19) we have, if $\overline{V} \subset D$

$$\tilde{\phi}(x) = \int_{\partial V} \tilde{\phi}(y) Q^x[X_{\tau_V} \in dy] = E^x[\tilde{\phi}(X_{\tau_V})].$$

We are now ready to formulate the weak, stochastic version:

The stochastic Dirichlet problem.

Given a bounded measurable function ϕ on ∂D , find a function $\tilde{\phi}$ on D such that

- (i)_s $\tilde{\phi}$ is X -harmonic
- (ii)_s $\lim_{t \uparrow \tau_D} \tilde{\phi}(X_t) = \phi(X_{\tau_D})$ a.s. $Q^x, x \in D$.

We first solve the stochastic Dirichlet problem (9.7) and then relate it to the original problem (9.1).

THEOREM 9.11. (Solution of the stochastic Dirichlet problem)

Let ϕ be a bounded measurable function on ∂D .

a) (Existence) Define

$$\tilde{\phi}(x) = E^x[\phi(X_{\tau_D})] \quad (9.8)$$

Then $\tilde{\phi}$ solves the stochastic Dirichlet problem (9.7)

b) (Uniqueness) Suppose g is a bounded function on D such that

- (1) g is X -harmonic
- (2) $\lim_{t \uparrow \tau_D} g(X_t) = \phi(X_{\tau_D})$ a.s. $Q^x, x \in D$.

Then $g(x) = E^x[\phi(X_{\tau_D})], x \in D$.

Proof. a) It follows from Lemma 9.10 that (i)_s holds. Fix $x \in D$. Let $\{D_k\}$ be an increasing sequence of open sets such that $\overline{D}_k \subset D$ and $D = \bigcup_k D_k$. Put $\tau_k = \tau_{D_k}, \tau = \tau_D$. Then by the strong Markov property

$$\begin{aligned} \tilde{\phi}(X_{\tau_k}) &= E^{X_{\tau_k}}[\phi(X_{\tau})] = E^x[\theta_{\tau_k}(\phi(X_{\tau}))|\mathcal{F}_{\tau_k}] \\ &= E^x[\phi(X_{\tau})|\mathcal{F}_{\tau_k}] \end{aligned} \quad (9.9)$$

Now $M_k = E^x[\phi(X_{\tau})|\mathcal{F}_{\tau_k}]$ is a bounded (discrete time) martingale so by the martingale convergence theorem Corollary C.9 (Appendix C) we get that

$$\lim_{k \rightarrow \infty} \tilde{\phi}(X_{\tau_k}) = \lim_{k \rightarrow \infty} E^x[\phi(X_{\tau})|\mathcal{F}_{\tau_k}] = \phi(X_{\tau}) \quad (9.10)$$

both pointwise for a.a. ω and in $L^p(Q^x)$, for all $p < \infty$. Moreover, by (9.9) it follows that for each k the process

$$N_t = \tilde{\phi}(X_{\tau_k \vee (t \wedge \tau_{k+1})}) - \tilde{\phi}(X_{\tau_k}); \quad t \geq 0$$

is a martingale w.r.t. $\mathcal{G}_t = \mathcal{F}_{\tau_k \vee (t \wedge \tau_{k+1})}$.

So by the martingale inequality

$$\begin{aligned} Q^x \left[\sup_{\tau_k \leq r \leq \tau_{k+1}} |\tilde{\phi}(X_r) - \tilde{\phi}(X_{\tau_k})| > \epsilon \right] &\leq \frac{1}{\epsilon^2} E^x [|\tilde{\phi}(X_{\tau_{k+1}}) - \tilde{\phi}(X_{\tau_k})|^2] \\ &\rightarrow 0 \text{ as } k \rightarrow \infty, \text{ for all } \epsilon > 0. \end{aligned} \quad (9.11)$$

From (9.10) and (9.11) we conclude that (ii)_s holds.

b) Let D_k, τ_k be as in a). Then since g is X -harmonic we have

$$g(x) = E^x[g(X_{\tau_k})]$$

for all k . So by (2) and bounded convergence

$$g(x) = \lim_{k \rightarrow \infty} E^x[g(X_{\tau_k})] = E^x[\phi(X_{\tau_D})], \text{ as asserted.}$$

Finally we return to the original Dirichlet problem (9.1). First we establish that if a solution of this problem exists, it must be the solution of the stochastic Dirichlet problem found in Theorem 9.11, provided that X_t satisfies *Hunt's condition (H)*:

$$(H) : \text{Every semipolar set for } X_t \text{ is polar for } X_t. \quad (9.12)$$

A *semipolar* set is a countable union of *thin* sets and a measurable set $G \subset \mathbf{R}^n$ is called *thin* (for X_t) if $Q^x[T_G = 0] = 0$ for all x , where $T_G = \inf\{t > 0; X_t \in G\}$ is the first hitting time of G . (Intuitively: For all starting points the process does not hit G immediately, a.s.). A measurable set $F \subset \mathbf{R}^n$ is called *polar* (for X_t) if $Q^x[T_F < \infty] = 0$ for all x . (Intuitively: For all starting points the process never hits F , a.s.). Clearly every polar set is semipolar, but the converse need not to be true (consider the process in Example 9.7). However, condition (H) does hold for Brownian motion (See Blumenthal and Getoor (1968)). It follows from the Girsanov theorem that condition (H) holds for all Ito diffusions whose diffusion coefficient matrix has a bounded inverse and whose drift coefficient satisfies the Novikov condition for all $T < \infty$.

We also need the following result, the proof of which can be found in Blumenthal and Getoor (1968, Prop. II.33):

LEMMA 9.12. Let $U \subset D$ be open and let I denote the set of irregular points of U . Then I is a semipolar set.

THEOREM 9.13. Suppose X_t satisfies Hunt's condition (H). Let ϕ be a bounded continuous function on ∂D . Suppose there exists a bounded $g \in C^2(D)$ such that

- (i) $Lg = 0$ in D
- (ii) $\lim_{\substack{x \rightarrow y \\ x \in D}} g(x) = \phi(y)$ for all regular $y \in \partial D$

Then $g(x) = E^x[\phi(X_{\tau_D})]$.

Proof. Let $\{D_k\}$ be as in the proof Theorem 9.11. By Lemma 9.9b) g is X -harmonic and therefore

$$g(x) = E^x[g(X_{\tau_k})] \quad \text{for all } x \in D_k \text{ and all } k.$$

If $k \rightarrow \infty$ then $X_{\tau_k} \rightarrow X_{\tau_D}$ and so $g(X_{\tau_k}) \rightarrow \phi(X_{\tau_D})$ if X_{τ_D} is regular. From the Lemma 9.12 we know that the set I of irregular points of ∂D is semipolar. So by condition (H) the set I is polar and therefore $X_{\tau_D} \notin I$ a.s. Q^x . Hence

$$g(x) = \lim E^x[g(X_{\tau_k})] = E^x[\phi(X_{\tau_D})], \text{ as claimed.}$$

Under what conditions will the solution $\tilde{\phi}$ of the stochastic Dirichlet problem (9.7) also a solution of the original Dirichlet problem (9.1)? This is a difficult question and we will content ourselves with the following partial answer:

THEOREM 9.14. Suppose L is uniformly elliptic in D , i.e. the eigenvalues of $[a_{ij}]$ are bounded away from 0 in D . (In addition we of course still assume that L is such that b and σ (given by (9.2)) satisfy (5.14) and (5.15)). Let ϕ be a bounded continuous function on ∂D . Put

$$\tilde{\phi}(x) = E^x[\phi(X_{\tau_D})].$$

Then $\tilde{\phi} \in C^{2+\alpha}(D)$ for all $\alpha < 1$ and $\tilde{\phi}$ solves the Dirichlet problem (9.1), i.e.

- (i) $L\tilde{\phi} = 0$ in D
- (ii) $\lim_{\substack{x \rightarrow y \\ x \in D}} \tilde{\phi}(x) = \phi(y)$ for all regular $y \in \partial D$.

Remark. If k is an integer, $\alpha > 0$ and G is an open set $C^{k+\alpha}(G)$ denotes the set of functions on G whose partial derivatives up to k 'th order is Lipschitz (Hölder) continuous with exponent α .

Proof. Choose an open ball Δ with $\overline{\Delta} \subset D$ and let $f \in C^\infty(\partial\Delta)$. Then, from the general theory of partial differential equations, for all $\alpha < 1$ there exists a continuous function u on $\overline{\Delta}$ such that $u|_\Delta \in C^{2+\alpha}(\Delta)$ and

$$\begin{aligned} Lu &= 0 \quad \text{in } \Delta \\ u &= f \quad \text{on } \partial\Delta \end{aligned} \tag{9.13}$$

(see e.g. Dynkin (1965 II, p. 226)). Since $u|_\Delta \in C^{2+\alpha}(\Delta)$ we have: If K is any compact subset of Δ there exists a constant C only depending on K and the C^α -norms of the coefficients of L such that

$$\|u\|_{C^{2+\alpha}(K)} \leq C(\|Lu\|_{C^\alpha(\Delta)} + \|u\|_{C(\Delta)}) \tag{9.14}$$

(See Bers, John and Schechter (1964, Theorem 3, p. 232).) Combining (9.13) and (9.14) we obtain

$$\|u\|_{C^{2+\alpha}(K)} \leq C\|f\|_{C(\partial\Delta)} \quad (9.15)$$

By uniqueness (Theorem 9.13) we know that

$$u(x) = \int f(y)d\mu_x(y), \quad (9.16)$$

where $d\mu_x = Q^x[X_{\tau_\Delta} \in dy]$ is the first exit distribution of X_t from Δ . From (9.15) it follows that

$$|\int f d\mu_{x_1} - \int f d\mu_{x_2}| \leq C\|f\|_{C(\partial\Delta)}|x_1 - x_2|^\alpha; \quad x_1, x_2 \in K. \quad (9.17)$$

By approximating a given continuous function on $\partial\Delta$ uniformly by functions in $C^\infty(\partial\Delta)$ we see that (9.17) holds for all functions $f \in C(\partial\Delta)$. Therefore

$$\|\mu_{x_1} - \mu_{x_2}\| \leq C|x_1 - x_2|^\alpha; \quad x_1, x_2 \in K \quad (9.18)$$

where $\|\cdot\|$ denotes the operator norm on measures on $\partial\Delta$. So if g is any bounded measurable function on $\partial\Delta$ we know that the function

$$\hat{g}(x) = \int g(y)d\mu_x(y) = E^x[g(X_{\tau_\Delta})]$$

belongs to the class $C^\alpha(K)$. Since $\tilde{\phi}(x) = E^x[\tilde{\phi}(X_{\tau_U})]$ for all open sets U with $\overline{U} \subset D$ and $x \in U$ (Lemma 9.10) this applies to $g = \tilde{\phi}$ and we conclude that $\tilde{\phi} \in C^\alpha(M)$ for any compact subset M of D .

We may therefore apply the solution to the problem (9.13) once more, this time with $f = \tilde{\phi}$ and this way we obtain that

$$u(x) = E^x[\tilde{\phi}(X_{\tau_D})] = \tilde{\phi}(x) \quad \text{belongs to } C^{2+\alpha}(M)$$

for any compact $M \subset D$. Therefore (i) holds by Lemma 9.9 a).

To obtain (ii) we apply a theorem from the theory of parabolic differential equations: The Kolmogorov backward equation

$$Lv = \frac{\partial v}{\partial t}$$

has a fundamental solution $v = p(t, x, y)$ jointly continuous in t, x, y for $t > 0$ and bounded in x, y for each fixed $t > 0$ (See Dynkin (1965 II), Theorem 0.4 p. 227). It follows (by bounded convergence) that the process X_t is a *strong Feller process*, in the sense that the function

$$x \rightarrow E^x[f(X_t)] = \int_{\mathbf{R}^n} f(y)p(t, x, y)dy$$

is continuous, for all $t > 0$ and all bounded, measurable functions f . In general we have:

If X_t is a strong Feller Ito diffusion and $D \subset \mathbf{R}^n$ is open then (9.19)

$$\lim_{\substack{x \rightarrow y \\ x \in D}} E^x[\phi(X_{\tau_D})] = \phi(y)$$

for all regular $y \in \partial D$ and bounded $\phi \in C(\partial D)$.

(See Theorem 13.3 p. 32–33 in Dynkin (1965 II)).

Therefore $\tilde{\phi}$ satisfies property (ii) and the proof is complete.

REMARK. One might ask why we don't require that

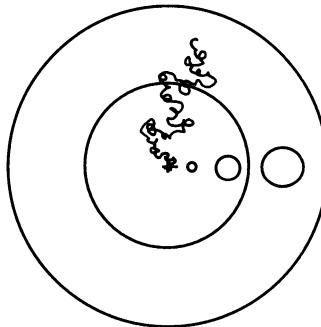
$$(ii)' \quad \lim_{\substack{x \rightarrow y \\ x \in D}} \tilde{\phi}(x) = \phi(y) \text{ for all } y \in \partial D.$$

(and not just for regular points $y \in \partial D$) in our formulation (9.1) of the Dirichlet problem. The answer is that (ii)' is too much to hope for in general, even when L is elliptic: Consider Example 9.6 again, in the case when the point 0 is not regular. Choose $\phi \in C(\partial D)$ such that

$$\phi(0) = 1, 0 \leq \phi(y) < 1 \text{ for } y \in \partial D \setminus \{0\}.$$

Since $\{0\}$ is polar for B_t (see Exercise 9.7 a) we have $B_{\tau_D}^0 \neq 0$ a.s and therefore

$$\tilde{\phi}(0) = E^0[\phi(B_{\tau_D})] < 1.$$



By a slight extension of the mean value property (7.19) (see Exercise 9.4) we get

$$E^0[\tilde{\phi}(X_{\sigma_k})] = E^0[\phi(X_{\tau_D})] = \tilde{\phi}(0) < 1 \quad (9.20)$$

where

$$\sigma_k = \inf\{t > 0; B_t \notin D \cap \{|x| < \frac{1}{k}\}\}, \quad k = 1, 2, \dots$$

This implies that it is impossible that $\tilde{\phi}(x) \rightarrow 1$ as $x \rightarrow 0$. Therefore (ii)' does not hold in this case.

In general one can show that the regular points for Brownian motion are exactly the regular points in the classical potential theoretic sense, i.e. the points y on ∂D where the limit of the generalized Perron-Wiener-Brelot solution coincide with $\phi(y)$, for all $\phi \in C(\partial D)$. See Doob (1984), Port and Stone (1979) or Rao (1977).

EXAMPLE 9.15. Let D denote the infinite strip

$$D = \{(t, x) \in \mathbf{R}^2; x < R\}, \text{ where } R \in \mathbf{R}$$

and let L be the differential operator

$$Lf(t, x) = \frac{\partial f}{\partial t} + \frac{1}{2} \frac{\partial^2 f}{\partial x^2}; \quad f \in C^2(D).$$

An Ito diffusion whose generator coincides with L on $C_0^2(\mathbf{R}^2)$ is (see Example 9.5)

$$X_t = (s + t, B_t); \quad t \geq 0,$$

and all the points of ∂D are regular for this process. It is not hard to see that in this case (9.5) holds, i.e.

$$\tau_D < \infty \text{ a.s.}$$

(see Exercise 7.4).

Assume that ϕ is a bounded continuous function on $\partial D = \{(t, R); t \in \mathbf{R}\}$. Then by Theorem 9.11 the function

$$\tilde{\phi}(s, x) = E^{s,x}[\phi(X_{\tau_D})]$$

is the solution of the stochastic Dirichlet problem (9.7), when $E^{s,x}$ denotes expectation w.r.t. the probability law $Q^{s,x}$ for X starting at (s, x) . Does $\tilde{\phi}$ also solve the problem (9.1)? Using the Laplace transform it is possible to find the distribution of the first exit point on ∂D for X , i.e. to find the distribution of the first time $t = \hat{\tau}$ that B_t reaches the value R . (See Karlin and Taylor (1975), p. 363). The result is

$$P^x[\hat{\tau} \in du] = g(x, u)du,$$

where

$$g(x, u) = \begin{cases} (R - x)(2\pi u^3)^{-1} \exp(-\frac{(R-x)^2}{2u}); & u > 0 \\ 0; & u \leq 0 \end{cases}$$

Thus the solution $\tilde{\phi}$ may be written

$$\tilde{\phi}(s, x) = \int_0^\infty \phi(s + t, R)g(x, t)dt = \int_s^\infty \phi(u, R)g(x, u - s)dt.$$

From the explicit formula for $\tilde{\phi}$ it is clear that $\frac{\partial \tilde{\phi}}{\partial s}$ and $\frac{\partial^2 \tilde{\phi}}{\partial x^2}$ are continuous and we conclude that $L\tilde{\phi} = 0$ in D by Lemma 9.9. So $\tilde{\phi}$ satisfies (i) of (9.1). What about property (ii) of (9.1)? It is not hard to see that for $t > 0$

$$E^{t_0,x}[f(X_t)] = (2\pi t)^{-\frac{1}{2}} \int_R f(t_0 + t, y) \exp(-\frac{|x - y|^2}{2t}) dy$$

for all bounded, (t, x) -measurable functions f . (See (2.2)). Therefore X_t is not a strong Feller process, so we cannot appeal to (9.19) to obtain (ii). However, it is easy to verify directly that if $|y| = R, t_1 > 0$ then for all $\epsilon > 0$ there exists $\delta > 0$ such that $|x - y| < \delta, |t - t_1| < \delta \Rightarrow Q^{t,x}[\tau_D \in N] \geq 1 - \epsilon$, where $N = [t_1 - \epsilon, t_1 + \epsilon] \times \{y\}$. And this is easily seen to imply (ii).

REMARK. As the above example (and Example 9.7) shows, an Ito diffusion need not be a strong Feller process. However, we have seen that it is always a Feller process (Lemma 8.4).

B) The Poisson problem

Let $L = \sum a_{ij} \frac{\partial^2}{\partial x_i \partial x_j} + \sum b_i \frac{\partial}{\partial x_i}$ be a semi-elliptic partial differential operator on a domain $D \subset \mathbf{R}^n$ as before and let X_t be an associated Ito diffusion, described by (9.2) and (9.3). The (generalized) *Poisson problem* is the following:

Given a continuous function g on D find a C^2 function f in D such that

$$Lf = -g \quad \text{in } D \tag{9.21}$$

$$\lim_{\substack{x \rightarrow y \\ x \in D}} f(x) = 0 \quad \text{for all regular } y \in \partial D. \tag{9.22}$$

Again we will first study a stochastic version of the problem and then investigate the relation between the corresponding stochastic solution and the deterministic solution (if it exists) of (9.21–22):

THEOREM 9.16. (Solution of the stochastic Poisson problem)

Assume that

$$E^x \left[\int_0^{\tau_D} |g(X_s)| ds \right] < \infty \quad \text{for all } x \in D \tag{9.23}$$

(This occurs, for example, if g is bounded and $E^x[\tau_D] < \infty$ for all $x \in D$). Define

$$\check{g}(x) = E^x \left[\int_0^{\tau_D} g(X_s) ds \right] \tag{9.24}$$

Then

$$\mathcal{A}\check{g} = -g \quad \text{in } D, \tag{9.25}$$

and

$$\lim_{t \uparrow \tau_D} \check{g}(X_t) = 0 \text{ a.s. } Q^x, \text{ for all } x \in D. \quad (9.26)$$

Proof. Choose U open, $x \in U \subset\subset D$. Put $\eta = \int_0^{\tau_D} g(X_s)ds$, $\tau = \tau_U$.

Then by the strong Markov property (7.15)

$$\begin{aligned} \frac{E^x[\check{g}(X_\tau)] - \check{g}(x)}{E^x[\tau]} &= \frac{1}{E^x[\tau]} (E^x[E^{X_\tau}[\eta]] - E^x[\eta]) \\ &= \frac{1}{E^x[\tau]} (E^x[E^x[\theta_\tau \eta | \mathcal{F}_\tau]] - E^x[\eta]) = \frac{1}{E^x[\tau]} (E^x[\theta_\tau \eta - \eta]) \end{aligned}$$

Approximate η by sums of the form

$$\eta^{(k)} = \sum g(X_{t_i}) \chi_{\{t_i < \tau_D\}} \Delta t_i$$

Since

$$\theta_t \eta^{(k)} = \sum g(X_{t_i+t}) \chi_{\{t_i + t < \tau_D^t\}} \Delta t_i \text{ for all } k$$

(see the argument for (7.16)) we see that

$$\theta_\tau \eta = \int_\tau^{\tau_D} g(X_s)ds. \quad (9.27)$$

Therefore

$$\frac{E^x[\check{g}(X_\tau)] - \check{g}(x)}{E^x[\tau]} = \frac{-1}{E^x[\tau]} E^x \left[\int_0^\tau g(X_s)ds \right] \rightarrow -g(x) \text{ as } U \downarrow x,$$

since g is continuous. This proves (9.25).

Put $H(x) = E^x \left[\int_0^{\tau_D} |g(X_s)|ds \right]$. Let D_k, τ_k be as in the proof of Theorem 9.11. Then by the same argument as above we get

$$\begin{aligned} E^x[H(X_{\tau_k \wedge t})] &= E^x \left[E^x \left[\int_{\tau_k \wedge t}^{\tau_D} |g(X_s)|ds | \mathcal{F}_{\tau_k \wedge t} \right] \right] \\ &= E^x \left[\int_{\tau_k \wedge t}^{\tau_D} |g(X_s)|ds \right] \rightarrow 0 \text{ as } t \rightarrow \tau_D, k \rightarrow \infty \end{aligned}$$

by dominated convergence. This implies (9.26).

REMARK. For functions g satisfying (9.23) define the operator \mathcal{R} by

$$(\mathcal{R}g)(x) = \check{g}(x) = E^x \left[\int_0^{\tau_D} g(X_s) ds \right]$$

Then (9.25) can be written

$$\mathcal{A}(\mathcal{R}g) = -g \quad (9.28)$$

i.e. the operator $-\mathcal{R}$ is a right inverse of the operator \mathcal{A} . Similarly, if we define

$$\mathcal{R}_\alpha g(x) = E^x \left[\int_0^{\tau_D} g(X_s) e^{-\alpha s} ds \right] \text{ for } \alpha \geq 0 \quad (9.29)$$

then the same method of proof as in Theorem 8.5 gives that

$$(\mathcal{A} - \alpha)\mathcal{R}_\alpha g = -g; \quad \alpha \geq 0 \quad (9.30)$$

(If $\alpha > 0$ then the assumption (9.23) can be replaced by the assumption that g is bounded (and continuous as before)).

Thus we may regard the operator \mathcal{R}_α as a generalization of the resolvent operator R_α discussed in Chapter VIII, and formula (9.30) as the analogue of Theorem 8.5 b).

Next we establish that if a solution f of the original problem (9.21–22) exists, then f is the solution \check{g} of the stochastic problem (9.25–26):

THEOREM 9.17. (Uniqueness theorem for the Poisson equation)

Assume that X_t satisfies Hunt's condition (H) ((9.12)). Assume that (9.23) holds and that there exists a function $f \in C^2(D)$ and a constant C such that

$$|f(x)| \leq C(1 + E^x \left[\int_0^{\tau_D} |g(X_s)| ds \right]) \text{ for all } x \in D \quad (9.31)$$

and with the properties

$$Lf = -g \text{ in } D, \quad (9.32)$$

$$\lim_{\substack{x \rightarrow y \\ x \in D}} f(x) = 0 \text{ for all regular points } y \in \partial D. \quad (9.33)$$

Then $f(x) = \check{g}(x) = E^x \left[\int_0^{\tau_D} g(X_s) ds \right]$.

Proof. Let D_k, τ_k be as in the proof of Theorem 9.11. Then by Dynkin's formula

$$E^x[f(X_{\tau_k})] - f(x) = E^x \left[\int_0^{\tau_k} (Lf)(X_s) ds \right] = -E^x \left[\int_0^{\tau_k} g(X_s) ds \right]$$

By dominated convergence we obtain

$$f(x) = \lim_{k \rightarrow \infty} (E^x[f(X_{\tau_k})] + E^x[\int_0^{\tau_k} g(X_s)ds]) = E^x[\int_0^{\tau_D} g(X_s)ds],$$

since X_{τ_D} is a regular point a.s. by condition (H) and Lemma 9.12. Finally we combine the Dirichlet and Poisson problem and obtain the following result:

THEOREM 9.18. (Solution of the combined stochastic Dirichlet and Poisson problem)

Let $\phi \in C(\partial D)$ be bounded and let $g \in C(D)$ satisfy

$$E^x[\int_0^{\tau_D} |g(X_s)|ds] < \infty \text{ for all } x \in D. \quad (9.34)$$

Define

$$h(x) = E^x[\int_0^{\tau_D} g(X_s)ds] + E^x[\phi(X_{\tau_D})], \quad x \in D. \quad (9.35)$$

a) Then

$$\mathcal{A}h = -g \text{ in } D \quad (9.36)$$

and

$$\lim_{t \uparrow \tau_D} h(X_t) = \phi(X_{\tau_D}) \text{ a.s. } Q^x, \quad \text{for all } x \in D. \quad (9.37)$$

Moreover, if there exists a function $h_1 \in C^2(D)$ and a constant C such that

$$|h_1(x)| < C(1 + E^x[\int_0^{\tau_D} |g(X_s)|ds]), \quad x \in D, \quad (9.38)$$

and h_1 satisfies (9.36) and (9.37), then $h_1 = h$.

REMARK. With an approach similar to the one used in Theorem 9.13 one can prove that if L is uniformly elliptic in D and $g \in C^\alpha(D)$ (for some $\alpha > 0$) is bounded, then the function h given by (9.35) solves the Dirichlet-Poisson problem, i.e.

$$Lh = -g \text{ in } D \quad (9.39)$$

and

$$\lim_{\substack{x \rightarrow y \\ x \in D}} h(x) = \phi(y) \text{ for all regular } y \in \partial D. \quad (9.40)$$

The Green measure

The solution \check{g} given by the formula (9.24) may be rewritten as follows:

DEFINITION 9.19. The *Green measure* (of X_t w.r.t. D at x), $G(x, \cdot)$ is defined by

$$G(x, H) = E^x \left[\int_0^{\tau_D} \chi_H(X_s) ds \right], \quad H \subset \mathbf{R}^n \text{ Borel set} \quad (9.41)$$

or

$$\int f(y) G(x, dy) = E^x \left[\int_0^{\tau_D} f(X_s) ds \right], \quad f \text{ bounded, continuous.} \quad (9.42)$$

In other words, $G(x, H)$ is the expected length of time the process stays in H before it exits from D . If X_t is Brownian motion, then

$$G(x, H) = \int_H G(x, y) dy,$$

where $G(x, y)$ is the classical Green function w.r.t. D and dy denotes Lebesgue measure. See Doob (1984), Port and Stone (1979) or Rao (1977).

Note that using the Fubini theorem we obtain the following relation between the Green measure G and the *transition measure* for X_t in D , $Q_t^D(x, H) = Q^x[X_t \in H, t < \tau_D]$:

$$G(x, H) = E^x \left[\int_0^{\infty} \chi_H(X_s) \cdot \chi_{[0, \tau_D)}(s) ds \right] = \int_0^{\infty} Q_t^D(x, H) dt. \quad (9.43)$$

From (9.42) we get

$$\check{g}(x) = E^x \left[\int_0^{\tau_D} g(X_s) ds \right] = \int_D g(y) G(x, dy), \quad (9.44)$$

which is the familiar formula for the solution of the Poisson equation in the classical case.

Also note that by using the Green function, we may regard the Dynkin formula as a generalization of the classical Green formula:

COROLLARY 9.20. (The Green formula)

Let $E^x[\tau_D] < \infty$ for all $x \in D$ and assume that $f \in C_0^2(\mathbf{R}^n)$. Then

$$f(x) = E^x[f(X_{\tau_D})] - \int_D (L_X f)(y) G(x, dy). \quad (9.45)$$

In particular, if $f \in C_0^2(D)$ we have

$$f(x) = - \int_D (L_X f)(y) G(x, dy). \quad (9.46)$$

(As before $L_X = \sum b_i \frac{\partial}{\partial x_i} + \frac{1}{2} \sum (\sigma \sigma^T)_{ij} \frac{\partial^2}{\partial x_i \partial x_j}$ when

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t.$$

Proof. By Dynkin's formula and (9.44) we have

$$E^x[f(X_{\tau_D})] = f(x) + E^x\left[\int_0^{\tau_D} (L_X f)(X_s) ds\right] = f(x) + \int_D (L_X f)(y) G(x, dy).$$

REMARK. Combining (9.28) with (9.46) we see that if $E^x[\tau_K] < \infty$ for all compacts $K \subset D$ and all $x \in D$, then $-\mathcal{R}$ is the inverse of the operator \mathcal{A} on $C_0^2(D)$:

$$\mathcal{A}(\mathcal{R}f) = \mathcal{R}(\mathcal{A}f) = -f, \text{ for all } f \in C_0^2(D) \quad (9.47)$$

More generally, for all $\alpha \geq 0$ we get the following analogue of Theorem 8.5:

$$(\mathcal{A} - \alpha)(\mathcal{R}_\alpha f) = \mathcal{R}_\alpha(\mathcal{A} - \alpha)f = -f \text{ for all } f \in C_0^2(D) \quad (9.48)$$

The first part of this is already established in (9.30) and the second part follows from the following useful extension of the Dynkin formula

$$E^x[e^{-\alpha\tau} f(X_\tau)] = f(x) + E^x\left[\int_0^\tau e^{-\alpha s} (\mathcal{A} - \alpha)f(X_s) ds\right]. \quad (9.49)$$

If $\alpha > 0$ this is valid for all stopping times $\tau \leq \infty$ and all $f \in C_0^2(\mathbf{R}^n)$. (See Exercise 9.6.)

EXERCISES

9.1.

In each of the cases below find an Ito diffusion whose generator coincides with L on C_0^2 :

- a) $Lf(t, x) = \alpha \frac{\partial f}{\partial t} + \frac{1}{2} \beta^2 \frac{\partial^2 f}{\partial x^2}; \alpha, \beta \text{ constants}$
- b) $Lf(x_1, x_2) = a \frac{\partial f}{\partial x_1} + b \frac{\partial f}{\partial x_2} + \frac{1}{2} \left(\frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2} \right); \alpha, \beta \text{ constants}$
- c) $Lf(x) = \alpha x f'(x) + \frac{1}{2} \beta^2 f''(x); \alpha, \beta \text{ constants}$
- d) $Lf(x) = \alpha f'(x) + \frac{1}{2} \beta^2 x^2 f''(x); \alpha, \beta \text{ constants}$
- e) $Lf(x_1, x_2) = \ln(1 + x_1^2) \frac{\partial f}{\partial x_1} + x_2 \frac{\partial f}{\partial x_2} + x_2^2 \frac{\partial^2 f}{\partial x_2^2} + 2x_1 x_2 \frac{\partial^2 f}{\partial x_1 \partial x_2} + 2x_1^2 \frac{\partial^2 f}{\partial x_1^2}$

9.2.

Use Theorem 9.11 to find the bounded solutions of the following boundary value problems:

$$(i) \quad \begin{cases} \frac{\partial u}{\partial t} + \frac{1}{2} \cdot \frac{\partial^2 u}{\partial x^2} = e^{\rho t} \phi(x); & 0 < t < T, x \in \mathbf{R} \\ u(T, x) = \psi(x); & x \in \mathbf{R} \end{cases}$$

where ϕ, ψ are given bounded, continuous functions.

$$(ii) \quad \begin{cases} \alpha x u'(x) + \frac{1}{2} \beta^2 x^2 u''(x) = 0; & 0 < x < x_0 \\ u(x_0) = x_0^2 \end{cases}$$

where α, β are given constants, $\alpha \geq \frac{1}{2} \beta^2$.

- (iii) If $\alpha < \frac{1}{2} \beta^2$ there are infinitely many bounded solutions of (ii), and an additional boundary condition e.g. at $x = 0$ is needed to provide uniqueness. Explain this in view of Theorem 9.11.

9.3.

Write down (using Brownian motion) and compare the solutions $u(t, x)$ of the following two boundary value problems:

$$a) \quad \begin{cases} \frac{\partial u}{\partial t} + \frac{1}{2} \Delta u = 0 & \text{for } 0 < t < T, x \in \mathbf{R}^n \\ u(T, x) = \phi(x) & \text{for } x \in \mathbf{R}^n \end{cases}$$

$$b) \quad \begin{cases} \frac{\partial u}{\partial t} - \frac{1}{2} \Delta u = 0 & \text{for } 0 < t < T, x \in \mathbf{R}^n \\ u(0, x) = \psi(x) & \text{for } x \in \mathbf{R}^n \end{cases}$$

9.4.

Let G and H be bounded open subsets of \mathbf{R}^n , $G \subset H$, and let B_t be n -dimensional Brownian motion. Use the property (H) for B_t to prove that

$$\inf\{t > 0; B_t \notin H\} = \inf\{t > \tau_G; B_t \notin H\}$$

i.e., with the terminology of (7.16),

$$\tau_H = \tau_H^\alpha \quad \text{where} \quad \alpha = \tau_G$$

Use this to prove that if $X_t = B_t$ then the mean value property (7.19) holds for all bounded open $G \subset H$, i.e. it is not necessary to require $G \subset\subset H$ in this case. This verifies the statement (9.20).

9.5. (The eigenvalues of the Laplacian)

Let $D \subset \mathbf{R}^n$ be open, bounded and let $\lambda \in \mathbf{R}$.

- a) Suppose there exists a solution $u \in C^2(D) \cap C(\overline{D})$, u not identically zero, such that

$$\begin{cases} -\frac{1}{2}\Delta u = \lambda u & \text{in } D \\ u = 0 & \text{on } \partial D \end{cases} \quad (9.50)$$

Show that we must have $\lambda > 0$. (Hint: If $\frac{1}{2}\Delta u = -\lambda u$ in D then

$$\left\langle \frac{1}{2}\Delta u, u \right\rangle = \langle -\lambda u, u \rangle$$

where

$$\langle u, v \rangle = \int_D u(x)v(x)dx.$$

Now use integration by parts.)

- b) It can be shown that if D is smooth then there exists $0 < \lambda_0 < \lambda_1 < \dots < \lambda_n < \dots$ where $\lambda_n \rightarrow \infty$ such that (9.50) holds for $\lambda = \lambda_n$, $n = 0, 1, 2, \dots$, and for no other values of λ . The numbers $\{\lambda_n\}$ are called the *eigenvalues* of the operator $-\frac{1}{2}\Delta$ in the domain D and the corresponding (nontrivial) solutions u_n of (9.50) are called the *eigenfunctions*. There is an interesting probabilistic interpretation of the lowest eigenvalue λ_0 . The following result indicates this:

Put $\tau = \tau_D = \inf\{t > 0; B_t \notin D\}$, choose $\rho > 0$ and define

$$w_\rho(x) = E^x[\exp(\rho\tau)]; \quad x \in D.$$

Prove that if $w_\rho(x) < \infty$ for all $x \in D$ then ρ is not an eigenvalue for $-\frac{1}{2}\Delta$. (Hint: Let u be a solution of (9.50) with $\lambda = \rho$. Apply Dynkin's formula to the process $dY_t = (dt, dB_t)$ and the function $f(t, x) = e^{\rho t}u(x)$ to deduce that $u(x) = 0$ for $x \in D$).

- c) Conclude that

$$\lambda_0 \geq \sup\{\rho; E^x[\exp(\rho\tau)] < \infty \text{ for all } x \in D\}$$

(We have in fact equality here. See for example Durrett (1984), Ch. 8B).

9.6.

Prove formula (9.49), for example by applying the Dynkin formula to the process

$$dY_t = \begin{bmatrix} dt \\ dX_t \end{bmatrix}$$

and the function $g(y) = g(t, x) = e^{-\alpha t}f(x)$.

9.7.

- a) Let B_t be Brownian motion in \mathbf{R}^2 . Prove that

$$P^x[\exists t > 0; B_t = y] = 0 \quad \text{for all } x, y \in \mathbf{R}^2$$

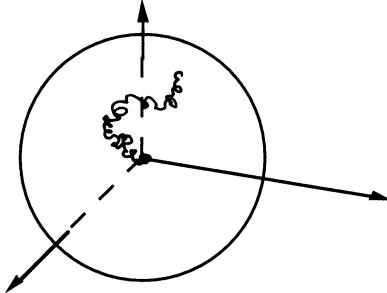
(Hint: First assume $x \neq y$. We may choose $y = 0$. One possible approach would be to apply Dynkin's formula with $f(u) = \ln|u|$ and $\tau = \inf\{t > 0; |B_t| \leq \rho \text{ or } |B_t| \geq R\}$, where $0 < \rho < R$. Let $\rho \rightarrow 0$ and then $R \rightarrow \infty$. If $x = y$ consider $P^x[\exists t > \epsilon; B_t = x]$ and use the Markov property.)

- b) Let $B_t = (B_t^{(1)}, B_t^{(2)})$ be Brownian motion in \mathbf{R}^2 . Prove that $\tilde{B}_t = (-B_t^{(1)}, B_t^{(2)})$ is also a Brownian motion.
- c) Prove that $0 \in \mathbf{R}^2$ is a *regular* boundary point (for Brownian motion) of the plane region

$$D = \{(x_1, x_2) \in \mathbf{R}^2; x_1^2 + x_2^2 < 1\} \setminus \{(x_1, 0); x_1 \geq 0\}$$

- d) Prove that $0 \in \mathbf{R}^3$ is an *irregular* boundary point (for Brownian motion) of the 3-dimensional region

$$U = \{(x_1, x_2, x_3) \in \mathbf{R}^3, x_1^2 + x_2^2 + x_3^2 < 1\} \setminus \{(x_1, 0, 0); x_1 \geq 0\}.$$



9.8.

- a) Find an Ito diffusion X_t and a measurable set G which is semipolar but not polar for X_t .
- b) Find an Ito diffusion X_t and a countable family of thin sets H_k ; $k = 1, 2, \dots$ such that $\bigcup_{k=1}^{\infty} H_k$ is not thin.

9.9.

- a) Let X_t be an Ito diffusion in \mathbf{R}^n and assume that g is a non-constant locally bounded real X_t -harmonic function on a connected open set $G \subset \mathbf{R}^n$. Prove that g satisfies the following weak form of the *maximum principle*: g does not have a (local or global) maximum at any point of G . (Similarly g satisfies the *minimum principle*).
- b) Give an example to show that a non-constant bounded X_t -harmonic function g can have a (non-strict) global maximum. (Hint: Consider uniform motion to the right.)

9.10.

Find the (stochastic) solution $f(t, x)$ of the boundary value problem

$$\begin{cases} K(x)e^{-\rho t} + \frac{\partial f}{\partial t} + \alpha x \frac{\partial f}{\partial x} + \frac{1}{2}\beta^2 x^2 \frac{\partial^2 f}{\partial x^2} = 0 & \text{for } x > 0, 0 < t < T \\ f(T, x) = e^{-\rho T} \phi(x) & \text{for } x > 0 \end{cases}$$

where K, ϕ are given functions and T, ρ, α, β are constants, $\rho > 0$, $T > 0$. (Hint: Consider $dY_t = (dt, dX_t)$ where X_t is a geometric Brownian motion).

9.11.

- a) The Poisson kernel is defined by

$$P_r(\theta) = \frac{1 - r^2}{1 - 2r \cos \theta + r^2} = \frac{1 - |z|^2}{|1 - z|^2}$$

where $r \geq 0$, $\theta \in [0, 2\pi]$, $z = re^{i\theta} \in \mathbf{C}$ ($i = \sqrt{-1}$).

The Poisson formula states that if D denotes the open unit disk in the plane $\mathbf{R}^2 = \mathbf{C}$ and $h \in C(\overline{D})$ satisfies $\Delta h = 0$ in D then

$$h(re^{i\theta}) = \frac{1}{2\pi} \int_0^{2\pi} P_r(t - \theta) h(e^{it}) dt$$

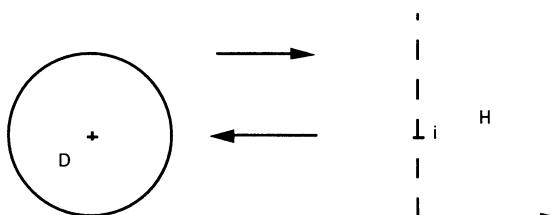
Prove that the probability that Brownian motion, starting from $z \in D$, first exits from D at a set $F \subset \partial D$ is given by

$$\frac{1}{2\pi} \int_F P_r(t - \theta) dt, \quad \text{where } z = re^{i\theta}.$$

- b) The function

$$w = \phi(z) = i \frac{1+z}{1-z}$$

maps the disc $D = \{|z| < 1\}$ conformally onto the half plane $H = \{w = u + iv; v > 0\}$, $\phi(\partial D) = \mathbf{R}$ and $\phi(0) = i$. Use Example 8.22 to prove that if μ denotes the harmonic measure for Brownian motion at the point $i = (0, 1)$ for the half plane H then



$$\int_{\mathbf{R}} f(\xi) d\mu(\xi) = \frac{1}{2\pi} \int_0^{2\pi} f(\phi(e^{it})) dt = \frac{1}{2\pi i} \int_{\partial D} \frac{f(\phi(z))}{z} dz$$

- c) Substitute $w = \phi(z)$ (i.e. $z = \psi(w) := \phi^{-1}(w) = \frac{w-i}{w+i}$) in the integral above to show that

$$\int_{\mathbf{R}} f(\xi) d\mu(\xi) = \frac{1}{\pi} \int_{\partial H} f(w) \frac{dw}{|w-i|^2} = \frac{1}{\pi} \int_{-\infty}^{\infty} f(x) \frac{dx}{x^2 + 1}$$

- d) Show that the harmonic measure μ_H^w for Brownian motion in H at the point $w = u + iv \in H$ is given by

$$d\mu_H^w(x) = \frac{1}{\pi} \cdot \frac{v}{(x-u)^2 + v^2} dx.$$

9.12. (A Feynman-Kac formula for boundary value problems)

Let X_t be an Ito diffusion on \mathbf{R}^n whose generator coincides with a given partial differential operator L on $C_0^2(\mathbf{R}^n)$. Let D, ϕ and g be as in Theorem 9.18 and let $q(x) \geq 0$ be a continuous function on \mathbf{R}^n . Consider the boundary value problem: Find $h \in C_2(D) \cap C(\overline{D})$ such that

$$\begin{cases} Lh(x) - q(x)h(x) = -g(x) & \text{on } D \\ \lim_{x \rightarrow y} h(x) = \phi(y); & y \in \partial D \end{cases}$$

Show that if a bounded solution h exists, then

$$h(x) = E^x \left[\int_0^{\tau_D} e^{- \int_0^t q(X_s) ds} g(X_t) dt + e^{- \int_0^{\tau_D} q(X_s) ds} \phi(X_{\tau_D}) \right].$$

(Compare with the Feynman-Kac formula.)

Hint: Proceed as in the proof of Theorem 8.6 b).

For more information on stochastic solutions of boundary value problems see Freidlin (1985).

X. Application to Optimal Stopping

Problem 5 in the introduction is a special case of a problem of the following type:

PROBLEM 10.1. (The optimal stopping problem)

Let X_t be an Ito diffusion on \mathbf{R}^n and let g (*the reward function*) be a given function on \mathbf{R}^n , satisfying

- a) $g(\xi) \geq 0$ for all $\xi \in \mathbf{R}^n$
 - b) g is continuous
- (10.1)

Find a stopping time $\tau^* = \tau^*(x, \omega)$ for $\{X_t\}$ such that

$$E^x[g(X_{\tau^*})] = \sup_{\tau} E^x[g(X_{\tau})] \text{ for all } x \in \mathbf{R}^n, \quad (10.2)$$

the sup being taken over all stopping times τ for $\{X_t\}$. We also want to find the corresponding optimal expected reward

$$g^*(x) = E^x[g(X_{\tau^*})]. \quad (10.3)$$

Here $g(X_{\tau})$ is to be interpreted as 0 at the points $\omega \in \Omega$ where $\tau(\omega) = \infty$ and as usual E^x denotes the expectation with respect to the probability law Q^x of the process $X_t; t \geq 0$ starting at $X_0 = x \in \mathbf{R}^n$.

We may regard X_t as the state of a game at time t , each ω corresponds to one sample of the game. For each time t we have the option of stopping the game, thereby obtaining the reward $g(X_t)$, or continue the game in the hope that stopping it at a later time will give a bigger reward. The problem is of course that we do not know what state the game is in at future times, only the probability distribution of the “future”. Mathematically, this means that the possible “stopping” times we consider really are stopping times in the sense of Definition 7.3: The decision whether $\tau \leq t$ or not should only depend on the behaviour of the Brownian motion B_r (driving the process X) up to time t , or perhaps only on the behaviour of X_r up to time t . So, among all possible stopping times τ we are asking for the optimal one, τ^* , which gives the best result “in the long run”, i.e. the biggest expected reward in the sense of (10.2).

In the following we will outline how a solution to this problem can be obtained using the material from the preceding chapter. Later in this chapter we shall see that our discussion of problem (10.2)–(10.3) also covers the apparently more general problems

$$g^*(s, x) = \sup_{\tau} E^{s,x}[g(\tau, X_\tau)] = E^{s,x}[g(\tau^*, X_{\tau^*})] \quad (10.4)$$

and

$$\begin{aligned} G^*(s, x) &= \sup_{\tau} E^{s,x}\left[\int_0^\tau f(t, X_t)dt + g(\tau, X_\tau)\right] \\ &= E^{s,x}\left[\int_0^{\tau^*} f(t, X_t)dt + g(\tau^*, X_{\tau^*})\right] \end{aligned} \quad (10.5)$$

where f is a given *reward rate function* (satisfying certain conditions).

We shall also discuss possible extensions of problem (10.2)–(10.3) to cases where g is not necessarily continuous or where g may assume negative values.

A basic concept in the solution of (10.2)–(10.3) is the following:

DEFINITION 10.2. A measurable function $f : \mathbf{R}^n \rightarrow [0, \infty]$ is called *supermeanvalued* (w.r.t. X_t) if

$$f(x) \geq E^x[f(X_\tau)] \quad (10.6)$$

for all stopping times τ and all $x \in \mathbf{R}^n$.

If, in addition, f is also *lower semicontinuous*, then f is called *l.s.c. superharmonic* or just *superharmonic* (w.r.t. X_t).

Note that if $f : \mathbf{R}^n \rightarrow [0, \infty]$ is lower semicontinuous then by the Fatou lemma

$$f(x) \leq E^x\left[\liminf_{k \rightarrow \infty} f(X_{\tau_k})\right] \leq \liminf_{k \rightarrow \infty} E^x[f(X_{\tau_k})], \quad (10.7)$$

for any sequence $\{\tau_k\}$ of stopping times such that $\tau_k \rightarrow 0$ a.s. Q^x . Combining this with (10.6) we see that if f is (l.s.c.) superharmonic, then

$$f(x) = \lim_{k \rightarrow \infty} E^x[f(X_{\tau_k})] \text{ for all } x, \quad (10.8)$$

for all such sequences τ_k .

REMARKS. 1) In the literature (see e.g. Dynkin (1965 II)) one often finds a weaker concept of X_t -superharmonicity, defined by the supermeanvalued property (10.6) plus the stochastic continuity requirement (10.8). This weaker concept corresponds to the X_t -harmonicity defined in Chapter IX.

2) If $f \in C^2(\mathbf{R}^n)$ it follows from Dynkin's formula that f is superharmonic w.r.t. X_t if and only if

$$\mathcal{A}f \leq 0$$

where \mathcal{A} is the characteristic operator of X_t . This is often a useful criterion (See e.g. Example 10.15).

3) If $X_t = B_t$ is Brownian motion in \mathbf{R}^n then the superharmonic functions for X_t coincide with the (nonnegative) superharmonic functions in classical potential theory. See Doob (1984) or Port and Stone (1979).

We state some useful properties of superharmonic and supermeanvalued functions.

LEMMA 10.3.

- a) If f is superharmonic (supermeanvalued) and $\alpha > 0$, then αf is superharmonic (supermeanvalued).
- b) If f_1, f_2 are superharmonic (supermeanvalued), then $f_1 + f_2$ is superharmonic (supermeanvalued).
- c) If $\{f_j\}_{j \in J}$ is a family of supermeanvalued functions, then $f(x) := \inf_{j \in J} f_j(x)$ is supermeanvalued (J is any set).
- d) If f_1, f_2, \dots are nonnegative superharmonic (supermeanvalued) functions and $f_k \uparrow f$ pointwise, then f is superharmonic (supermeanvalued).
- e) If f is supermeanvalued and $\sigma \leq \tau$ are stopping times, then $E^x[f(X_\sigma)] \geq E^x[f(X_\tau)]$.
- f) If f is supermeanvalued and H is a Borel set, then $\tilde{f}(x) := E^x[f(X_{\tau_H})]$ is supermeanvalued.

Proof of Lemma 10.3.

a) and b) are straightforward.

c) Suppose f_j is supermeanvalued for all $j \in J$. Then

$$f_j(x) \geq E^x[f_j(X_\tau)] \geq E^x[f(X_\tau)] \text{ for all } j.$$

so $f(x) = \inf f_j(x) \geq E^x[f(X_\tau)]$, as required.

d) Suppose f_j is supermeanvalued, $f_j \uparrow f$. Then

$$\begin{aligned} f(x) &\geq f_j(x) \geq E^x[f_j(X_\tau)] \text{ for all } j, \text{ so} \\ f(x) &\geq \lim_{j \rightarrow \infty} E^x[f_j(X_\tau)] = E^x[f(X_\tau)], \end{aligned}$$

by monotone convergence. Hence f is supermeanvalued. If each f_j also satisfies (10.8) then

$$f_j(x) = \lim_{k \rightarrow \infty} E^x[f_j(X_{\tau_k})] \leq \lim_{k \rightarrow \infty} E^x[f(X_{\tau_k})]$$

and so

$$f(x) \leq \lim_{k \rightarrow \infty} E^x[f(X_{\tau_k})].$$

e) If f is supermeanvalued we have by the Markov property when $t > s$

$$E^x[f(X_t)|\mathcal{F}_s] = E^{X_s}[f(X_{t-s})] \leq f(X_s), \quad (10.9)$$

i.e. the process

$$\zeta_t = f(X_t)$$

is a *supermartingale* w.r.t. the σ -algebras \mathcal{F}_t generated by $\{B_r; r \leq t\}$. (Appendix C). Therefore, by Doob's optional sampling theorem (see Gihman and Skorohod (1975, Theorem 6 p.11)) we have

$$E^x[f(X_\sigma)] \geq E^x[f(X_\tau)]$$

for all stopping times σ, τ with $\sigma \leq \tau$ a.s. Q^x .

f) Suppose f is supermeanvalued. By the strong Markov property (7.15) and formula (7.16) we have, for any stopping time α ,

$$\begin{aligned} E^x[\tilde{f}(X_\alpha)] &= E^x[E^{X_\alpha}[f(X_{\tau_H})]] = E^x[E^x[\theta_\alpha f(X_{\tau_H})|\mathcal{F}_\alpha]] \\ &= E^x[\theta_\alpha f(X_{\tau_H})] = E^x[f(X_{\tau_H^\alpha})] \end{aligned} \quad (10.10)$$

where $\tau_H^\alpha = \inf\{t > \alpha; X_t \notin H\}$. Since $\tau_H^\alpha \geq \tau_H$ we have by e)

$$E^x[\tilde{f}(X_\alpha)] \leq E^x[f(X_{\tau_H})] = \tilde{f}(x),$$

so \tilde{f} is supermeanvalued. □

The following concepts are fundamental:

DEFINITION 10.4. Let h be a real measurable function on \mathbf{R}^n . If f is a superharmonic (supermeanvalued) function and $f \geq h$ we say that f is a *superharmonic (supermeanvalued) majorant* of h (w.r.t. X_t). The function

$$\bar{h}(x) = \inf_f f(x); \quad x \in \mathbf{R}^n, \quad (10.11)$$

the inf being taken over all supermeanvalued majorants f of h , is again supermeanvalued (Lemma 10.3 c)) and therefore \bar{h} is the *least supermeanvalued majorant* of h .

Similarly, suppose there exists a function \hat{h} such that

- (i) \hat{h} is a superharmonic majorant of h and
- (ii) if f is any other superharmonic majorant of h then $\hat{h} \leq f$.

Then \hat{h} is called *the least superharmonic majorant of h* .

Note that by (10.11) the least supermeanvalued majorant \bar{h} of h always exists ($\bar{h} \equiv \infty$ is possible). Moreover, if \bar{h} is lower semicontinuous, then \hat{h} exists and $\hat{h} = \bar{h}$. Later we will prove that if g is nonnegative (or lower bounded) and lower semicontinuous, then \hat{g} exists (Theorem 10.7).

Let $g \geq 0$ and let f be a supermeanvalued majorant of g . Then if τ is a stopping time

$$f(x) \geq E^x[f(X_\tau)] \geq E^x[g(X_\tau)].$$

So

$$f(x) \geq \sup_{\tau} E^x[g(X_\tau)] = g^*(x).$$

Therefore we always have

$$\hat{g}(x) \geq g^*(x) \quad \text{for all } x \in \mathbf{R}^n. \quad (10.12)$$

What is not so easy to see is that the converse inequality also holds, i.e. that in fact

$$\hat{g} = g^*. \quad (10.13)$$

We will prove this after we have established a useful iterative procedure for calculating \hat{g} . Before we give such a procedure let us introduce a concept which is related to superharmonic functions:

DEFINITION 10.5. A lower semicontinuous function $f : \mathbf{R}^n \rightarrow [0, \infty]$ is called *excessive* (w.r.t. X_t) if

$$f(x) \geq E^x[f(X_s)] \quad \text{for all } s \geq 0, x \in \mathbf{R}^n. \quad (10.14)$$

It is clear that a superharmonic function must be excessive. What is not so obvious, is that the converse also holds:

THEOREM 10.6. Let $f : \mathbf{R}^n \rightarrow [0, \infty]$. Then f is excessive w.r.t. X_t if and only if f is superharmonic w.r.t. X_t .

Proof in a special case. Let L be the differential operator associated to X as in (7.29), so that L coincides with the generator A of X on C_0^2 . We only prove the theorem in the special case when $f \in C^2(\mathbf{R}^n)$ and Lf is bounded: Then by Dynkin's formula we have

$$E^x[f(X_t)] = f(x) + E^x\left[\int_0^t Lf(X_r)dr\right] \quad \text{for all } t \geq 0,$$

so if f is excessive then $Lf \leq 0$. Therefore, if τ is a stopping time we get

$$E^x[f(X_{t \wedge \tau})] \leq f(x) \quad \text{for all } t \geq 0.$$

Letting $t \rightarrow \infty$ we see that f is superharmonic.

A proof in the general case can be found in Dynkin (1965 II, p. 5).

The first iterative procedure for the least superharmonic majorant \hat{g} of g is the following:

THEOREM 10.7. (Construction of the least superharmonic majorant). Let $g = g_0$ be a nonnegative, lower semicontinuous function on \mathbf{R}^n and define inductively

$$g_n(x) = \sup_{t \in S_n} E^x[g_{n-1}(X_t)], \quad (10.15)$$

where $S_n = \{k \cdot 2^{-n}; 0 \leq k \leq 4^n\}$, $n = 1, 2, \dots$. Then $g_n \uparrow \hat{g}$ and \hat{g} is the least superharmonic majorant of g . Moreover, $\hat{g} = \bar{g}$.

Proof. Note that $\{g_n\}$ is increasing. Define $\check{g}(x) = \lim_{n \rightarrow \infty} g_n(x)$. Then

$$\check{g}(x) \geq g_n(x) \geq E^x[g_{n-1}(X_t)] \text{ for all } n \text{ and all } t \in S_n.$$

Hence

$$\check{g}(x) \geq \lim_{n \rightarrow \infty} E^x[g_{n-1}(X_t)] = E^x[\check{g}(X_t)] \quad (10.16)$$

for all $t \in S = \bigcup_{n=1}^{\infty} S_n$.

Since \check{g} is an increasing limit of lower semicontinuous functions (Lemma 8.4) \check{g} is lower semicontinuous. Fix $t \in \mathbf{R}$ and choose $t_k \in S$ such that $t_k \rightarrow t$. Then by (10.16), the Fatou lemma and lower semicontinuity

$$\check{g}(x) \geq \lim_{k \rightarrow \infty} E^x[\check{g}(X_{t_k})] \geq E^x[\lim_{k \rightarrow \infty} \check{g}(X_{t_k})] \geq E^x[\check{g}(X_t)].$$

So \check{g} is an excessive function. Therefore \check{g} is superharmonic by Theorem 10.6 and hence \check{g} is a superharmonic majorant of g . On the other hand, if f is any supermeanvalued majorant of g , then clearly by induction

$$f(x) \geq g_n(x) \text{ for all } n$$

and so $f(x) \geq \check{g}(x)$. This proves that \check{g} is the least supermeanvalued majorant \bar{g} of g . So $\check{g} = \hat{g}$. \square

It is a consequence of Theorem 10.7 that we may replace the finite sets S_n by the whole interval $[0, \infty]$:

COROLLARY 10.8. Define $h_0 = g$ and inductively

$$h_n(x) = \sup_{t \geq 0} E^x[h_{n-1}(X_t)]; \quad n = 1, 2, \dots$$

Then $h_n \uparrow \hat{g}$.

Proof. Let $h = \lim h_n$. Then clearly $h \geq \check{g} = \hat{g}$. On the other hand, since \hat{g} is excessive we have

$$\hat{g}(x) \geq \sup_{t \geq 0} E^x[\hat{g}(X_t)].$$

So by induction

$$\hat{g} \geq h_n \text{ for all } n.$$

Thus $\hat{g} = h$ and the proof is complete.

We are now ready for our first main result on the optimal stopping problem. The following result is basically due to Dynkin (1963) (and, in a martingale context, Snell (1952)):

THEOREM 10.9. (Existence theorem for optimal stopping)

Let g^* denote the optimal reward and \hat{g} the least superharmonic majorant of a continuous reward function $g \geq 0$.

a) Then

$$g^*(x) = \hat{g}(x). \quad (10.17)$$

b) For $\epsilon > 0$ let

$$D_\epsilon = \{x; g(x) < \hat{g}(x) - \epsilon\}. \quad (10.18)$$

Suppose g is bounded. Then stopping at the first time τ_ϵ of exit from D_ϵ is close to being optimal, in the sense that

$$|g^*(x) - E^x[g(X_{\tau_\epsilon})]| < 2\epsilon \quad (10.19)$$

for all x .

c) For arbitrary continuous $g \geq 0$ let

$$D = \{x; g(x) < g^*(x)\} \quad (\text{the continuation region}). \quad (10.20)$$

For $N = 1, 2, \dots$ define $g_N = g \wedge N$, $D_N = \{x; g_N(x) < (g_N)^\wedge(x)\}$ and $\sigma_N = \tau_{D_N}$. Then $D_N \subset D_{N+1}$, $D_N \subset D \cap g^{-1}([0, N])$, $D = \bigcup_N D_N$. If $\sigma_N < \infty$ a.s. Q^x for all N then

$$g^*(x) = \lim_{N \rightarrow \infty} E^x[g(X_{\sigma_N})]. \quad (10.21)$$

d) In particular, if $\tau_D < \infty$ a.s. Q^x and the family $\{g(X_{\sigma_N})\}_N$ is uniformly integrable w.r.t. Q^x (Appendix C), then

$$g^*(x) = E^x[g(X_{\tau_D})]$$

and $\tau^* = \tau_D$ is an optimal stopping time.

Proof. First assume that g is bounded and define

$$\tilde{g}_\epsilon(x) = E^x[\hat{g}(X_{\tau_\epsilon})] \text{ for } \epsilon > 0. \quad (10.22)$$

Then \tilde{g}_ϵ is supermeanvalued by Lemma 10.3 f). We claim that

$$g(x) \leq \tilde{g}_\epsilon(x) + \epsilon \quad \text{for all } x. \quad (10.23)$$

To see this suppose

$$\beta := \sup_x \{g(x) - \tilde{g}_\epsilon(x)\} > \epsilon. \quad (10.24)$$

Then for all $\eta > 0$ we can find x_0 such that

$$g(x_0) - \tilde{g}_\epsilon(x_0) \geq \beta - \eta. \quad (10.25)$$

On the other hand, since $\tilde{g}_\epsilon + \beta$ is a supermeanvalued majorant of g , we have

$$\hat{g}(x_0) \leq \tilde{g}_\epsilon(x_0) + \beta. \quad (10.26)$$

Combining (10.25) and (10.26) we get

$$\hat{g}(x_0) \leq g(x_0) + \eta. \quad (10.27)$$

Consider the two possible cases:

Case 1: $\tau_\epsilon > 0$ a.s. Q^{x_0} . Then by (10.27) and the definition of D_ϵ

$$g(x_0) + \eta \geq \hat{g}(x_0) \geq E^{x_0}[\hat{g}(X_{t \wedge \tau_\epsilon})] \geq E^{x_0}[(g(X_t) + \epsilon) \mathcal{X}_{\{t < \tau_\epsilon\}}] \text{ for all } t > 0.$$

Hence by the Fatou lemma and lower semicontinuity of g

$$\begin{aligned} g(x_0) + \eta &\geq \liminf_{t \rightarrow 0} E^{x_0}[(g(X_t) + \epsilon) \mathcal{X}_{\{t < \tau_\epsilon\}}] \\ &\geq E^{x_0}[\liminf_{t \rightarrow 0} (g(X_t) + \epsilon) \mathcal{X}_{\{t < \tau_\epsilon\}}] \geq g(x_0) + \epsilon. \end{aligned}$$

This is a contradiction if $\eta < \epsilon$.

Case 2: $\tau_\epsilon = 0$ a.s. Q^{x_0} . Then $\tilde{g}_\epsilon(x_0) = \hat{g}(x_0)$, so $g(x_0) \leq \tilde{g}_\epsilon(x_0)$, contradicting (10.25) for $\eta \leq \beta$.

Therefore (10.24) leads to a contradiction. Thus (10.23) is proved and we conclude that $\tilde{g} + \epsilon$ is a supermeanvalued majorant of g . Therefore

$$\hat{g} \leq \tilde{g}_\epsilon + \epsilon = E[\hat{g}(X_{\tau_\epsilon})] + \epsilon \leq E[(g + \epsilon)(X_{\tau_\epsilon})] + \epsilon \leq g^* + 2\epsilon \quad (10.28)$$

and since ϵ was arbitrary we have by (10.12)

$$\hat{g} = g^*.$$

If g is not bounded, let

$$g_N = \min(N, g), \quad N = 1, 2, \dots$$

Then

$$g^* \geq (g_N)^* = (g_N)^\wedge \uparrow h \text{ as } N \rightarrow \infty, \quad \text{where } h \geq \hat{g}$$

since h is a superharmonic majorant of g . Thus $h = \hat{g} = g^*$ and this proves (10.17) for general g . From (10.28) and (10.17) we obtain (10.19).

Finally, to obtain c) and d) let us again first assume that g is bounded. Then, since

$$\tau_\epsilon \uparrow \tau_D \text{ as } \epsilon \downarrow 0$$

and $\tau_D < \infty$ a.s we have

$$E^x[g(X_{\tau_\epsilon})] \rightarrow E^x[g(X_{\tau_D})] \quad \text{as } \epsilon \downarrow 0, \quad (10.29)$$

and hence by (10.28) and (10.17)

$$g^*(x) = E^x[g(X_{\tau_D})] \text{ if } g \text{ is bounded.} \quad (10.30)$$

Finally, if g is not bounded define

$$h = \lim_{N \rightarrow \infty} (g_N)^\wedge.$$

Then h is superharmonic by Lemma 10.3 d) and since $(g_N)^\wedge \leq \hat{g}$ for all N we have $h \leq \hat{g}$. On the other hand $g_N \leq (g_N)^\wedge \leq h$ for all N and therefore $g \leq h$. Since \hat{g} is the least superharmonic majorant of g we conclude that

$$h = \hat{g}. \quad (10.31)$$

Hence by (10.30–31) we obtain (10.21):

$$g^*(x) = \lim_{N \rightarrow \infty} (g_N)^\wedge(x) = \lim_{N \rightarrow \infty} E^x[g_N(X_{\sigma_N})] \leq \lim_{N \rightarrow \infty} E^x[g(X_{\sigma_N})] \leq g^*(x).$$

Note that $(g_N)^\wedge \leq N$ everywhere, so if $g_N(x) < (g_N)^\wedge(x)$ then $g_N(x) < N$ and therefore $g(x) = g_N(x) < (g_N)^\wedge(x) \leq \hat{g}(x)$ and $g_{N+1}(x) = g_N(x) < (g_N)^\wedge(x) \leq (g_{N+1})^\wedge(x)$. Hence $D_N \subset D \cap \{x; g(x) < N\}$ and $D_N \subset D_{N+1}$ for all N . So by (10.31) we conclude that D is the increasing union of the sets $D_N; N = 1, 2, \dots$. Therefore

$$\tau_D = \lim_{N \rightarrow \infty} \sigma_N.$$

So by (10.21) and uniform integrability we have

$$\begin{aligned} \hat{g}(x) &= \lim_{N \rightarrow \infty} (g_N)^\wedge(x) = \lim_{N \rightarrow \infty} E^x[g_N(X_{\sigma_N})] \\ &= E^x[\lim_{N \rightarrow \infty} g_N(X_{\sigma_N})] = E^x[g(X_{\tau_D})], \end{aligned}$$

and the proof of Theorem 10.9 is complete.

REMARKS. 1) Note that the sets D, D_ϵ and D_N are open, since $\hat{g} = g^*$ is lower semicontinuous and g is continuous.

2) By inspecting the proof of a) we see that (10.17) holds under the weaker assumption that $g \geq 0$ is lower semicontinuous.

The following consequence of Theorem 10.9 is often useful:

COROLLARY 10.10. Suppose there exists a Borel set H such that

$$\tilde{g}_H(x) = E^x[g(X_{\tau_H})]$$

is supermeanvalued majorant of g . Then

$$g^*(x) = \tilde{g}_H(x), \text{ so } \tau^* = \tau_H \text{ is optimal.}$$

Proof. If \tilde{g}_H is a supermeanvalued majorant of g then clearly

$$\bar{g}(x) \leq \tilde{g}_H(x)$$

On the other hand we of course have

$$\tilde{g}_H(x) \leq \sup_{\tau} E^x[g(X_{\tau})] = g^*(x),$$

so $g^* = \tilde{g}_H$ by Theorem 10.7 and Theorem 10.9 a).

COROLLARY 10.11. Let

$$D = \{x; g(x) < \hat{g}(x)\}$$

and put

$$\tilde{g}(x) = \tilde{g}_D(x) = E^x[g(X_{\tau_D})].$$

If $\tilde{g} \geq g$ then $\tilde{g} = g^*$.

Proof. Since $X_{\tau_D} \notin D$ we have $g(X_{\tau_D}) \geq \hat{g}(X_{\tau_D})$ and therefore $g(X_{\tau_D}) = \hat{g}(X_{\tau_D})$, a.s. Q^x . So $\tilde{g}(x) = E^x[\hat{g}(X_{\tau_D})]$ is supermeanvalued since \hat{g} is, and the result follows from Corollary 10.10.

Theorem 10.9 gives a sufficient condition for the existence of an optimal stopping time τ^* . Unfortunately, τ^* need not exist in general. For example, if

$$X_t = t \text{ for } t \geq 0 \quad (\text{deterministic})$$

and

$$g(\xi) = \frac{\xi^2}{1 + \xi^2}; \quad \xi \in \mathbf{R}$$

then $g^*(x) = 1$, but there is no stopping time τ such that

$$E^x[g(X_{\tau})] = 1.$$

However, we can prove that if an optimal stopping time τ^* exists, then it is of the form given in Theorem 10.9:

THEOREM 10.12. (Uniqueness theorem for optimal stopping)

Define as before

$$D = \{x; g(x) < g^*(x)\} \subset \mathbf{R}^n.$$

Suppose there exists an optimal time τ^* for the problem (10.2). Then

$$\tau^* \geq \tau_D, \quad (10.33)$$

where τ_D is the first exit time from D of X_t and

$$g^*(x) = E^x[g(X_{\tau_D})] \text{ for all } x. \quad (10.34)$$

Hence τ_D is an optimal stopping time for the problem (10.2).

Moreover, every point $x \in \partial D$ is a *regular* boundary point for X_t .

Proof. Let τ be an \mathcal{F}_t -stopping time and assume $Q^x[\tau < \tau_D] > 0$. Since $g(X_\tau) < g^*(X_\tau)$ if $\tau < \tau_D$ and $g \leq g^*$ always, we have

$$\begin{aligned} E^x[g(X_\tau)] &= \int_{\tau < \tau_D} g(X_\tau) dQ^x + \int_{\tau \geq \tau_D} g(X_\tau) dQ^x \\ &< \int_{\tau < \tau_D} g^*(X_\tau) dQ^x + \int_{\tau \geq \tau_D} g^*(X_\tau) dQ^x \\ &= E^x[g^*(X_\tau)] \leq g^*(x), \end{aligned}$$

since g^* is superharmonic. This proves (10.33). To obtain (10.34) we note that since \hat{g} is superharmonic we have by Lemma 10.3 e)

$$\begin{aligned} g^*(x) &= E^x[g(X_{\tau^*})] \leq E^x[\hat{g}(X_{\tau^*})] \leq E^x[\hat{g}(X_{\tau_D})] \\ &= E^x[g(X_{\tau_D})] \leq g^*(x). \end{aligned}$$

Finally, to prove that ∂D is regular for X_t define

$$\hat{\tau}(x, \omega) = \begin{cases} 0 & \text{if } x \notin D \\ \tau_D(\omega) & \text{if } x \in D \end{cases}$$

Then

$$E^x[g(X_{\hat{\tau}})] = g(x) = g^*(x) \text{ for } x \notin D$$

and

$$E^x[g(X_{\hat{\tau}})] = E^x[g(X_{\tau_D})] = g^*(x) \text{ for } x \in D \text{ by (10.34).}$$

Hence $\hat{\tau}$ is optimal and therefore $\hat{\tau} \geq \tau_D$ by (10.33). Since clearly $\hat{\tau} \leq \tau_D$ we must have $\hat{\tau} = \tau_D$, a.s. In particular,

$$\tau_D = 0 \text{ a.s. } Q^x \text{ for all } x \in \partial D$$

and hence every $x \in \partial D$ is regular for X_t . \square

REMARK. The following observation is sometimes useful:

Let \mathcal{A} be the characteristic operator of X . Assume $g \in C^2(\mathbf{R}^n)$. Define

$$U = \{x; \mathcal{A}g(x) > 0\}. \quad (10.35)$$

Then, with D as before, (10.20),

$$U \subset D. \quad (10.36)$$

Consequently, from (10.33) we conclude that it is *never optimal to stop the process before it exits from U* . But there may be cases when $U \neq D$, so that it is optimal to proceed beyond U before stopping. (This is in fact the typical situation.) See e.g. Example 10.16.

To prove (10.36) choose $x \in U$ and let τ_0 be the first exit time from a bounded open set $W \ni x$, $W \subset U$. Then by Dynkin's formula, for $u > 0$

$$E^x[g(X_{\tau_0 \wedge u})] = g(x) + E^x\left[\int_0^{\tau_0 \wedge u} \mathcal{A}g(X_s)ds\right] > g(x)$$

so $g(x) < g^*(x)$ and therefore $x \in D$.

EXAMPLE 10.13. Let $X_t = B_t$ be a Brownian motion in \mathbf{R}^2 . Using that B_t is recurrent in \mathbf{R}^2 (Example 7.13) one can show that the only nonnegative superharmonic functions in \mathbf{R}^2 are the constants (Exercise 10.2).

Therefore

$$g^*(x) = \|g\|_\infty := \sup\{g(y); y \in \mathbf{R}^2\} \text{ for all } x.$$

So if g is unbounded then $g^* = \infty$ and no optimal stopping time exists. Assume therefore that g is bounded. The continuation region is

$$D = \{x; g(x) < \|g\|_\infty\},$$

so if ∂D is a *polar set* i.e. $\text{cap}(\partial D) = 0$, where cap denotes the *logarithmic capacity* (see Port and Stone (1979)), then $\tau_D = \infty$ a.s. and no optimal stopping exists. On the other hand, if $\text{cap}(\partial D) > 0$ then $\tau_D < \infty$ a.s. and

$$E^x[g(B_{\tau_D})] = \|g\|_\infty = g^*(x),$$

so $\tau^* = \tau_D$ is optimal.

EXAMPLE 10.14. The situation is different in \mathbf{R}^n for $n \geq 3$.

a) To illustrate this let $X_t = B_t$ be Brownian motion in \mathbf{R}^3 and let the reward function be

$$g(\xi) = \begin{cases} |\xi|^{-1} & \text{for } |\xi| \geq 1 \\ 1 & \text{for } |\xi| < 1 \end{cases}; \quad \xi \in \mathbf{R}^3.$$

Then g is superharmonic (in the classical sense) in \mathbf{R}^3 , so $g^* = g$ everywhere and the best policy is to stop immediately, no matter where the starting point is.

b) Let us change g to

$$h(x) = \begin{cases} |x|^{-\alpha} & \text{for } |x| \geq 1 \\ 1 & \text{for } |x| < 1 \end{cases}$$

for some $\alpha > 1$. Let $H = \{x; |x| > 1\}$ and define

$$\tilde{h}(x) = E^x[h(B_{\tau_H})] = P^x[\tau_H < \infty].$$

Then by Example 7.13

$$\tilde{h}(x) = \begin{cases} 1 & \text{if } |x| \leq 1 \\ |x|^{-1} & \text{if } |x| > 1, \end{cases}$$

i.e. $\tilde{h} = g$ (defined in a)), which is a superharmonic majorant of h . Therefore by Corollary 10.10

$$h^* = \tilde{h} = g,$$

$H = D$ and $\tau^* = \tau_H$ is an optimal stopping time.

Reward functions assuming negative values

The results we have obtained so far regarding the problem (10.2)–(10.3) are based on the assumptions (10.1). To some extent these assumptions can be relaxed, although neither can be removed completely. For example, we have noted that Theorem 10.9 a) still holds if $g \geq 0$ is only assumed to be lower semicontinuous.

The nonnegativity assumption on g can also be relaxed. First of all, note that if g is bounded below, say $g \geq -M$ where $M > 0$ is a constant, then we can put

$$g_1 = g + M \geq 0$$

and apply the theory to g_1 . Since

$$E^x[g(X_\tau)] = E^x[g_1(X_\tau)] - M \quad \text{if } \tau < \infty \text{ a.s.,}$$

we have $g^*(x) = g_1^*(x) - M$, so the problem can be reduced to the optimal stopping problem for the nonnegative function g_1 . (See Exercise 10.4.)

If g is not bounded below, then problem (10.2)–(10.3) is not well-defined unless

$$E^x[g^-(X_\tau)] < \infty \quad \text{for all } \tau \tag{10.37}$$

where

$$g^-(x) = -\min(g(x), 0).$$

If we assume that g satisfies the stronger condition that

the family $\{g^-(X_\tau); \tau \text{ stopping time}\}$ is uniformly integrable (10.38)

then basically all the theory from the nonnegative case carries over. We refer to the reader to Shirayev (1978) for more information. See also Theorem 10.18.

The time-inhomogeneous case

Let us now consider the case when the reward function g depends on both time and space, i.e.

$$g = g(t, x) : \mathbf{R} \times \mathbf{R}^n \rightarrow [0, \infty), \quad g \text{ is continuous.} \quad (10.39)$$

Then the problem is to find $g_0(x)$ and τ^* such that

$$g_0(x) = \sup_{\tau} E^x[g(\tau, X_\tau)] = E^x[g(\tau^*, X_{\tau^*})]. \quad (10.40)$$

To reduce this case to the original case (10.2)–(10.3) we proceed as follows:

Suppose the Ito diffusion $X_t = X_t^x$ has the form

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t; \quad t \geq 0, \quad X_0 = x$$

where $b : \mathbf{R}^n \rightarrow \mathbf{R}^n$ and $\sigma : \mathbf{R}^n \rightarrow \mathbf{R}^{n \times m}$ are given functions satisfying the conditions of Theorem 5.5 and B_t is m -dimensional Brownian motion. Define the Ito diffusion $Y_t = Y_t^{(s,x)}$ in \mathbf{R}^{n+1} by

$$Y_t = \begin{bmatrix} s+t \\ X_t^x \end{bmatrix}; \quad t \geq 0 \quad (10.41)$$

Then

$$dY_t = \begin{bmatrix} 1 \\ b(X_t) \end{bmatrix} dt + \begin{bmatrix} 0 \\ \sigma(X_t) \end{bmatrix} dB_t = \hat{b}(Y_t)dt + \hat{\sigma}(Y_t)dB_t \quad (10.42)$$

where

$$\hat{b}(\eta) = \hat{b}(t, \xi) = \begin{bmatrix} 1 \\ b(\xi) \end{bmatrix} \in \mathbf{R}^{n+1}, \quad \hat{\sigma}(\eta) = \hat{\sigma}(t, \xi) = \begin{bmatrix} 0 \cdots 0 \\ \dots \\ \sigma(\xi) \end{bmatrix} \in \mathbf{R}^{(n+1) \times m},$$

with $\eta = (t, \xi) \in \mathbf{R} \times \mathbf{R}^n$.

So Y_t is an Ito diffusion starting at $y = (s, x)$. Let $R^y = R^{(s,x)}$ denote the probability law of $\{Y_t\}$ and let $E^y = E^{(s,x)}$ denote the expectation w.r.t. R^y . In terms of Y_t the problem (10.40) can be written

$$g_0(x) = g^*(0, x) = \sup_{\tau} E^{(0,x)}[g(Y_{\tau})] = E^{(0,x)}[g(Y_{\tau^*})] \quad (10.43)$$

which is a special case of the problem

$$g^*(s, x) = \sup_{\tau} E^{(s,x)}[g(Y_{\tau})] = E^{(s,x)}[g(Y_{\tau^*})], \quad (10.44)$$

which is of the form (10.2)–(10.3) with X_t replaced by Y_t .

Note that the characteristic operator $\hat{\mathcal{A}}$ of Y_t is given by

$$\hat{\mathcal{A}}\phi(s, x) = \frac{\partial \phi}{\partial s}(s, x) + \mathcal{A}\phi(s, x); \quad \phi \in C^2(\mathbf{R} \times \mathbf{R}^n) \quad (10.45)$$

where \mathcal{A} is the characteristic operator of X_t (working on the x -variables).

EXAMPLE 10.15. Let $X_t = B_t$ be 1-dimensional Brownian motion and let the reward function be

$$g(t, \xi) = e^{-\alpha t + \beta \xi}, \quad \xi \in \mathbf{R}$$

where $\alpha, \beta \geq 0$ are constants. The characteristic operator $\hat{\mathcal{A}}$ of $Y_t^{s,x} = [s+t, B_t^x]$ is given by

$$\hat{\mathcal{A}}f(s, x) = \frac{\partial f}{\partial s} + \frac{1}{2} \cdot \frac{\partial^2 f}{\partial x^2}; \quad f \in C^2.$$

Thus

$$\mathcal{A}g = (-\alpha + \frac{1}{2}\beta^2)g,$$

so if $\beta^2 \leq 2\alpha$ then $g^* = g$ and the best policy is to stop immediately. If $\beta^2 > 2\alpha$ we have

$$U := \{(s, x); \hat{\mathcal{A}}g(s, x) > 0\} = \mathbf{R}^2$$

and therefore by (10.36) $D = \mathbf{R}^2$ and hence τ^* does not exist. If $\beta^2 > 2\alpha$ we can use Theorem 10.7 to prove that $g^* = \infty$:

$$\begin{aligned} \sup_{t \in S_n} E^{(s,x)}[g(Y_t)] &= \sup_{t \in S_n} E[e^{-\alpha(s+t) + \beta B_t^x}] \\ &= \sup_{t \in S_n} [e^{-\alpha(s+t)} \cdot e^{\beta x + \frac{1}{2}\beta^2 t}] \quad (\text{see the remark following (5.6)}) \\ &= \sup_{t \in S_n} g(s, x) \cdot e^{(-\alpha + \frac{1}{2}\beta^2)t} = g(s, x) \cdot \exp((-\alpha + \frac{1}{2}\beta^2)2^n), \end{aligned}$$

so $g_n(s, x) \rightarrow \infty$ as $n \rightarrow \infty$.

Hence no optimal stopping exists in this case.

EXAMPLE 10.16. (When is the right time to sell the stocks?)

We now return to a specified version of Problem 5 in the introduction:

Suppose the price X_t at time t of a person's assets (e.g. a house, stocks, oil ...) varies according to a stochastic differential equation of the form

$$dX_t = rX_t dt + \alpha X_t dB_t, X_0 = x > 0,$$

where B_t is 1-dimensional Brownian motion and r, α are known constants. (The problem of estimating α and r from a series of observations can be approached using the quadratic variation $\langle X, X \rangle_t$ of the process $\{X_t\}$ (Exercise 4.7) and filtering theory (Example 6.13), respectively. Suppose that connected to the sale of the assets there is a fixed fee/tax or transaction cost $a > 0$. Then if the person decides to sell at time t the discounted net of the sale is

$$e^{-\rho t}(X_t - a),$$

where $\rho > 0$ is given discounting factor. The problem is to find a stopping time τ that maximizes

$$E^{(s,x)}[e^{-\rho\tau}(X_\tau - a)] = E^{(s,x)}[g(\tau, X_\tau)],$$

where

$$g(t, \xi) = e^{-\rho t}(\xi - a).$$

The characteristic operator $\hat{\mathcal{A}}$ of the process $Y_t = (s+t, X_t)$ is given by

$$\hat{\mathcal{A}}f(s, x) = \frac{\partial f}{\partial s} + rx \frac{\partial f}{\partial x} + \frac{1}{2}\alpha^2 x^2 \frac{\partial^2 f}{\partial x^2}; \quad f \in C^2(\mathbf{R}^2).$$

Hence $\hat{\mathcal{A}}g(s, x) = -\rho e^{-\rho s}(x - a) + rxe^{-\rho s} = e^{-\rho s}((r - \rho)x + \rho a)$. So

$$U := \{(s, x); \hat{\mathcal{A}}g(s, x) > 0\} = \begin{cases} \mathbf{R} \times \mathbf{R}_+ & \text{if } r \geq \rho \\ \{(s, x); x < \frac{a\rho}{\rho-r}\} & \text{if } r < \rho \end{cases}$$

So if $r \geq \rho$ we have $U = D = \mathbf{R} \times \mathbf{R}_+$ so τ^* does not exist. If $r > \rho$ then $g^* = \infty$ while if $r = \rho$ then

$$g^*(s, x) = xe^{-\rho s}.$$

(The proofs of these statements are left as Exercise 10.5.)

It remains to examine the case $r < \rho$. First we establish that the region D must be invariant w.r.t. t , in the sense that

$$D + (t_0, 0) = D \quad \text{for all } t_0. \tag{10.46}$$

To prove (10.46) consider

$$\begin{aligned} D + (t_0, 0) &= \{(t + t_0, x); (t, x) \in D\} = \{(s, x); (s - t_0, x) \in D\} \\ &= \{(s, x); g(s - t_0, x) < g^*(s - t_0, x)\} = \{(s, x); e^{\rho t_0} g(s, x) < e^{\rho t_0} g^*(s, x)\} \\ &= \{(s, x); g(s, x) < g^*(s, x)\} = D, \end{aligned}$$

where we have used that

$$\begin{aligned} g^*(s - t_0, x) &= \sup_{\tau} E^{(s-t_0, x)}[e^{-\rho\tau}(X_{\tau} - a)] = \sup_{\tau} E[e^{-\rho(\tau+(s-t_0))}(X_{\tau}^x - a)] \\ &= e^{\rho t_0} \sup_{\tau} E[e^{-\rho(\tau+s)}(X_{\tau}^x - a)] = e^{\rho t_0} g^*(s, x). \end{aligned}$$

Therefore the connected component of D that contains U must have the form

$$D(x_0) = \{(t, x); 0 < x < x_0\} \quad \text{for some } x_0 \geq \frac{a\rho}{\rho - r}.$$

Note that D cannot have any other components, for if V is a component of D disjoint from U then $\hat{\mathcal{A}}g < 0$ in V and so, if $y \in V$,

$$E^y[g(Y_{\tau})] = g(y) + E^y[\int_0^{\tau} \hat{\mathcal{A}}g(Y_t)dt] < g(y)$$

for all exit times τ bounded by the exit time from an x -bounded strip in V . From this we conclude by Theorem 10.9 c) that $g^*(y) = g(y)$, which implies $V = \emptyset$.

Put $\tau(x_0) = \tau_{D(x_0)}$ and let us compute

$$\tilde{g}(s, x) = \tilde{g}_{x_0}(s, x) = E^{(s, x)}[g(Y_{\tau(x_0)})]. \quad (10.47)$$

From Chapter IX we know that $f = \tilde{g}$ is the solution of the boundary value problem

$$\begin{aligned} \frac{\partial f}{\partial s} + rx \frac{\partial f}{\partial x} + \frac{1}{2}\alpha^2 x^2 \frac{\partial^2 f}{\partial x^2} &= 0 \quad \text{for } 0 < x < x_0 \\ f(s, x_0) &= e^{-\rho s}(x_0 - a). \end{aligned} \quad (10.48)$$

(Note that $\mathbf{R} \times \{0\}$ does not contain any regular boundary points of D w.r.t. $Y_t = (s+t, X_t)$.)

If we try a solution of (10.48) of the form

$$f(s, x) = e^{-\rho s}\phi(x)$$

we get the following 1-dimensional problem

$$\begin{aligned} -\rho\phi + rx\phi'(x) + \frac{1}{2}\alpha^2 x^2 \phi''(x) &= 0 \quad \text{for } 0 < x < x_0 \\ \phi(x_0) &= x_0 - a \end{aligned} \quad (10.49)$$

The general solution ϕ of (10.49) is

$$\phi(x) = C_1 x^{\gamma_1} + C_2 x^{\gamma_2},$$

where C_1, C_2 are arbitrary constants and

$$\gamma_i = \alpha^{-2} \left[\frac{1}{2} \alpha^2 - r \pm \sqrt{\left(r - \frac{1}{2} \alpha^2 \right)^2 + 2\rho\alpha^2} \right] \quad (i = 1, 2), \quad \gamma_2 < 0 < \gamma_1.$$

Since $\phi(x)$ is bounded as $x \rightarrow 0$ we must have $C_2 = 0$ and the boundary requirement $\phi(x_0) = x_0 - a$ gives $C_1 = x_0^{-\gamma_1}(x_0 - a)$. We conclude that the bounded solution f of (10.48) is

$$\tilde{g}_{x_0}(s, x) = f(s, x) = e^{-\rho s} (x_0 - a) \left(\frac{x}{x_0} \right)^{\gamma_1}. \quad (10.50)$$

If we fix (s, x) then the value of x_0 which maximizes $\tilde{g}_{x_0}(s, x)$ is easily seen to be given by

$$x_0 = x_{\max} = \frac{a\gamma_1}{\gamma_1 - 1} \quad (10.51)$$

(note that $\gamma_1 > 1$ if and only if $r < \rho$).

So by (10.47–51) we can conclude that

$$\begin{aligned} g^*(s, x) &= \sup_{\tau} E^{(s, x)}[g(\tau, X_{\tau})] = \sup_{x_0} E^{(s, x)}[g(\tau(x_0), X_{\tau(x_0)})] \\ &= \sup_{x_0} \tilde{g}_{x_0}(s, x) = \tilde{g}_{x_{\max}}(s, x). \end{aligned}$$

The conclusion is therefore that one should sell the assets the first time the price of them reaches the value $x_{\max} = \frac{a\gamma_1}{\gamma_1 - 1}$. The expected discounted profit obtained from this strategy is

$$g^*(s, x) = e^{-\rho s} \left(\frac{\gamma_1 - 1}{a} \right)^{\gamma_1 - 1} \left(\frac{x}{\gamma_1} \right)^{\gamma_1}.$$

REMARK. The reader is invited to check that the value $x_0 = x_{\max}$ is the only value of x_0 which makes the function

$$x \rightarrow \tilde{g}_{x_0}(s, x) \quad (\text{given by (10.47)})$$

continuously differentiable at x_0 . This is not a coincidence. In fact, it illustrates a general phenomenon which is known as the *high contact* (or smooth fit) principle. See Samuelson (1965), McKean (1965), Bather (1970) and Shirayev (1978). This principle is the basis of the fundamental connection between optimal stopping and *variational inequalities*. Later in this chapter we will discuss some aspects of this connection. More information can be found in Bensoussan and Lions (1978) and Friedman (1976). See also Brekke and Øksendal (1991).

Optimal stopping problems involving an integral

Let

$$dY_t = b(Y_t)dt + \sigma(Y_t)dB_t, \quad Y_0 = y \quad (10.52)$$

be an Ito diffusion in \mathbf{R}^k . Let $g : \mathbf{R}^k \rightarrow [0, \infty)$ be continuous and let $f : \mathbf{R}^k \rightarrow [0, \infty)$ be Lipschitz continuous with at most linear growth. (These conditions can be relaxed. See (10.38) and Theorem 10.18.) Consider the optimal stopping problem:

Find $G^*(y)$ and τ^* such that

$$G^*(y) = \sup_{\tau} E^y \left[\int_0^{\tau} f(Y_t) dt + g(Y_{\tau}) \right] = E^y \left[\int_0^{\tau^*} f(Y_t) dt + g(Y_{\tau^*}) \right] \quad (10.53)$$

This problem can be reduced to our original problem (10.2)–(10.3) by proceeding as follows: Define the Ito diffusion Z_t in \mathbf{R}^{k+1} by

$$dZ_t = \begin{bmatrix} dY_t \\ dW_t \end{bmatrix} := \begin{bmatrix} b(Y_t) \\ f(Y_t) \end{bmatrix} dt + \begin{bmatrix} \sigma(Y_t) \\ 0 \end{bmatrix} dB_t; \quad Z_0 = z = (y, w). \quad (10.54)$$

Then we see that

$$G^*(y) = \sup_{\tau} E^{(y,0)} [W_{\tau} + g(Y_t)] = \sup_{\tau} E^{(y,0)} [G(Y_t)] \quad (10.55)$$

with

$$G(z) := G(y, w) := g(y) + w; \quad z = (y, w) \in \mathbf{R}^k \times \mathbf{R}. \quad (10.56)$$

This is again a problem of the type (10.2)–(10.3) with X_t replaced by Z_t and g replaced by G . Note that the connection between the characteristic operators \mathcal{A}_Y of Y_t and \mathcal{A}_Z of Z_t is given by

$$\mathcal{A}_Z \phi(z) = \mathcal{A}_Z \phi(y, w) = \mathcal{A}_Y \phi(y, w) + f(y) \frac{\partial \phi}{\partial w}, \quad \phi \in C^2(\mathbf{R}^{k+1}). \quad (10.57)$$

In particular, if $G(y, w) = g(y) + w \in C^2(\mathbf{R}^{k+1})$ then

$$\mathcal{A}_Z G(y, w) = \mathcal{A}_Y g(y, w) + f(y). \quad (10.58)$$

EXAMPLE 10.17. Consider the optimal stopping problem

$$\gamma(x) = \sup_{\tau} E^x \left[\int_0^{\tau} \theta e^{-\rho t} X_t dt + e^{-\rho \tau} X_{\tau} \right],$$

where

$$dX_t = \alpha X_t dt + \beta X_t dB_t; \quad X_0 = x > 0$$

is geometric Brownian motion (α, β, θ constants). We put

$$dY_t = \begin{bmatrix} dt \\ dX_t \end{bmatrix} = \begin{bmatrix} 1 \\ \alpha X_t \end{bmatrix} dt + \begin{bmatrix} 0 \\ \beta X_t \end{bmatrix} dB_t; \quad Y_0 = (s, x)$$

and

$$dZ_t = \begin{bmatrix} dY_t \\ dW_t \end{bmatrix} = \begin{bmatrix} 1 \\ \alpha X_t \\ e^{-\rho t} X_t \end{bmatrix} dt + \begin{bmatrix} 0 \\ \beta X_t \\ 0 \end{bmatrix} dB_t; \quad Z_0 = (s, x, w).$$

Then with

$$\begin{aligned} f(y) &= f(s, x) = \theta e^{-\rho s} x, & g(y) &= e^{-\rho s} x \quad \text{and} \\ G(s, x, w) &= g(s, x) + w = e^{-\rho s} x + w \end{aligned}$$

we have

$$\mathcal{A}_Z G = \frac{\partial G}{\partial s} + \alpha x \frac{\partial G}{\partial x} + \frac{1}{2} \beta^2 x^2 \frac{\partial^2 G}{\partial x^2} + \theta e^{-\rho s} x \frac{\partial G}{\partial w} = (-\rho + \alpha + \theta) e^{-\rho s} x.$$

Hence

$$U = \{(s, x, w); \mathcal{A}_Z G(s, x, w) > 0\} = \begin{cases} \mathbf{R}^3 & \text{if } \rho < \alpha + \theta \\ \emptyset & \text{if } \rho \geq \alpha + \theta \end{cases} \quad (10.59)$$

From this we conclude (see Exercise 10.6):

$$\text{If } \rho < \alpha + \theta \text{ then } \tau^* \text{ does not exist and } G^*(x) = \infty. \quad (10.60)$$

$$\text{If } \rho \geq \alpha + \theta \text{ then } \tau^* = 0 \text{ and } G^*(s, x, w) = G(s, x, w) = e^{-\rho s} x + w. \quad (10.61)$$

Connection with variational inequalities

The 'high contact principle' says, roughly, that – under certain conditions – the solution g^* of (10.2)–(10.3) is a C^1 function on \mathbf{R}^n if $g \in C^2(\mathbf{R}^n)$. This is a useful information which can help us to determine g^* . Indeed, this principle is so useful that it is frequently applied in the literature also in cases where its validity has not been rigorously proved.

Fortunately it turns out to be easy to prove a *sufficiency* condition of high contact type, i.e. a kind of verification theorem for optimal stopping, which makes it easy to verify that a given candidate for g^* (that we may have found by guessing or intuition) is actually equal to g^* . The result below is a simplified variant of a result in Brekke and Øksendal (1991):

In the following we fix a domain V in \mathbf{R}^k and we let

$$dY_t = b(Y_t)dt + \sigma(Y_t)dB_t; \quad Y_0 = y \quad (10.62)$$

be an Ito diffuction in \mathbf{R}^k . Define

$$T = T(y, \omega) = \inf\{t > 0; Y_t(\omega) \notin V\}. \quad (10.63)$$

Let $f : \mathbf{R}^k \rightarrow \mathbf{R}$, $g : \mathbf{R}^k \rightarrow \mathbf{R}$ be continous functions satisfying

$$(a) \quad E^y \left[\int_0^T |f(Y_t)| dt \right] < \infty \quad \text{for all } y \in \mathbf{R}^k \quad (10.64)$$

and

- (b) the family $\{g^-(Y_\tau); \tau \text{ stopping time, } \tau \leq T\}$ is uniformly integrable w.r.t. R^y (the probability law of Y_t), for all $y \in \mathbf{R}^k$. (10.65)

Consider the following problem: Find $\Phi(y)$ and $\tau^* \leq T$ such that

$$\Phi(y) = \sup_{\tau \leq T} J^\tau(y) = J^{\tau^*}(y), \quad (10.66)$$

where

$$J^\tau(y) = E^y \left[\int_0^\tau f(Y_t) dt + g(Y_\tau) \right] \quad \text{for } \tau \leq T.$$

Note that since $J^0(y) = g(y)$ we have

$$\Phi(y) \geq g(y) \quad \text{for all } y \in V. \quad (10.67)$$

We can now formulate the variational inequalities. As usual we let

$$L = L_Y = \sum_{i=1}^k b_i(y) \frac{\partial}{\partial y_i} + \frac{1}{2} \sum_{i,j=1}^k (\sigma \sigma^T)_{ij}(y) \frac{\partial^2}{\partial y_i \partial y_j}$$

be the partial differential operator which coincides with the generator A_Y of Y_t on $C_0^2(\mathbf{R}^k)$.

THEOREM 10.18. (Variational inequalities for optimal stopping)

Suppose we can find a function $\phi : V \rightarrow \mathbf{R}$ such that

- (i) $\phi \in C^1(V)$
- (ii) $\phi \geq g$ on V

Define

$$D = \{x \in V; \phi(x) > g(x)\}.$$

Suppose Y_t spends 0 time on ∂D a.s., i.e.

$$(iii) \quad E^y \left[\int_0^T \mathcal{X}_{\partial D}(Y_t) dt \right] = 0 \quad \text{for all } y \in V.$$

Moreover, suppose the following:

- (iv) $\phi \in C^2(V \setminus \partial D)$
- (v) $L\phi + f \leq 0$ on $V \setminus \overline{D}$
- (vi) $L\phi + f = 0$ on D
- (vii) $\tau_D := \inf\{t > 0; Y_t \notin D\} < \infty$ a.s. Q^y for all $y \in V$
and
- (viii) the family $\{\phi(X_\tau); \tau \leq \tau_D\}$ is uniformly integrable w.r.t. Q^y , for all $y \in V$.

Then

$$\phi(y) = \Phi(y) = \sup_{\tau \leq T} E^y \left[\int_0^\tau f(Y_t) dt + g(Y_\tau) \right]; \quad y \in V \quad (10.68)$$

and

$$\tau^* = \tau_D \quad (10.69)$$

is an optimal stopping time for this problem.

Proof. We can find a sequence of functions $\phi_j \in C^2(V)$, $j = 1, 2, \dots$, such that

- (a) $\phi_j \geq \phi$ on V
- (b) $\phi_j = \phi$ outside $H_j := \{y \in V; d(y, \partial D) < \frac{1}{j}\}$
- (c) $\{L\phi_j + f\}_{j=1}^\infty$ is upper bounded on V .

Let $K = \sup\{L\phi_j(y) + f(y); y \in V, j = 1, 2, \dots\}$. For $R > 0$ put

$T_R = \min(R, \inf\{t > 0; |Y_t| \geq R\})$ and let $\tau \leq T$ be a stopping time. Then by Dynkin's formula

$$\begin{aligned} E^y[\phi_j(Y_{\tau \wedge T_R})] &= \phi_j(y) + E^y \left[\int_0^{\tau \wedge T_R} L\phi_j(Y_t) dt \right] \\ &\leq \phi_j(y) - E^y \left[\int_0^{\tau \wedge T_R} f(Y_t) dt \right] + KE^y \left[\int_0^R \chi_{H_j}(Y_t) dt \right] \end{aligned} \quad (10.70)$$

By dominated convergence

$$E^y \left[\int_0^R \chi_{H_j}(Y_t) dt \right] \rightarrow E^y \left[\int_0^R \chi_{\partial D}(Y_t) dt \right] = 0 \quad \text{as } j \rightarrow \infty. \quad (10.71)$$

If $y \notin \partial D$ then $y \notin H_j$ and hence $\phi_j(y) = \phi(y)$ for j large enough. Therefore, by (10.65), (10.70) and (10.71),

$$\phi(y) \geq \lim_{j \rightarrow \infty} E^y \left[\int_0^{\tau \wedge T_R} f(Y_t) dt + \phi_j(Y_{\tau \wedge T_R}) \right] \geq E^y \left[\int_0^{\tau \wedge T_R} f(Y_t) dt + g(Y_{\tau \wedge T_R}) \right].$$

Hence by the Fatou lemma and (10.64), (10.65)

$$\phi(y) \geq \lim_{R \rightarrow \infty} E^y \left[\int_0^{\tau \wedge T_R} f(Y_t) dt + g(Y_{\tau \wedge T_R}) \right] \geq E^y \left[\int_0^\tau f(Y_t) dt + g(Y_\tau) \right].$$

Since $\tau \leq T$ was arbitrary, we conclude that

$$\phi(y) \geq \Phi(y) \quad \text{for all } y \in V \setminus \partial D.$$

So by lower semicontinuity

$$\phi(y) \geq \Phi(y) \quad \text{for all } y \in V. \quad (10.72)$$

If $y \notin D$ then $\phi(y) = g(y) \leq \Phi(y)$ so by (10.72) we have

$$\phi(y) = g^*(y) \quad \text{and} \quad \hat{\tau} = \hat{\tau}(y, \omega) := 0 \quad \text{is optimal for } y \notin D. \quad (10.73)$$

If $y \in D$ choose j_0 such that $y \notin H_{j_0}$ and define

$$\tau^{(j)} = \inf\{t > 0; Y_t \notin D \setminus \overline{H}_j\}; \quad j = 1, 2, \dots$$

By Dynkin's formula we have for $j \geq j_0$, using (vi),

$$E^y[\phi(Y_{\tau^{(j)} \wedge T_R})] = \phi(y) + E^y\left[\int_0^{\tau^{(j)} \wedge T_R} L\phi(Y_t)dt\right] = \phi(y) - E^y\left[\int_0^{\tau^{(j)} \wedge T_R} f(Y_t)dt\right].$$

So by uniform integrability and (vi), (vii) we get

$$\begin{aligned} \phi(y) &= \lim_{R, j \rightarrow \infty} E^y\left[\int_0^{\tau^{(j)} \wedge T_R} f(Y_t)dt + \phi(Y_{\tau^{(j)} \wedge T_R})\right] \\ &= E^y\left[\int_0^{\tau_D} f(Y_t)dt + g(Y_{\tau_D})\right] = J^{\tau_D}(y) \leq \Phi(y). \end{aligned} \quad (10.74)$$

Combining (10.72) and (10.74) we get

$$\phi(y) \geq \Phi(y) \geq J^{\tau_D}(y) = \phi(y)$$

so

$$\phi(y) = \Phi(y) \quad \text{and} \quad \hat{\tau}(y, \omega) := \tau_D \quad \text{is optimal when } y \in D. \quad (10.75)$$

From (10.73) and (10.75) we conclude that

$$\phi(y) = \Phi(y) \quad \text{for all } y \in V.$$

Moreover, the stopping time τ defined by

$$\hat{\tau}(y, \omega) = \begin{cases} 0 & \text{for } y \notin D \\ \tau_D & \text{for } y \in D \end{cases}$$

is optimal. By Theorem 10.12 we have $\hat{\tau} \geq \tau_D$. Since obviously $\hat{\tau} \leq \tau_D$ we conclude that $\hat{\tau} = \tau_D$. \square

EXAMPLE 10.19. To illustrate Theorem 10.18 let us apply it to reconsider Example 10.16:

Rather than proving (10.46) and the following properties of D , we now simply guess/assume that D has the form

$$D = \{(s, x); 0 < x < x_0\}$$

for some $x_0 > 0$, which is intuitively reasonable. Then we solve (10.48) for arbitrary x_0 and we arrive at the following candidate ϕ for g^* :

$$\phi(s, x) = \begin{cases} e^{-\rho s}(x_0 - a)(\frac{x}{x_0})^{\gamma_1} & \text{for } 0 < x < x_0 \\ e^{-\rho s}(x - a) & \text{for } x \geq x_0. \end{cases}$$

The requirement that $\phi \in C^1$ (Theorem 10.18 (i)) gives the value (10.51) for x_0 . It is clear that $\phi \in C^2$ outside ∂D and by construction $L\phi = 0$ on D . Moreover, conditions (iii), (vii) and (viii) clearly hold. It remains to verify that

- (ii) $\phi(s, x) \geq g(s, x)$ for $0 < x < x_0$, i.e. $\phi(s, x) \geq e^{-\rho s}(x - a)$ for $0 < x < x_0$
and
- (v) $L\phi(s, x) \leq 0$ for $x > x_0$, i.e. $Lg(s, x) \leq 0$ for $x > x_0$.

This is easily done by direct calculation (assuming $r < \rho$).

We conclude that $\phi = g^*$ and $\tau^* = \tau_D$ is optimal (with the value (10.51) for x_0).

EXERCISES

10.1.

In each of the optimal stopping problems below find the supremum g^* and – if it exists – an optimal stopping time τ^* . (Here B_t denotes 1-dimensional Brownian motion)

a) $g^*(x) = \sup_{\tau} E^x[B_{\tau}^2]$

b) $g^*(x) = \sup_{\tau} E^x[|B_{\tau}|^p],$
where $p > 0$.

c) $g^*(x) = \sup_{\tau} E^x[e^{-B_{\tau}^2}]$

d) $g^*(s, x) = \sup_{\tau} E^{(s, x)}[e^{-\rho \tau} \cosh B_{\tau}]$
where $\rho > 0$ and $\cosh x = \frac{1}{2}(e^x + e^{-x})$.

10.2.

- a) Prove that the only nonnegative (B_t -) superharmonic functions in \mathbf{R}^2 are the constants.

(Hint: Suppose u is a nonnegative superharmonic function and that there exist $x, y \in \mathbf{R}^2$ such that

$$u(x) < u(y).$$

Consider

$$E^x[u(B_\tau)],$$

where τ is the first hitting time for B_t of a small disc centered at y).

- b) Prove that the only nonnegative superharmonic functions in \mathbf{R} are the constants and use this to find $g^*(x)$ when

$$g(x) = \begin{cases} xe^{-x} & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases}$$

- c) Let $\gamma \in \mathbf{R}, n \geq 3$ and define, for $x \in \mathbf{R}^n$,

$$f_\gamma(x) = \begin{cases} |x|^\gamma & \text{for } |x| \geq 1 \\ 1 & \text{for } |x| < 1 \end{cases}$$

For what values of γ is $f_\gamma(\cdot)$ (B_t -) harmonic for $|x| > 1$? Prove that f_γ is superharmonic in \mathbf{R}^n iff $\gamma \in [2 - n, 0]$.

10.3.

Find g^*, τ^* such that

$$g^*(s, x) = \sup_{\tau} E^{(s, x)}[e^{-\rho\tau} B_\tau^2] = E^{(s, x)}[e^{-\rho\tau^*} B_{\tau^*}^2],$$

where B_t is 1-dimensional Brownian motion, $\rho > 0$ is constant.

Hint: First assume that the continuation region has the form

$$D = \{(s, x); -x_0 < x < x_0\}$$

for some x_0 and then try to determine x_0 . Then apply Theorem 10.18.

10.4.

Let X_t be an Ito diffusion on \mathbf{R}^n and $g : \mathbf{R}^n \rightarrow \mathbf{R}^+$ a continuous reward function. Define

$$g^\diamond(x) = \sup\{E^x[g(X_\tau)]; \tau \text{ stopping time, } E^x[\tau] < \infty\}.$$

Show that $g^\diamond = g^*$.

(Hint: If τ is a stopping time put $\tau_k = \tau \wedge k$ for $k = 1, 2, \dots$ and consider

$$E^x[g(X_\tau) \cdot \chi_{\tau < \infty}] \leq E^x[\liminf_{k \rightarrow \infty} g(X_{\tau_k})])$$

10.5.

With g, r, ρ as in Example 10.16 prove that

- a) if $r > \rho$ then $g^* = \infty$,
- b) if $r = \rho$ then $g^*(s, x) = xe^{-\rho s}$.

10.6.

Prove statements (10.60), (10.61) in Example 10.17.

10.7.

As a supplement to Exercise 10.4 it is worth noting that if g is not bounded below then the two problems

$$g^*(x) = \sup\{E^x[g(X_\tau)]; \tau \text{ stopping time}\}$$

and

$$g^\diamond(x) = \sup\{E^x[g(X_\tau)]; \tau \text{ stopping time, } E^x[\tau] < \infty\}$$

need not have the same solution. For example, if $g(x) = x, X_t = B_t \in \mathbf{R}$ prove that

$$g^*(x) = \infty \quad \text{for all } x \in \mathbf{R}$$

while

$$g^\diamond(x) = x \quad \text{for all } x \in \mathbf{R}.$$

(See Exercise 7.4.)

10.8.

Give an example with g not bounded below where Theorem 10.9 a) fails.
(Hint: See Exercise 10.7.)

10.9.

Solve the optimal stopping problem

$$\gamma(x) = \sup_{\tau} E^x \left[\int_0^{\tau} e^{-\rho t} B_t^2 dt + e^{-\rho \tau} B_\tau^2 \right].$$

10.10.

Prove the following simple, but useful, observation, which can be regarded as an extension of (10.36):

Let $W = \{(s, x); \exists \tau \text{ with } g(s, x) < E^{(s,x)}[g(\tau, X_\tau)]\}$.

Then $W \subset D$.

10.11.

Consider the optimal stopping problem

$$g^*(s, x) = \sup_{\tau} E^{(s, x)}[e^{-\rho\tau} B_{\tau}^+],$$

where $B_t \in \mathbf{R}$ and $x^+ = \max\{x, 0\}$.

- a) Use the argument for (10.46) and Exercise 10.10 to prove that the continuation region D has the form

$$D = \{(s, x); x < x_0\}$$

for some $x_0 > 0$.

- b) Determine x_0 and find g^* .

- c) Verify the *high contact principle*:

$$\frac{\partial g^*}{\partial x} = \frac{\partial g}{\partial x} \quad \text{when } (s, x) = (s, x_0),$$

where $g(t, x) = e^{-\rho t} x^+$.

10.12.

The first time the high contact principle was formulated seems to be in a paper by Samuelson (1965), who studied the optimal time for selling an asset, if the reward obtained by selling at the time t and when price is ξ is given by

$$g(t, \xi) = e^{-\rho t}(\xi - 1)^+.$$

The price process is assumed to be a geometric Brownian motion X_t given by

$$dX_t = rX_t dt + \alpha X_t dB_t, \quad X_0 = x > 0,$$

where $r < \rho$.

In other words, the problem is to find g^*, τ^* such that

$$g^*(s, x) = \sup_{\tau} E^{(s, x)}[e^{-\rho\tau}(X_{\tau} - 1)^+] = E^{(s, x)}[e^{-\rho\tau^*}(X_{\tau^*} - 1)^+].$$

- a) Use the argument for (10.46) and Exercise 10.10 to prove that the continuation region D has the form

$$D = \{(s, x); 0 < x < x_0\}$$

for some $x_0 > \frac{\rho}{\rho-r}$.

- b) For a given $x_0 > \frac{\rho}{\rho-r}$ solve the boundary value problem

$$\begin{cases} \frac{\partial f}{\partial s} + rx \frac{\partial f}{\partial x} + \frac{1}{2}\alpha^2 x^2 \frac{\partial^2 f}{\partial x^2} = 0 & \text{for } 0 < x < x_0 \\ f(s, 0) = 0 \\ f(s, x_0) = e^{-\rho t}(x_0 - 1)^+ \end{cases}$$

by trying $f(s, x) = e^{-\rho s} \phi(x)$.

- c) Determine x_0 by using the high contact principle, i.e. by using that

$$\frac{\partial f}{\partial x} = \frac{\partial g}{\partial x} \quad \text{when } x = x_0.$$

- d) With f, x_0 as in b), c) define

$$\gamma(s, x) = \begin{cases} f(s, x); & x < x_0 \\ e^{-\rho s}(x - 1)^+; & x \geq x_0. \end{cases}$$

Use Theorem 10.18 to verify that $\gamma = g^*$ and that $\tau^* = \tau_D$ is optimal.

10.13. (A resource extraction problem)

Suppose the price P_t of one unit of a resource (e.g. gas, oil) at time t is varying like a geometric Brownian motion

$$dP_t = \alpha P_t dt + \beta P_t dB_t; \quad P_0 = p$$

where B_t is 1-dimensional Brownian motion and α, β are constants.

Let Q_t denote the amount of remaining resources at time t . Assume that the rate of extraction is proportional to the remaining amount, so that

$$dQ_t = -\lambda Q_t dt; \quad Q_0 = q$$

where $\lambda > 0$ is a constant.

If the running cost rate is $K > 0$ and we stop the extraction at the time $\tau = \tau(\omega)$ then the expected total discounted profit is given by

$$J^\tau(s, p, q) = E^{(s, p, q)} \left[\int_0^\tau (\lambda P_t Q_t - K) e^{-\rho(s+t)} dt + e^{-\rho(s+\tau)} g(P_\tau, Q_\tau) \right],$$

where $\rho > 0$ is the discounting exponent and $g(p, q)$ is a given bequest function giving the value of the remaining resource amount q when the price is p .

- a) Write down the characteristic operator \mathcal{A} of the diffusion process

$$dX_t = \begin{bmatrix} dt \\ dP_t \\ dQ_t \end{bmatrix}; \quad X_0 = (s, p, q)$$

and formulate the variational inequalities of Theorem 10.18 corresponding to the following optimal stopping problem

$$G^*(s, p, q) = \sup_{\tau} J^\tau(s, p, q) = J^{\tau^*}(s, p, q).$$

- b) Assume that $g(p, q) = pq$ and find the domain U corresponding to (10.35), (10.58), i.e.

$$U = \{(s, p, q); \mathcal{A}(e^{-\rho s}g(p, q) + f(s, p, q)) > 0\},$$

where

$$f(s, p, q) = e^{-\rho s}(\lambda pq - K).$$

Conclude that

- (i) if $\rho \geq \alpha$ then $\tau^* = 0$ and $G^*(p, q) = pq$
- (ii) if $p < \alpha$ then $D \supset \{(s, p, q); pq > \frac{K}{\alpha-\rho}\}$.

- c) As a candidate for G^* when $\rho < \alpha$ we try a function of the form

$$\phi(s, p, q) = \begin{cases} e^{-\rho s}pq; & 0 < pq \leq y_0 \\ e^{-\rho s}\psi(pq); & pq > y_0 \end{cases}$$

for a suitable $\psi : \mathbf{R} \rightarrow \mathbf{R}$, and a suitable y_0 . Use Theorem 10.18 to determine ψ, y_0 and to verify that with this choice of ψ, y_0 we have $\phi = G^*$ and $\tau^* = \inf\{t > 0; P_t Q_t \leq y_0\}$, if $\rho < \alpha < \rho + \lambda$.

- d) What happens if $\rho + \lambda \leq \alpha$?

10.14. (Finding the optimal investment time)

Solve the optimal stopping problem

$$G^*(s, p) = \sup_{\tau} E^{(s, p)} \left[\int_{\tau}^{\infty} e^{-\rho(s+t)} P_t dt - C e^{-\rho(s+\tau)} \right],$$

where

$$dP_t = \alpha P_t dt + \beta P_t dB_t; \quad P_0 = p,$$

B_t is 1-dimensional Brownian motion and α, β, ρ, C are constants, $0 < \alpha < \rho$ and $C > 0$. (We may interpret this as the problem of finding the optimal time for investment in a project. The profit rate after investment is P_t and the cost of the investment is C . Thus G^* gives the maximal expected discounted net profit.)

Hint: Write $\int_{\tau}^{\infty} e^{-\rho(s+t)} P_t dt = e^{-\rho s} \left[\int_0^{\infty} e^{-\rho t} P_t dt - \int_0^{\tau} e^{-\rho t} P_t dt \right]$. Compute $E \left[\int_0^{\infty} e^{-\rho t} P_t dt \right]$ by using the solution formula for P_t (see Chapter V) and then apply Theorem 10.18 to the problem

$$\Phi(s, p) = \sup_{\tau} E^{(s, p)} \left[- \int_0^{\tau} e^{-\rho(s+t)} P_t dt - C e^{-\rho(s+\tau)} \right].$$

XI. Application to Stochastic Control

Statement of the problem

Suppose that the state of a system at time t is described by an Ito process X_t of the form

$$dX_t = dX_t^u = b(t, X_t, u_t)dt + \sigma(t, X_t, u_t)dB_t, \quad (11.1)$$

where $X_t \in \mathbf{R}^n$, $b : \mathbf{R} \times \mathbf{R}^n \times U \rightarrow \mathbf{R}^n$, $\sigma : \mathbf{R} \times \mathbf{R}^n \times U \rightarrow \mathbf{R}^{n \times m}$ and B_t is m -dimensional Brownian motion. Here $u_t \in U \subset \mathbf{R}^k$ is a parameter whose value we can choose in the given Borel set U at any instant t in order to control the process X_t . Thus $u_t = u(t, \omega)$ is a stochastic process. Since our decision at time t must be based upon what has happened up to time t , the function $\omega \rightarrow u(t, \omega)$ must (at least) be measurable w.r.t. \mathcal{F}_t , i.e. the process u_t must be \mathcal{F}_t -adapted. Thus the right hand side of (11.1) is well-defined as a stochastic integral, under suitable assumptions on the functions b and σ . At the moment we will not specify the conditions on b and σ further, but simply assume that the process X_t satisfying (11.1) exists. See further comments on this in the end of this chapter.

Let $\{X_h^{s,x}\}_{h \geq s}$ be the solution of (11.1) such that $X_s^{s,x} = x$, i.e.

$$X_h^{s,x} = x + \int_s^h b(r, X_r^{s,x}, u_r)dr + \int_s^h \sigma(r, X_r^{s,x}, u_r)dB_r; \quad h \geq s$$

and let the probability law of X_t starting at x for $t = s$ be denoted by $Q^{s,x}$, so that

$$Q^{s,x}[X_{t_1} \in F_1, \dots, X_{t_k} \in F_k] = P^0[X_{t_1}^{s,x} \in F_1, \dots, X_{t_k}^{s,x} \in F_k] \quad (11.2)$$

for $s \leq t_i$, $F_i \subset \mathbf{R}^n$; $1 \leq i \leq k$, $k = 1, 2, \dots$

To obtain an easier notation we introduce

$$Y_t = (s + t, X_{s+t}^{s,x}) \quad \text{for } t \geq 0, \quad Y_0 = (t, x)$$

and we observe that if we substitute this in (11.1) we get the equation

$$dY_t = dY_t^u = b(Y_t, u_t)dt + \sigma(Y_t, u_t)dB_t, \quad (11.3)$$

i.e. the stochastic integral is time-homogeneous. (Strictly speaking, the u, b and σ in (11.3) are slightly different from the u, b and σ in (11.1).) The probability law of Y_t starting at $y = (s, x)$ for $t = 0$ is (with slight abuse of notation) also denoted by $Q^{s,x} = Q^y$.

Let $F : \mathbf{R} \times \mathbf{R}^n \times U \rightarrow \mathbf{R}$ (the “utility rate” function) and $K : \mathbf{R} \times \mathbf{R}^n \rightarrow \mathbf{R}$ (the “bequest” function) be given continuous functions, let G be a fixed domain in $\mathbf{R} \times \mathbf{R}^n$ and let \hat{T} be the first exit time after s from G for the process $\{X_r^{s,x}\}_{r \geq s}$, i.e.

$$\hat{T} = \hat{T}^{s,x}(\omega) = \inf\{r > s; (r, X_r^{s,x}(\omega)) \notin G\}. \quad (11.4)$$

Suppose

$$E^{s,x} \left[\int_s^{\hat{T}} |F^{u_r}(r, X_r)| dr + |K(\hat{T}, X_{\hat{T}})| \chi_{\{\hat{T} < \infty\}} \right] < \infty \quad \text{for all } s, x, u \quad (11.5)$$

where $F^u(r, z) = F(r, z, u)$. Then we define the *performance function* $J^u(s, x)$ by

$$J^u(s, x) = E^{s,x} \left[\int_s^{\hat{T}} F^{u_r}(r, X_r) dr + K(\hat{T}, X_{\hat{T}}) \chi_{\{\hat{T} < \infty\}} \right]. \quad (11.6)$$

Note that

$$\int_s^{\hat{T}} F^{u_r}(r, X_r) dr = \int_0^{\hat{T}-s} F^{u_{s+t}}(s+t, X_{s+t}) dt = \int_0^T F^{u_{s+t}}(Y_t) dt,$$

where

$$T := \inf\{t > 0; Y_t \notin G\} = \hat{T} - s. \quad (11.7)$$

Moreover,

$$K(\hat{T}, X_{\hat{T}}) = K(Y_{\hat{T}-s}) = K(Y_T).$$

Therefore the performance function may be written in terms of Y as follows, with $y = (s, x)$,

$$J^u(y) = E^y \left[\int_0^T F^{u_t}(Y_t) dt + K(Y_T) \chi_{\{T < \infty\}} \right]. \quad (11.8)$$

(Strictly speaking this u_t is a time shift of the u_t in (11.6).)

The problem is – for each $y \in G$ – to find the number $\Phi(y)$ and a control $u^* = u^*(t, \omega) = u^*(y, t, \omega)$ such that

$$\Phi(y) := \sup_{u(t,\omega)} J^u(y) = J^{u^*}(y) \quad (11.9)$$

where the supremum is taken over all \mathcal{F}_t -adapted processes $\{u_t\}$ with values in U . Such a control u^* – if it exists – is called an *optimal control* and Φ is called the *optimal performance or the value function*. Examples of types of control functions that may be considered are:

- 1) Functions of the form $u(t,\omega) = u(t)$ i.e. not depending on ω . These controls are sometimes called *deterministic* or *open loop controls*.
- 2) Processes $\{u_t\}$ which are \mathcal{M}_t -adapted, i.e. for each t the function $\omega \rightarrow u(t,\omega)$ is \mathcal{M}_t -measurable, where \mathcal{M}_t is the σ -algebra generated by $\{X_s^u; s \leq t\}$. These controls are called *closed loop* or *feedback controls*.
- 3) The controller has only *partial knowledge* of the state of the system. More precisely, to the controller's disposal are only (noisy) observations R_t of X_t , given by an Ito process of the form

$$dR_t = a(t, X_t)dt + \gamma(t, X_t)d\hat{B}_t,$$

where \hat{B} is a Brownian motion (not necessarily related to B). Hence the control process $\{u_t\}$ must be adapted w.r.t. the σ -algebra \mathcal{N}_t generated by $\{R_s; s \leq t\}$. In this situation the stochastic control problem is linked to the filtering problem (Chapter VI). In fact, if the equation (11.1) is linear and the performance function is integral quadratic (i.e. F and K are quadratic) then the stochastic control problem splits into a linear filtering problem and a corresponding deterministic control problem. This is called the Separation Principle. See Example 11.4.

- 4) Functions $u(t,\omega)$ of the form $u(t,\omega) = u_0(t, X_t(\omega))$ for some function $u_0 : \mathbf{R}^{n+1} \rightarrow U \subset \mathbf{R}^k$. In this case we assume that u does not depend on the starting point $y = (s,x)$: The value we choose at time t only depends on the state of the system at this time. These are called *Markov controls*, because with such u the corresponding process X_t becomes an Ito diffusion, in particular a Markov process. In the following we will not distinguish between u and u_0 . Thus we will identify a function $u : \mathbf{R}^{n+1} \rightarrow U$ with the Markov control $u(Y) = u(t, X_t)$ and simply call such functions Markov controls.

The Hamilton-Jacobi-Bellman equation

Let us first consider only *Markov controls*

$$u = u(t, X_t(\omega))$$

Introducing $Y_t = (s+t, X_{s+t})$ (as explained earlier) the system equation becomes

$$dY_t = b(Y_t, u(Y_t))dt + \sigma(Y_t, u(Y_t))dB_t. \quad (11.10)$$

For $v \in U$ and $f \in C_0^2(\mathbf{R} \times \mathbf{R}^n)$ define

$$(L^v f)(y) = \frac{\partial f}{\partial s}(y) + \sum_{i=1}^n b_i(y, v) \frac{\partial f}{\partial x_i} + \sum_{i,j=1}^n a_{ij}(y, v) \frac{\partial^2 f}{\partial x_i \partial x_j} \quad (11.11)$$

where $a_{ij} = \frac{1}{2}(\sigma \sigma^T)_{ij}$, $y = (s, x)$ and $x = (x_1, \dots, x_n)$. Then for each choice of the function u the solution $Y_t = Y_t^u$ is an Ito diffusion with generator A given by

$$(Af)(y) = (L^{u(y)} f)(y) \quad \text{for } f \in C_0^2(\mathbf{R} \times \mathbf{R}^n) \text{ (see Theorem 7.9).}$$

For $v \in U$ define $F^v(y) = F(y, v)$. The first fundamental result in stochastic control theory is the following:

THEOREM 11.1. (The Hamilton-Jacobi-Bellman (HJB) equation (I))

Define

$$\Phi(y) = \sup\{J^u(y); u = u(Y) \text{ Markov control}\}.$$

Suppose that $\Phi \in C^2(G)$, $T < \infty$ a.s. Q^y for all $y \in G$ and that an optimal Markov control u^* exists. Then

$$\sup_{v \in U} \{F^v(y) + (L^v \Phi)(y)\} = 0 \quad \text{for all } y \in G \quad (11.12)$$

and

$$\Phi(y) = K(y) \quad \text{for all } y \in \partial_R G, \quad (11.13)$$

where $\partial_R G$ is the set of regular boundary points of G for $Y_t^{u^*}$. The supremum in (11.12) is obtained if $v = u^*(y)$ where $u^*(y)$ is optimal. In other words,

$$F(y, u^*(y)) + (L^{u^*} \Phi)(y) = 0 \quad \text{for all } y \in G. \quad (11.14)$$

Proof. The last two statements are easy to prove: Since $u^* = u^*(y)$ is optimal we have

$$\Phi(y) = J^{u^*}(y) = E^y \left[\int_0^T F(Y_s, u^*(Y_s)) ds + K(Y_T) \right].$$

If $y \in \partial_R G$ then $T = 0$ a.s. Q^y and (11.13) follows. By the solution of the Dirichlet-Poisson problem (Theorem 9.18)

$$(L^{u^*(y)} \Phi)(y) = -F(y, u^*(y)) \quad \text{for all } y \in G,$$

which is (11.14). We proceed to prove (11.12). Fix $y = (s, x) \in G$ and choose a Markov control u . Let $\alpha \leq T$ be a stopping time.

Since

$$J^u(y) = E^y \left[\int_0^\alpha F^u(Y_r) dr + K(Y_\alpha) \right],$$

we get by the strong Markov property, (7.16) and (9.27)

$$\begin{aligned}
 E^y[J^u(Y_\alpha)] &= E^y[E^{Y_\alpha}[\int_0^T F^u(Y_r)dr + K(Y_T)]] \\
 &= E^y[E^y[\theta_\alpha(\int_0^T F^u(Y_r)dr + K(Y_T))|\mathcal{F}_\alpha]] \\
 &= E^y[E^y[\int_\alpha^T F^u(Y_r)dr + K(Y_T)|\mathcal{F}_\alpha]] \\
 &= E^y[\int_0^T F^u(Y_r)dr + K(Y_T) - \int_0^\alpha F^u(Y_r)dr] \\
 &= J^u(y) - E^y[\int_0^\alpha F^u(Y_r)dr].
 \end{aligned}$$

So

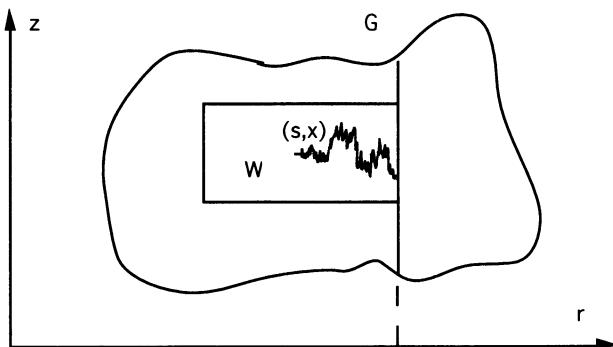
$$J^u(y) = E^y[\int_0^\alpha F^u(Y_r)dr] + E^y[J^u(Y_\alpha)]. \quad (11.15)$$

Now let $W \subset G$ be of the form $W = \{(r, z) \in G; t_0 < r < t_1 \text{ and } |z - x| < \varepsilon\}$ where $t_0 < s < t_1$ and $\varepsilon > 0$ are chosen such that $\overline{W} \subset G$. Put $\alpha = \tau_W$, the first exit time from W of Y_t . Suppose an optimal control $u^*(y) = u^*(r, z)$ exists and choose

$$u(r, z) = \begin{cases} v & \text{if } (r, z) \in W \\ u^*(r, z) & \text{if } (r, z) \in G \setminus W \end{cases}$$

where $v \in U$ is arbitrary. Then

$$\Phi(Y_\alpha) = J^{u^*}(Y_\alpha) = J^u(Y_\alpha) \quad (11.16)$$



and therefore, combining (11.15) and (11.16) we obtain

$$\Phi(y) \geq J^u(y) = E^y \left[\int_0^\alpha F^v(Y_r) dr \right] + E^y[\Phi(Y_\alpha)]. \quad (11.17)$$

Since $\Phi \in C^2(G)$ we get by Dynkin's formula

$$E^y[\Phi(Y_\alpha)] = \Phi(y) + E^y \left[\int_0^\alpha L^u \Phi(Y_r) dr \right],$$

which substituted in (11.17) gives

$$\Phi(y) \geq E^y \left[\int_0^\alpha F^v(Y_r) dr \right] + \Phi(y) + E^y \left[\int_0^\alpha (L^v \Phi)(Y_r) dr \right]$$

or

$$E^y \left[\int_0^\alpha (F^v(Y_r) + (L^v \Phi)(Y_r)) dr \right] \leq 0$$

So

$$\frac{E^y \left[\int_0^\alpha (F^v(Y_r) + (L^v \Phi)(Y_r)) dr \right]}{E^y[\alpha]} \leq 0 \quad \text{for all such } W.$$

Letting $t_1 \downarrow s$ we obtain, since $F^v(\cdot)$ and $(L^v \Phi)(\cdot)$ are continuous at y , that $F^v(y) + (L^v \Phi)(y) \leq 0$, which combined with (11.14) gives (11.12). That completes the proof.

REMARK. The HJB (I) equation states that if an optimal control u^* exists, then we know that its value v at the point y is a point v where the function

$$v \rightarrow F^v(y) + (L^v \Phi)(y); \quad v \in U$$

attains its maximum (and this maximum is 0). Thus the original stochastic control problem is associated to the easier problem of finding the maximum of a real function in $U \subset \mathbf{R}^k$. However, the HJB (I) equation only states that it is necessary that $v = u^*(y)$ is the maximum of this function. It is just as important to know if this is also sufficient: If at each point y we have found $v = v_0(y)$ such that $F^v(y) + (L^v \Phi)(y)$ is maximal and this maximum is 0, will $u_0(Y)$ be an optimal control? The next result states that (under some conditions) this is actually the case:

THEOREM 11.2. (The HJB (II) equation – a converse of HJB (I))

Let ϕ be a function in $C^2(G) \cap C(\overline{G})$ such that, for all $v \in U$,

$$F^v(y) + (L^v \phi)(y) \leq 0; \quad y \in G \quad (11.18)$$

with boundary values

$$\lim_{t \rightarrow T} \phi(Y_t) = K(Y_T) \cdot \mathcal{X}_{\{T < \infty\}} \quad \text{a.s. } Q^y \quad (11.19)$$

and such that

$$\{\phi(Y_\tau)\}_{\tau \leq T} \text{ is uniformly } Q^y\text{-integrable for all Markov controls } u \text{ and all } y \in G. \quad (11.20)$$

Then

$$\phi(y) \geq J^u(y) \quad \text{for all Markov controls } u \text{ and all } y \in G. \quad (11.21)$$

Moreover, if for each $y \in G$ we have found $u_0(y)$ such that

$$F^{u_0(y)}(y) + (L^{u_0(y)}\phi)(y) = 0 \quad (11.22)$$

then $u_0 = u_0(y)$ is a Markov control such that

$$\phi(y) = J^{u_0(y)}(y)$$

and hence u_0 must be an *optimal* control and $\phi(y) = \bar{\Phi}(y)$.

Proof. Assume that ϕ satisfies (11.18) and (11.19) above. Let u be a Markov control. Since $L^u\phi \leq -F^u$ in G we have by Dynkin's formula

$$\begin{aligned} E^y[\phi(Y_{T_R})] &= \phi(y) + E^y\left[\int_0^{T_R} (L^u\phi)(Y_r)dr\right] \\ &\leq \phi(y) - E^y\left[\int_0^{T_R} F^u(Y_r)dr\right] \end{aligned}$$

where

$$T_r = \min\{R, T, \inf\{t > 0; |Y_t| \geq R\}\} \quad (11.23)$$

for all $R < \infty$. This gives, by (11.5),

$$\phi(y) \geq E^y\left[\int_0^{T_R} F^u(Y_r)dr + \phi(Y_{T_R})\right] \rightarrow E^y\left[\int_0^T F^u(Y_r)dr + K(Y_T)\mathcal{X}_{\{T < \infty\}}\right] = J^u(y)$$

as $R \rightarrow \infty$, which proves (11.21). If u_0 is such that (11.22) holds, then the calculations above give equality and the proof is complete.

The HJB equations (I), (II) provide a very nice solution to the stochastic control problem in the case where only Markov controls are considered. One might feel that considering only Markov controls is too restrictive, but fortunately one can

always obtain as good performance with a Markov control as with an arbitrary \mathcal{F}_t -adapted control, at least if some extra conditions are satisfied:

THEOREM 11.3. Let

$$\Phi_M(y) = \sup\{J^u(y); u = u(Y) \text{ Markov control}\}$$

and

$$\Phi_a(y) = \sup\{J^u(y); u = u(t, \omega) \text{ } \mathcal{F}_t\text{-adapted control}\}.$$

Suppose there exists an optimal Markov control $u_0 = u_0(Y)$ for the Markov control problem (i.e. $\Phi_M(y) = J^{u_0}(y)$ for all $y \in G$) such that all the boundary points of G are regular w.r.t. $Y_t^{u_0}$ and that Φ_M is a bounded function in $C^2(G) \cap C(\overline{G})$. Then

$$\Phi_M(y) = \Phi_a(y) \quad \text{for all } y \in G.$$

Proof. Let ϕ be a bounded function in $C^2(G) \cap C(\overline{G})$ satisfying

$$F^v(y) + (L^v\phi)(y) \leq 0 \quad \text{for all } y \in G, v \in U \quad (11.24)$$

and

$$\phi(y) = K(y) \quad \text{for all } y \in \partial G. \quad (11.25)$$

Let $u_t(\omega) = u(t, \omega)$ be an \mathcal{F}_t -adapted control. Then Y_t is an Ito process given by

$$dY_t = b(Y_t, u_t)dt + \sigma(Y_t, u_t)dB_t$$

so by Lemma 7.8, with T_R as in (11.23),

$$E^y[\phi(Y_{T_R})] = \phi(y) + E^y\left[\int_0^{T_R} (L^{u(t, \omega)}\phi)(Y_t)dt\right], \quad (11.26)$$

where

$$(L^{u(t, \omega)}\phi)(y) = \frac{\partial \phi}{\partial t}(y) + \sum_{i=1}^n b_i(y, u(t, \omega)) \frac{\partial \phi}{\partial x_i}(y) + \sum_{i,j=1}^n a_{ij}(y, u(t, \omega)) \frac{\partial^2 \phi}{\partial x_i \partial x_j}(y),$$

with $a_{ij} = \frac{1}{2}(\sigma\sigma^T)_{ij}$. Thus by (11.24) and (11.25) this gives

$$E^y[K(Y_{T_R})] \leq \phi(y) - E^y\left[\int_0^{T_R} F(Y_t, u(t, \omega))dt\right]. \quad (11.27)$$

Letting $R \rightarrow \infty$ we obtain

$$\phi(y) \geq J^u(y). \quad (11.28)$$

But by Theorem 11.1 the function $\phi(y) = \Phi_M(y)$ satisfies (11.24) and (11.25). So by (11.28) we have $\Phi_M(y) \geq \Phi_a(y)$ and Theorem 11.3 follows.

REMARK. The theory above also applies to the corresponding *minimum* problem

$$\Psi(y) = \inf_u J^u(y) = J^{u^*}(y). \quad (11.29)$$

To see the connection we note that

$$\Psi(y) = -\sup_u \{-J^u(y)\} = -\sup_u \{E^y \left[\int_0^T -F^u(Y_t) dt - K(Y_t) \right] \}$$

so $-\Psi$ coincides with the solution Φ of the problem (11.9), but with F replaced by $-F$ and K replaced by $-K$. Using this, we see that the HJB equations apply to Ψ also but with reverse inequalities. For example, equation (11.12) for Φ gets for Ψ the form

$$\inf_{v \in U} \{F^v(y) + (L^v \Psi)(y)\} = 0 \quad \text{for all } y \in G. \quad (11.30)$$

We now illustrate the results by some examples:

EXAMPLE 11.4. (The linear stochastic regulator problem)

Suppose that the state X_t of the system at time t is given by a linear stochastic differential equation:

$$dX_t = (H_t X_t + M_t u_t) dt + \sigma_t dB_t, \quad t \geq s; \quad X_s = x \quad (11.31)$$

and the cost is of the form

$$J^u(s, x) = E^{s, x} \left[\int_s^{t_1} \{X_t^T C_t X_t + u_t^T D_t u_t\} dt + X_{t_1}^T R X_{t_1} \right], \quad s \leq t_1 \quad (11.32)$$

where all the coefficients $H_t \in \mathbf{R}^{n \times n}$, $M_t \in \mathbf{R}^{n \times k}$, $\sigma_t \in \mathbf{R}^{n \times m}$, $C_t \in \mathbf{R}^{n \times n}$, $D_t \in \mathbf{R}^{k \times k}$ and $R \in \mathbf{R}^{n \times n}$ are t -continuous and deterministic. We assume that C_t and R are symmetric, nonnegative definite and D_t is symmetric, positive definite, for all t . We also assume that t_1 is a deterministic time.

The problem is then to choose the control $u = u(t, X_t)$ such that it minimizes $J^u(s, x)$. We may interpret this as follows: The aim is to find a control u which makes $|X_t|$ small fast and such that the energy used ($\sim u^T D u$) is small. The sizes of C_t and R reflect the cost of having large values of $|X_t|$, while the size of D_t reflects the cost (energy) of applying large values of $|u_t|$.

In this case the HJB-equation for $\Psi(s, x) = \inf_u J^u(s, x)$ becomes

$$\begin{aligned}
0 &= \inf_v \{F^v(s, x) + (L^v \Psi)(s, x)\} \\
&= \frac{\partial \Psi}{\partial s} + \inf_v \{x^T C_t x + v^T D_t v + \sum_{i=1}^n (H_t x + M_t v)_i \frac{\partial \Psi}{\partial x_i} \\
&\quad + \frac{1}{2} \sum_{i,j=1}^n (\sigma \sigma^T)_{ij} \frac{\partial^2 \Psi}{\partial x_i \partial x_j}\} \quad \text{for } s < t_1
\end{aligned} \tag{11.33}$$

and

$$\Psi(t_1, x) = x^T R x. \tag{11.34}$$

Let us try to find a solution ψ of (11.33)–(11.34) of the form

$$\psi(t, x) = x^T S_t x + a_t \tag{11.35}$$

where $S(t) = S_t \in \mathbf{R}^{n \times n}$ is symmetric, nonnegative definite, $a_t \in \mathbf{R}$ and both a_t and S_t are continuously differentiable w.r.t. t (and deterministic). In order to use Theorem 11.2 we need to determine S_t and a_t such that

$$\inf_v \{F^v(t, x) + (L^v \psi)(t, x)\} = 0 \quad \text{for } t < t_1 \tag{11.36}$$

and

$$\psi(t_1, x) = x^T R x \tag{11.37}$$

To obtain (11.37) we put

$$S_{t_1} = R \tag{11.38}$$

$$a_{t_1} = 0 \tag{11.39}$$

Using (11.35) we get

$$\begin{aligned}
F^v(t, x) + (L^v \psi)(t, x) &= x^T S'_t x + a'_t + x^T C_t + v^T D_t v + \\
&\quad + (H_t x + M_t v)^T (S_t x + S_t^T x) + \sum_{i,j} (\sigma \sigma^T)_{ij} S_{ij},
\end{aligned} \tag{11.40}$$

where $S'_t = \frac{d}{dt} S_t$, $a'_t = \frac{d}{dt} a_t$. The minimum of this expression is obtained when

$$\frac{\partial}{\partial v_i} (F^v(t, x) + (L^v \psi)(t, x)) = 0; \quad i = 1, \dots, k$$

i.e. when

$$2D_t v + 2M_t^T S_t x = 0$$

i.e. when

$$v = -D_t^{-1} M_t^T S_t x. \tag{11.41}$$

We substitute this value of v in (11.40) and obtain

$$\begin{aligned}
F^v(t, x) + (L^v \psi)(t, x) &= \\
&= x^T S'_t x + a'_t + x^T C_t x + x^T S_t M_t D_t^{-1} D_t D_t^{-1} M_t^T S_t x \\
&+ (H_t x - M_t D_t^{-1} M_t^T S_t x)^T 2S_t x + \text{tr}(\sigma \sigma^T S)_t \\
&= x^T (S'_t + C_t - S_t M_t D_t^{-1} M_t^T S_t + 2H_t^T S_t) x + a'_t + \text{tr}(\sigma \sigma^T S)_t,
\end{aligned}$$

where tr denotes the (matrix) trace. We obtain that this is 0 if we choose S_t such that

$$S'_t = -2H_t^T S_t + S_t M_t D_t^{-1} M_t^T S_t - C_t; t < t_1 \quad (11.42)$$

and a_t such that

$$a'_t = -\text{tr}(\sigma \sigma^T S)_t; \quad t < t_1. \quad (11.43)$$

We recognize (11.42) as a Riccati type equation from linear filtering theory (see (6.30)). Equation (11.42) with boundary condition (11.38) determines S_t uniquely. Combining (11.43) with the boundary condition (11.39) we obtain

$$a_t = \int_t^{t_1} \text{tr}(\sigma \sigma^T S)_s ds. \quad (11.44)$$

With such a choice of S_t and a_t we see that (11.36) and (11.37) hold, so by Theorem 11.2 we conclude that

$$u^*(t, x) = -D_t^{-1} M_t^T S_t x, \quad t < t_1 \quad (11.45)$$

is an optimal control and the minimum cost is

$$\Psi(s, x) = x^T S_s x + \int_s^{t_1} \text{tr}(\sigma \sigma^T S)_t dt, \quad s < t_1. \quad (11.46)$$

This formula shows that the extra cost due to the noise in the system is given by

$$a_s = \int_s^{t_1} \text{tr}(\sigma \sigma^T S)_t dt.$$

The Separation Principle (see Davis (1977), Davis and Vinter (1985) or Fleming and Rishel (1975)) states that if we had only partial knowledge of the state X_t of the system, i.e. if we only had noisy observations

$$dZ_t = g_t X_t dt + \gamma_t d\tilde{B}_t \quad (11.47)$$

to our disposal, then the optimal control $u^*(t, \omega)$ (required to be \mathcal{G}_t -adapted, where \mathcal{G}_t is the σ -algebra generated by $\{Z_r; r \leq t\}$), would be given by

$$u^*(t, \omega) = -D_t^{-1} M_t^T S_t \hat{X}_t(\omega), \quad (11.48)$$

where \hat{X}_t is the filtered estimate of X_t based on the observations $\{Z_r; r \leq t\}$, given by the Kalman-Bucy filter (6.43). Comparing with (11.45) we see that the stochastic control problem in this case splits into a linear filtering problem and a deterministic control problem.

An important field of applications of the stochastic control theory is economics and finance. Therefore we illustrate the results above by applying them to a simple case of optimal portfolio diversification. This problem has been considered in more general settings by many authors, see for example Markowitz (1976), Merton (1971), Harrison and Pliska (1981), Aase (1984), Karatzas, Lehoczky and Shreve (1987) and the survey article Duffie (1994) and the references therein.

EXAMPLE 11.5. (An optimal portfolio selection problem)

Let X_t denote the wealth of a person at time t . Suppose that the person has the choice of two different investments. The price $p_1(t)$ at time t of one of the assets is assumed to satisfy the equation

$$\frac{dp_1}{dt} = p_1(a + \alpha W_t) \quad (11.49)$$

where W_t denotes white noise and $a, \alpha > 0$ are constants measuring the average relative rate of change of p and the size of the noise, respectively. As we have discussed earlier we interpret (11.49) as the (Ito) stochastic differential equation

$$dp_1 = p_1 adt + p_1 \alpha dB_t. \quad (11.50)$$

This investment is called *risky*, since $\alpha > 0$. We assume that the price p_2 of the other asset satisfies a similar equation, but with no noise:

$$dp_2 = p_2 bdt. \quad (11.51)$$

This investment is called *safe*. So it is natural to assume $b < a$. At each instant the person can choose how big fraction u of his wealth he will invest in the risky asset, thereby investing the fraction $1 - u$ in the safe one. This gives the following stochastic differential equation for the wealth $X_t = X_t^u$:

$$\begin{aligned} dX_t &= uX_t adt + uX_t \alpha dB_t + (1 - u)X_t bdt \\ &= X_t(au + b(1 - u))dt + \alpha uX_t dB_t. \end{aligned} \quad (11.52)$$

Suppose that, starting with the wealth $X_t = x > 0$ at time t , the person wants to maximize the expected utility of the wealth at some future time $t_0 > t$. If we allow no borrowing (i.e. require $X \geq 0$) and are given a utility function $N : [0, \infty) \rightarrow [0, \infty)$, $N(0) = 0$ (usually assumed to be increasing and concave) the problem is to find $\Phi(s, x)$ and a (Markov) control $u^* = u^*(t, X_t)$, $0 \leq u^* \leq 1$, such that

$$\Phi(s, x) = \sup\{J^u(s, x); u \text{ Markov control, } 0 \leq u \leq 1\} = J^{u^*}(s, x), \\ \text{where } J^u(s, x) = E^{s, x}[N(X_T^u)] \quad (11.53)$$

and T is the first exit time from the region $G = \{(r, z); r < t_0, z > 0\}$. This is a performance criterion of the form (11.6)/(11.8) with $F = 0$ and $K = N$. The differential operator L^v has the form (see (11.11))

$$(L^v f)(t, x) = \frac{\partial f}{\partial t} + x(av + b(1 - v)) \frac{\partial f}{\partial x} + \frac{1}{2} \alpha^2 v^2 x^2 \frac{\partial^2 f}{\partial x^2}. \quad (11.54)$$

The HJB equation becomes

$$\sup_v \{(L^v \Phi)(t, x)\} = 0, \quad \text{for } (t, x) \in G; \quad (11.55)$$

and

$$\Phi(t, x) = N(x) \quad \text{for } t = t_0, \quad \Phi(t, 0) = N(0) \quad \text{for } t < t_0. \quad (11.56)$$

Therefore, for each (t, x) we try to find the value $v = u(t, x)$ which maximizes the function

$$\eta(v) = L^v \Phi = \frac{\partial \Phi}{\partial t} + x(b + (a - b)v) \frac{\partial \Phi}{\partial x} + \frac{1}{2} \alpha^2 v^2 x^2 \frac{\partial^2 \Phi}{\partial x^2}. \quad (11.57)$$

If $\Phi_x := \frac{\partial \Phi}{\partial x} > 0$ and $\Phi_{xx} := \frac{\partial^2 \Phi}{\partial x^2} < 0$, the solution is

$$v = u(t, x) = -\frac{(a - b)\Phi_x}{x\alpha^2\Phi_{xx}}. \quad (11.58)$$

If we substitute this into the HJB equation (11.57) we get the following non-linear boundary value problem for Φ :

$$\Phi_t + bx\Phi_x - \frac{(a - b)^2\Phi_x^2}{2\alpha^2\Phi_{xx}} = 0 \quad \text{for } t < t_0, x > 0 \quad (11.59)$$

$$\Phi(t, x) = N(x) \quad \text{for } t = t_0 \text{ or } x = 0. \quad (11.60)$$

The problem (11.59–60) is hard to solve for general N . Important examples of increasing and concave functions are the power functions

$$N(x) = x^r \quad \text{where } 0 < r < 1. \quad (11.61)$$

If we choose such a utility function N , we try to find a solution of (11.59–60) of the form

$$\Phi(t, x) = f(t)x^r.$$

Substituting we obtain

$$\Phi(t, x) = e^{\lambda(t_0 - t)}x^r, \quad (11.62)$$

where $\lambda = br + \frac{(a-b)^2 r}{2\alpha^2(1-r)}$.

Using (11.58) we obtain the optimal control

$$u^*(t, x) = \frac{a - b}{\alpha^2(1 - r)}. \quad (11.63)$$

If $\frac{a - b}{\alpha^2(1 - r)} \in (0, 1)$ this is the solution to the problem, in virtue of Theorem 11.2. Note that u^* is in fact constant.

Another interesting choice of the utility function is $N(x) = \log x$, called the Kelly criterion. As noted by Aase (1984) (in a more general setting) we may in this case obtain the optimal control directly by evaluating $E^{s,x}[\log(X_T)]$ using Dynkin's formula:

$$\begin{aligned} E^{s,x}[\log(X_T)] &= \\ &= \log x + E^{s,x} \left[\int_s^T \{au(t, X_t) + b(1 - u(t, X_t)) - \frac{1}{2}\alpha^2 u^2(t, X_t)\} dt \right] \end{aligned}$$

since $L^v(\log x) = av + b(1 - v) - \frac{1}{2}\alpha^2 v^2$.

So it is clear that $J^u(s, x) = E^{s,x}[\log(X_T)]$ is maximal if we for all r, z choose $u(s, z)$ to have the value of v which maximizes

$$av + b(1 - v) - \frac{1}{2}\alpha^2 v^2$$

i.e. we choose

$$v = u(t, X_t) = \frac{a - b}{\alpha^2} \quad \text{for all } t, \omega. \quad (11.64)$$

So this is the optimal control if the Kelly criterion is used. Similarly, this direct method also gives the optimal control when $N(x) = x^r$ (See Exercise 11.8).

EXAMPLE 11.6.

Finally we include an example which shows that even quite simple – and apparently innocent – stochastic control problems can lead us beyond the reach of the theory developed in this chapter:

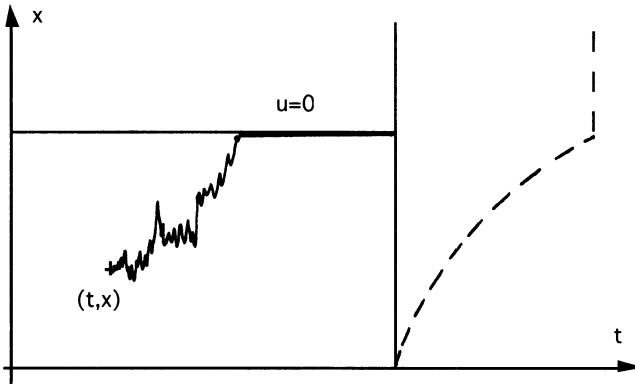
Suppose the system is a 1-dimensional Ito integral

$$dX_t = dX_t^u = u(t, \omega) dB_t, \quad t \geq s; \quad X_s = x > 0 \quad (11.65)$$

and consider the stochastic control problem

$$\Phi(t, x) = \sup_u E^{t,x}[K(X_\tau)], \quad (11.66)$$

where τ is the first exit time from $G = \{(r, z); r \leq t_1, z > 0\}$ for $Y_t = (s + t, X_{s+t}^{s,x})$ and K is a given bounded continuous function.



Intuitively, we can think of the system as the state of a game which behaves like an “excited” Brownian motion, where we can control the size u of the excitation at every instant. The purpose of the control is to maximize the expected payoff $K(X_{t_1})$ of the game at a fixed future time t_1 .

Assuming that $\Phi \in C^2$ and that u^* exists we get by the HJB (I) equation

$$\sup_{v \in \mathbf{R}} \left\{ \frac{\partial \Phi}{\partial t} + \frac{1}{2} v^2 \frac{\partial^2 \Phi}{\partial x^2} \right\} = 0 \quad \text{for } t < t_1, \quad \Phi(t_1, x) = K(x). \quad (11.67)$$

From this we see that we necessarily have

$$\frac{\partial^2 \Phi}{\partial x^2} \leq 0, \quad v^* \frac{\partial^2 \Phi}{\partial x^2} = 0 \quad \text{and} \quad \frac{\partial \Phi}{\partial t} = 0 \quad \text{for } t < t_1, \quad (11.68)$$

where v^* is the value of $v \in \mathbf{R}$ which gives the supremum in (11.67). But if $\frac{\partial \Phi}{\partial t} = 0$, then $\Phi(t, x) = \Phi(t_1, x) = K(x)$. However, this cannot possibly be the solution in general, because we have not assumed that $\frac{\partial^2 K}{\partial x^2} \leq 0$ – in fact, K was not even assumed to be differentiable.

What went wrong? Since the conclusion of the HJB (I) equation was wrong, the assumptions cannot hold. So either Φ is not C^2 or u^* does not exist, or both.

To simplify the problem assume that

$$K(x) = \begin{cases} x^2; & 0 \leq x \leq 1 \\ 1; & x > 1 \end{cases}$$

Then considering the figure above and using some intuition we see that it is optimal to excite as much as possible if X_t is in the strip $0 < x < 1$ to avoid exiting from G in the interval $\{t_1\} \times (0, 1)$. Using that X_t is just a time change of Brownian motion (see Chapter VIII) we conclude that this optimal control leads to a process X^* which jumps immediately to the value 1 with probability

x and to the value 0 with probability $1 - x$, if the starting point is $x \in (0, 1)$. If the starting point is $x \in [1, \infty)$ we simply choose our control to be zero. In other words, heuristically we should have

$$u^*(t, x) = \begin{cases} \infty & \text{if } x \in (0, 1) \\ 0 & \text{if } x \in [1, \infty) \end{cases} \quad (11.69)$$

with corresponding expected payoff

$$\phi^*(s, x) = E^{s,x}[K(X_{t_1}^*)] = \begin{cases} x & \text{if } 0 \leq x \leq 1 \\ 1 & \text{if } x > 1 \end{cases} \quad (11.70)$$

Thus we see that our candidate u^* for optimal control is not continuous (not even finite!) and the corresponding optimal process X_t^* is not an Ito diffusion (it is not even continuous). So to handle this case mathematically it is necessary to enlarge the family of admissible controls (and the family of corresponding processes). For example, one can prove an extended version of Theorem 11.2 which allows us to conclude that our choice of u^* above does indeed give at least as good performance as any other Markov control u and that ϕ^* given by (11.70) does coincide with the maximal expected payoff Φ defined by (11.66).

This last example illustrates the importance of the question of *existence* in general, both of the optimal control u^* and of the corresponding solution X_t of the stochastic differential equation (11.1). We briefly outline some results in this direction:

With certain conditions on $b, \sigma, F, \partial G$ and assuming that the set of control values is compact, one can show, using general results from nonlinear partial differential equations, that a smooth function ϕ exists such that

$$\sup_v \{F^v(y) + (L^v \phi)(y)\} = 0 \quad \text{for } y \in G$$

and

$$\phi(y) = K(y) \quad \text{for } y \in \partial G.$$

Then by a measurable selection theorem one can find a (measurable) function $u^*(y)$ such that

$$F^{u^*}(y) + (L^{u^*} \phi)(y) = 0, \quad (11.71)$$

for a.a. $y \in G$ w.r.t. Lebesgue measure in \mathbf{R}^{n+1} . Even if u^* is only known to be measurable, one can show that the corresponding solution $X_t = X_t^{u^*}$ of (11.1) exists (see Stroock and Varadhan (1979) for general results in this direction). Then by inspecting the proof of Theorem 11.2 one can see that it suffices to have (11.71) satisfied outside a subset of G with Green measure 0 (see Definition 9.19). Under suitable conditions on b and σ one can in fact show that the Green measure is absolutely continuous w.r.t. Lebesgue measure. Thus by (11.71) (and a strengthened Theorem 11.2) u^* is an optimal control. We refer

the reader to Fleming and Rishel (1975), Bensoussan and Lions (1978), Dynkin and Yushkevich (1979) and Krylov (1980) for details and further studies.

Stochastic control problems with terminal conditions

In many applications there are constraints on the types of Markov controls u to be considered, for example in terms of the probabilistic behaviour of Y_t^u at the terminal time $t = T$. Such problems can often be handled by applying a kind of “Lagrange multiplier” method, which we now describe:

Consider the problem of finding $\Phi(y)$ and $u^*(y)$ such

$$\Phi(y) = \sup_{u \in \mathcal{K}} J^u(y) \quad (11.72)$$

where

$$J^u(y) = E^y \left[\int_0^T F^u(Y_t^u) dt + K(Y_T^u) \right], \quad (11.73)$$

and where the supremum is taken over the space \mathcal{K} of all Markov controls $u : \mathbf{R}^{n+1} \rightarrow U \subset \mathbf{R}^k$ such that

$$E^y[M_i(Y_T^u)] = 0, \quad i = 1, 2, \dots, l, \quad (11.74)$$

where $M = (M_1, \dots, M_l) : \mathbf{R}^{n+1} \rightarrow \mathbf{R}^l$ is a given continuous function,

$$E^y[|M|(Y_T^u)] < \infty \quad \text{for all } y, u. \quad (11.75)$$

Now we introduce a related, but unconstrained problem as follows:

For each $\lambda \in \mathbf{R}^l$ and each Markov control u define

$$J_\lambda^u(y) = E^y \left[\int_0^T F^u(Y_t^u) dt + K(Y_T^u) + \lambda \cdot M(Y_T^u) \right] \quad (11.76)$$

where \cdot denotes the inner product in \mathbf{R}^l . Find $\Phi_\lambda(y)$ and $u_\lambda^*(y)$ such that

$$\Phi_\lambda(y) = \sup_u J_\lambda^u(y) = J_\lambda^{u_\lambda^*}(y), \quad (11.77)$$

without terminal conditions.

THEOREM 11.4.

Suppose that we for all $\lambda \in \Lambda \subset \mathbf{R}^l$ can find $\Phi_\lambda(y)$ and u_λ^* solving the (unconstrained) stochastic control problem (11.76)–(11.77). Moreover, suppose that there exists $\lambda_0 \in \Lambda$ such that

$$E^y[M(Y_T^{u_{\lambda_0}^*})] = 0. \quad (11.78)$$

Then $\Phi(y) := \Phi_{\lambda_0}(y)$ and $u^* := u_{\lambda_0}^*$ solves the constrained stochastic control problem (11.72)–(11.74).

Proof. Let u be a Markov control, $\lambda \in \Lambda$. Then by the definition of u_λ^* we have

$$\begin{aligned} E^y \left[\int_0^T F^{u_\lambda^*}(Y_t^{u_\lambda^*}) dt + K(Y_T^{u_\lambda^*}) + \lambda \cdot M(Y_T^{u_\lambda^*}) \right] &= J_\lambda^{u_\lambda^*}(y) \\ &\geq J_\lambda^u(y) = E^y \left[\int_0^T F^u(Y_t^u) dt + K(Y_T^u) + \lambda \cdot M(Y_T^u) \right]. \end{aligned} \quad (11.79)$$

In particular, if $\lambda = \lambda_0$ and $u \in \mathcal{K}$ then

$$E^y[M(Y_T^{u_{\lambda_0}^*})] = 0 = E^y[M(Y_T^u)]$$

and hence by (11.79)

$$J^{u_{\lambda_0}^*}(y) \geq J^u(y) \quad \text{for all } u \in \mathcal{K}.$$

Since $u_{\lambda_0}^* \in \mathcal{K}$ the proof is complete. \square

For an application of this result, see Exercise 11.11.

EXERCISES

11.1.

Write down the HJB equation for the problem

$$\Psi(s, x) = \inf_u E^{s,x} \left[\int_s^\infty e^{-\alpha t} (g(X_t) + u_t^2) dt \right]$$

where

$$dX_t = u_t dt + dB_t; \quad X_t, u_t, B_t \in \mathbf{R},$$

$\alpha > 0$ is a constant and $g : \mathbf{R} \rightarrow \mathbf{R}$ is a given bounded, continuous function. Show that if $\Psi \in C^2$ and u^* exists then

$$u^*(t, x) = -\frac{1}{2} e^{\alpha t} \frac{\partial \Psi}{\partial x}.$$

11.2.

Consider the stochastic control problem

$$\Psi_0(s, x) = \inf_u E^{s,x} \left[\int_s^{\infty} e^{-\rho t} f(u_t, X_t) dt \right],$$

where

$$dX_t = dX_t^u = b(u_t, X_t)dt + \sigma(u_t, X_t)dB_t; \quad X_t \in \mathbf{R}^n, u_t \in \mathbf{R}^k, B_t \in \mathbf{R}^m,$$

f is a given bounded continuous real function, $\rho > 0$ and the inf is taken over all *time-homogeneous* Markov controls u , i.e. controls u of the form $u = u(X_t)$. Prove that

$$\Psi_0(s, x) = e^{-\rho s} \xi(x), \quad \text{where } \xi(x) = \Psi(0, x).$$

(Hint: By definition of $E^{s,x}$ we have

$$E^{s,x} \left[\int_s^{\infty} e^{-\rho t} f(u(X_t), X_t) dt \right] = E \left[\int_0^{\infty} e^{-\rho(s+t)} f(u(X_{s+t}^{s,x}), X_{s+t}^{s,x}) dt \right]$$

where E denotes expectation w.r.t. P^0 .)

11.3.

Define

$$dX_t = r u_t X_t dt + \alpha u_t X_t dB_t; \quad X_t, u_t, B_t \in \mathbf{R}$$

and

$$\Phi(s, x) = \sup_u E^{s,x} \left[\int_s^{\infty} e^{-\rho t} f(X_t) dt \right],$$

where r, α, ρ are constants, $\rho > 0$ and f is a bounded continuous real function.

Assume that $\Phi \in C^2$ and that an optimal Markov control u^* exists.

a) Show that

$$\sup_{v \in \mathbf{R}} \left\{ e^{-\rho t} f(x) + \frac{\partial \Phi}{\partial t} + rvx \frac{\partial \Phi}{\partial x} + \frac{1}{2} \alpha^2 v^2 x^2 \frac{\partial^2 \Phi}{\partial x^2} \right\} = 0.$$

Deduce that

$$\frac{\partial^2 \Phi}{\partial x^2} \leq 0.$$

b) Assume that $\frac{\partial^2 \Phi}{\partial x^2} < 0$. Prove that

$$u^*(t, x) = - \frac{r \frac{\partial \Phi}{\partial x}}{\alpha^2 x \frac{\partial^2 \Phi}{\partial x^2}}$$

and that

$$2\alpha^2(e^{-\rho t}f + \frac{\partial\Phi}{\partial t})\frac{\partial^2\Phi}{\partial x^2} - r^2(\frac{\partial\Phi}{\partial x})^2 = 0.$$

c) Assume that $\frac{\partial^2\Phi}{\partial x^2} = 0$. Prove that $\frac{\partial\Phi}{\partial x} = 0$ and

$$e^{-\rho t}f(x) + \frac{\partial\Phi}{\partial t} = 0.$$

d) Assume that $u_t^* = u^*(X_t)$ and that b) holds. Prove that $\Phi(t, x) = e^{-\rho t}\xi(x)$ and

$$2\alpha^2(f - \rho\xi)\xi'' - r^2(\xi')^2 = 0.$$

(See Exercise 11.2)

11.4.

The assumptions in Theorem 11.1 often fail (see e.g. Exercise 11.10), so it is useful to have results in such cases also. For example, if we define Φ_a as in Theorem 11.3 then, without assuming that u^* exists and without smoothness conditions on Φ , we have the *Bellman principle* (compare with (11.15)–(11.16))

$$\Phi_a(y) = \sup_u E^y \left[\int_0^\alpha F^u(Y_r^u) dr + \Phi_a(Y_\alpha^u) \right]$$

for all $y \in G$ and all stopping times $\alpha \leq \tau$, the sup being taken over all \mathcal{F}_t -adapted controls u . (See Krylov (1980, Th. 6, p. 150).)

Deduce that if $\Phi_a \in C^2(G)$ then

$$F^v(y) + L^v\Phi_a(y) \leq 0 \quad \text{for all } y \in G, v \in U.$$

11.5.

Assume that $F = 0$ in (11.8) and that an optimal Markov control u^* exists. Prove that the function Φ is *superharmonic* in G w.r.t. the process Y_t^u , for any Markov control u . (Hint: See (11.15)–(11.16).)

11.6.

Let X_t denote your wealth at time t . Suppose that at any time t you have a choice between two investments:

- 1) A risky investment where the unit price $p_1 = p_1(t, \omega)$ satisfies the equation

$$dp_1 = a_1 p_1 dt + \sigma_1 p_1 dB_t$$

- 2) A safer (less risky) investment where the unit price $p_2 = p_2(t, \omega)$ satisfies

$$dp_2 = a_2 p_2 dt + \sigma_2 p_2 d\tilde{B}_t$$

where a_i, σ_i are constants such that

$$a_1 > a_2, \quad \sigma_1 > \sigma_2$$

and B_t, \tilde{B}_t are independent 1-dimensional Brownian motions.

- a) Let $u(t, \omega)$ denote the fraction of the fortune $X_t(\omega)$ which is placed in the risky investment at time t . Show that

$$dX_t = dX_t^{(u)} = X_t(a_1 u + a_2(1 - u))dt + X_t(\sigma_1 u dB_t + \sigma_2(1 - u)d\tilde{B}_t)$$

- b) Assuming that u is a Markov control, $u = u(t, X_t)$, find the generator A^u of (t, X_t^u) .

- c) Write down the HJB equation for the stochastic control problem

$$\Phi(s, x) = \sup_u E^{s,x}[\sqrt{X_T^{(u)}}]$$

where $T = \min(t_1, \tau_0)$, $\tau_0 = \inf\{t > s; X_t = 0\}$ and t_1 is a given future time (constant).

- d) Find the optimal control u^* for the problem in c).

11.7.

Consider the stochastic control problem

$$(system) \quad dX_t = audt + u dB_t; \quad X_0 = x > 0$$

where $B_t \in \mathbf{R}$ and $a \in \mathbf{R}$ is a given constant, and

$$(performance) \quad \Phi(s, x) = \sup_u E^{s,x}[(X_T)^r],$$

where $0 < r < 1$ and

$$T = \inf\{t > s; X_t = 0\} \wedge t_1,$$

t_1 being a given future time (constant).

Show that this problem has the optimal control

$$u^*(t, x) = \frac{ax}{1 - r}$$

with corresponding optimal performance

$$\Phi(s, x) = x^r \exp\left(\frac{a^2(T - s)r}{2(1 - r)}\right).$$

11.8.

Use Dynkin's formula to prove directly that

$$u^*(t, x) = \min\left(\frac{a - b}{\alpha^2(1 - r)}, 1\right)$$

is the optimal control for the problem in Example 11.5, with utility function $N(x) = x^r$. (Hint: See the argument leading to (11.64).)

11.9.

In Benes (1974) the following stochastic control problem is considered:

$$\Psi(s, x) = \inf_u E^{s,x} \left[\int_s^\infty e^{-\rho t} X_t^2 dt \right],$$

where

$$dX_t = dX_t^{(u)} = au_t dt + dB_t; \quad X_t, B_t \in \mathbf{R}$$

and a, ρ are (known) constants, $\rho > 0$. Here the controls u are restricted to take values in $U = [-1, 1]$.

a) Show that the HJB equation for this problem is

$$\inf_{v \in [-1, 1]} \left\{ e^{-\rho s} x^2 + \frac{\partial \Psi}{\partial s} + av \frac{\partial \Psi}{\partial x} + \frac{1}{2} \cdot \frac{\partial^2 \Psi}{\partial x^2} \right\} = 0.$$

b) If $\Psi \in C^2$ and u^* exists, show that

$$u^*(x) = -\text{sign}(ax),$$

where

$$\text{sign } z = \begin{cases} 1 & \text{if } z > 0 \\ -1 & \text{if } z \leq 0 \end{cases}$$

(Hint: Explain why $x > 0 \Rightarrow \frac{\partial \Psi}{\partial x} > 0$ and $x < 0 \Rightarrow \frac{\partial \Psi}{\partial x} < 0$)

11.10.

Let

$$f(x) = \begin{cases} x^2 & \text{for } 0 \leq x \leq 1 \\ \sqrt{x} & \text{for } x > 1 \end{cases}$$

and put

$$J^u(s, x) = E^{s,x} \left[\int_s^T e^{-\rho t} f(X_t^u) dt \right], \Phi(s, x) = \sup_u J^u(s, x)$$

where

$$dX_t^u = u_t dB_t; \quad t \geq s$$

with control values $u_t \in \mathbf{R}$, $B_t \in \mathbf{R}$ and

$$T = \inf\{t > s; X_t^u \leq 0\}.$$

a) Define

$$\phi(s, x) = \frac{1}{\rho} e^{-\rho s} \hat{f}(x) \quad \text{for } x \geq 0, s \in \mathbf{R}$$

where

$$\hat{f}(x) = \begin{cases} x & \text{for } 0 \leq x \leq 1 \\ \sqrt{x} & \text{for } x > 1 \end{cases}$$

Prove that

$$J^u(s, x) \leq \phi(s, x)$$

for all s, x and all (finite) Markov controls u .

(Hint: Put $\phi_1(s, x) = \frac{1}{\rho} e^{-\rho s} x$ for all s, x and $\phi_2(s, x) = \frac{1}{\rho} e^{-\rho s} \sqrt{x}$ for all s, x . Then

$$J^u(s, x) \leq \phi_i(s, x) \quad \text{for } i = 1, 2$$

by Theorem 11.2.)

b) Show that

$$\Phi(s, x) = \phi(s, x)$$

(Hint: Consider $J^{u_k}(s, x)$, where

$$u_k(x) = \begin{cases} k & \text{for } 0 \leq x < 1 \\ 0 & \text{for } x \geq 1 \end{cases}$$

and let $k \rightarrow \infty$).

Thus u^* does not exist and H is not a C^2 function. Hence both conditions for the HJB (I) equation fail in this case.

11.11.

Consider a 1-dimensional version of the stochastic linear regulator problem of Example 11.4:

$$\Psi(s, x) = \inf_{u \in \mathcal{K}} E^{s, x} \left[\int_s^{t_1} ((X_r^u)^2 + \theta u_r^2) dr \right] \quad (11.80)$$

where

$$dX_t^u = u_t dt + \sigma dB_t; \quad \text{for } t \geq s, X_s = x,$$

$u_t, B_t \in \mathbf{R}$, σ, θ constants, $\theta > 0$, the infimum being over the space \mathcal{K} of all Markov controls u satisfying

$$E^{s, x}[(X_{t_1}^u)^2] = m^2, \quad \text{where } m \text{ is a constant.} \quad (11.81)$$

Solve this problem by using Theorem 11.4.

(Hint: Solve for each $\lambda \in \mathbf{R}$ the unconstrained problem

$$\Psi_\lambda(s, x) = \inf_u E^{s,x} \left[\int_s^{t_1} ((X_r^u) + \theta u_r^2) dr + \lambda (X_{t_1}^u)^2 \right]$$

with optimal control u_λ^* . Then try to find λ_0 such that

$$E^{s,x}[(X_{t_1}^{u_{\lambda_0}^*})^2] = m^2.$$

11.12.

Solve the stochastic control problem

$$\Psi(s, x) = \inf_u J^u(s, x) = J^{u^*}(s, x)$$

where

$$J^u(s, x) = E^{s,x} \left[\int_s^\infty e^{-\rho r} (X_r^2 + \theta u_r^2) dr \right]$$

and

$$dX_t = u_t dt + \sigma dB_t,$$

with $u_t, B_t \in \mathbf{R}$ and $\sigma \in \mathbf{R}$, $\rho > 0$, $\theta > 0$ are constants. (Hint: Try $\psi(s, x) = e^{-\rho s}(ax^2 + b)$ for suitable constants a, b and apply Theorem 11.2.)

Appendix A: Normal Random Variables

Here we recall some basic facts which are used in the text.

DEFINITION A.1. Let (Ω, \mathcal{F}, P) be a probability space. A random variable $X : \Omega \rightarrow \mathbf{R}$ is *normal* if the distribution of X has a density of the form

$$p_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right), \quad (A.1)$$

where $\sigma > 0$ and m are constants. In other words,

$$P[X \in G] = \int_G p_X(x)dx, \text{ for all Borel sets } G \subset \mathbf{R}.$$

If this is the case, then

$$E[X] = \int_{\Omega} XdP = \int_{\mathbf{R}} xp_X(x)dx = m \quad (A.2)$$

and

$$\text{var}[X] = E[(X - m)^2] = \int_{\mathbf{R}} (x - m)^2 p_X(x)dx = \sigma^2. \quad (A.3)$$

More generally, a random variable $X : \Omega \rightarrow \mathbf{R}^n$ is called (*multi*) *normal* $\mathcal{N}(m, C)$ if the distribution of X has a density of the form

$$p_X(x_1, \dots, x_n) = \frac{\sqrt{|A|}}{(2\pi)^{n/2}} \cdot \exp\left(-\frac{1}{2} \cdot \sum_{j,k} (x_j - m_j)a_{jk}(x_k - m_k)\right) \quad (A.4)$$

where $m = (m_1, \dots, m_n) \in \mathbf{R}^n$ and $C^{-1} = A = [a_{jk}] \in \mathbf{R}^{n \times n}$ is a positive definite matrix.

If this is the case then

$$E[X] = m \quad (A.5)$$

and

$$\begin{aligned} A^{-1} = C = [c_{jk}] &\text{ is the covariance matrix of } X, \text{ i.e.} \\ c_{jk} &= E[(X_j - m_j)(X_k - m_k)]. \end{aligned} \quad (A.6)$$

DEFINITION A.2. The characteristic function of a random variable $X : \Omega \rightarrow \mathbf{R}^n$ is the function $\phi_X : \mathbf{R}^n \rightarrow \mathbf{C}$ (where \mathbf{C} denotes the complex numbers) defined by

$$\phi_X(u_1, \dots, u_n) = E[\exp(i(u_1 X_1 + \dots + u_n X_n))] = \int_{\mathbf{R}^n} e^{i \langle u, x \rangle} \cdot P[X \in dx], \quad (A.7)$$

where $\langle u, x \rangle = u_1 x_1 + \dots + u_n x_n$ (and $i \in \mathbf{C}$ is the imaginary unit). In other words, ϕ_X is the Fourier transform of X (or, more precisely, of the measure $P[X \in dx]$). Therefore we have

THEOREM A.3. The characteristic function of X determines the distribution of X uniquely.

It is not hard to verify the following:

THEOREM A.4. If $X : \Omega \rightarrow \mathbf{R}^n$ is normal $\mathcal{N}(m, C)$, then

$$\phi_X(u_1, \dots, u_n) = \exp\left(-\frac{1}{2} \sum_{j,k} u_j c_{jk} u_k + i \sum_j u_j m_j\right). \quad (A.8)$$

Theorem A.4 is often used as a basis for an extended concept of a normal random variable: We define $X : \Omega \rightarrow \mathbf{R}^n$ to be normal (in the extended sense) if ϕ_X satisfies (A.8) for some symmetric non-negative definite matrix $C = [c_{jk}] \in \mathbf{R}^{n \times n}$ and some $m \in \mathbf{R}^n$. So by this definition it is not required that C be invertible. From now on we will use this extended definition of normality. In the text we often use the following result:

THEOREM A.5. Let $X_j : \Omega \rightarrow \mathbf{R}$ be random variables; $1 \leq j \leq n$. Then

$$X = (X_1, \dots, X_n) \text{ is normal}$$

if and only if

$$Y = \lambda_1 X_1 + \dots + \lambda_n X_n \text{ is normal for all } \lambda_1, \dots, \lambda_n \in \mathbf{R}.$$

Proof. If X is normal, then

$$\begin{aligned} E[\exp(iu(\lambda_1 X_1 + \dots + \lambda_n X_n))] &= \exp\left(-\frac{1}{2} \sum_{j,k} u \lambda_j c_{jk} u \lambda_k + i \sum_j u \lambda_j m_j\right) \\ &= \exp\left(-\frac{1}{2} u^2 \sum_{j,k} \lambda_j c_{jk} \lambda_k + i u \sum_j \lambda_j m_j\right), \end{aligned}$$

so Y is normal with $E[Y] = \sum \lambda_j m_j$, $\text{var}[Y] = \sum \lambda_j c_{jk} \lambda_k$.

Conversely, if $Y = \lambda_1 X_1 + \dots + \lambda_n X_n$ is normal with $E[Y] = m$ and $\text{var}[Y] = \sigma^2$, then

$$E[\exp(iu(\lambda_1 X_1 + \dots + \lambda_n X_n))] = \exp\left(-\frac{1}{2}u^2\sigma^2 + ium\right),$$

where

$$\begin{aligned} m &= \sum_j \lambda_j E[X_j], \sigma^2 = E[(\sum_j \lambda_j X_j - \sum_j \lambda_j E[X_j])^2] \\ &= E[(\sum_j \lambda_j (X_j - m_j))^2] = \sum_{j,k} \lambda_j \lambda_k E[(X_j - m_j)(X_k - m_k)], \end{aligned}$$

where $m_j = E[X_j]$. Hence X is normal.

THEOREM A.6. Let Y_0, Y_1, \dots, Y_n be real, random variables on Ω . Assume that $X = (Y_0, Y_1, \dots, Y_n)$ is normal and that Y_0 and Y_j are uncorrelated for each $j \geq 1$, i.e

$$E[(Y_0 - E[Y_0])(Y_j - E[Y_j])] = 0; \quad 1 \leq j \leq n.$$

Then Y_0 is independent of $\{Y_1, \dots, Y_n\}$.

Proof. We have to prove that

$$P[Y_0 \in G_0, Y_1 \in G_1, \dots, Y_n \in G_n] = P[Y_0 \in G_0] \cdot P[Y_1 \in G_1, \dots, Y_n \in G_n], \quad (A.9)$$

for all Borel sets $G_0, G_1, \dots, G_n \subset \mathbf{R}$.

We know that in the first line (and the first column) of the covariance matrix $c_{jk} = E[(Y_j - E[Y_j])(Y_k - E[Y_k])]$ only the first entry $c_{00} = \text{var}[Y_0]$, is non-zero. Therefore the density of X satisfies

$$p_X(x_0, \dots, x_n) = p_{Y_0}(x_0) \cdot p_{Y_1, \dots, Y_n}(x_1, \dots, x_n),$$

and this is equivalent to (A.9).

Finally we establish the following:

THEOREM A.7. Suppose $X_k : \Omega \rightarrow \mathbf{R}^n$ is normal for all k and that $X_k \rightarrow X$ in $L^2(\Omega)$, i.e.

$$E[|X_k - X|^2] \rightarrow 0 \text{ as } k \rightarrow \infty$$

Then X is normal.

Proof. Since $|e^{i\langle u, x \rangle} - e^{i\langle u, y \rangle}| < |u| \cdot |x - y|$, we have

$$E[\{\exp(i \langle u, X_k \rangle) - \exp(i \langle u, X \rangle)\}^2] \leq |u|^2 \cdot E[|X_k - X|^2] \rightarrow 0 \text{ as } k \rightarrow \infty.$$

Therefore

$$E[\exp(i \langle u, X_k \rangle)] \rightarrow E[\exp(i \langle u, X \rangle)] \text{ as } k \rightarrow \infty.$$

So X is normal, with mean $E[X] = \lim E[X_k]$ and covariance matrix $C = \lim C_k$, where C_k is the covariance matrix of X_k .

Appendix B: Conditional Expectations

Let (Ω, \mathcal{F}, P) be a probability space and let $X : \Omega \rightarrow \mathbf{R}^n$ be a random variable such that $E[|X|] < \infty$. If $\mathcal{H} \subset \mathcal{F}$ is a σ -algebra, then the *conditional expectation* of X given \mathcal{H} , denoted by $E[X|\mathcal{H}]$, is defined as follows:

DEFINITION B.1. $E[X|\mathcal{H}]$ is the (a.s. unique) function from Ω to \mathbf{R}^n satisfying:

- (1) $E[X|\mathcal{H}]$ is \mathcal{H} -measurable
- (2) $\int_H E[X|\mathcal{H}]dP = \int_H XdP$, for all $H \in \mathcal{H}$.

The existence and uniqueness of $E[X|\mathcal{H}]$ comes from the Radon-Nikodym theorem: Let μ be the measure on \mathcal{H} defined by

$$\mu(H) = \int_H XdP; \quad H \in \mathcal{H}.$$

Then μ is absolutely continuous w.r.t. $P|\mathcal{H}$, so there exists a $P|\mathcal{H}$ -unique \mathcal{H} -measurable function F on Ω such that

$$\mu(H) = \int_H FdP \text{ for all } H \in \mathcal{H}.$$

Thus $E[X|\mathcal{H}] := F$ does the job and this function is unique a.s. w.r.t. the measure $P|\mathcal{H}$.

We list some of the basic properties of the conditional expectation:

THEOREM B.2. Suppose $Y : \Omega \rightarrow \mathbf{R}^n$ is another random variable with $E[|Y|] < \infty$ and let $a, b \in \mathbf{R}$. Then

- a) $E[aX + bY|\mathcal{H}] = aE[X|\mathcal{H}] + bE[Y|\mathcal{H}]$
- b) $E[E[X|\mathcal{H}]] = E[X]$
- c) $E[X|\mathcal{H}] = X$ if X is \mathcal{H} -measurable
- d) $E[X|\mathcal{H}] = E[X]$ if X is independent of \mathcal{H}

- e) $E[Y \cdot X | \mathcal{H}] = Y \cdot E[X | \mathcal{H}]$ if Y is \mathcal{H} -measurable, where \cdot denotes the usual inner product in \mathbf{R}^n .

Proof. d): If X is independent of \mathcal{H} we have for $H \in \mathcal{H}$

$$\int_H X dP = \int_{\Omega} X \cdot \chi_H dP = \int_{\Omega} X dP \cdot \int_{\Omega} \chi_H dP = E[X] \cdot P(H),$$

so the constant $E[X]$ satisfies (1) and (2).

- e): We first establish the result in the case when $Y = \chi_H$ (where χ denotes the indicator function), for some $H \in \mathcal{H}$.

Then for all $G \in \mathcal{G}$

$$\int_G Y \cdot E[X | \mathcal{H}] dP = \int_{G \cap H} E[X | \mathcal{H}] dP = \int_{G \cap H} X dP = \int_G Y X dP,$$

so $Y \cdot E[X | \mathcal{H}]$ satisfies both (1) and (2). Similarly we obtain that the result is true if Y is a simple function

$$Y = \sum_{j=1}^m c_j \chi_{H_j}, \text{ where } H_j \in \mathcal{H}.$$

The result in the general case then follows by approximating Y by such simple functions.

THEOREM B.3. Let \mathcal{G}, \mathcal{H} be σ -algebras such that $\mathcal{G} \subset \mathcal{H}$. Then

$$E[X | \mathcal{G}] = E[E[X | \mathcal{H}] | \mathcal{G}].$$

Proof. If $G \in \mathcal{G}$ then $G \in \mathcal{H}$ and therefore

$$\int_G E[X | \mathcal{H}] dP = \int_G X dP.$$

Hence $E[E[X | \mathcal{H}] | \mathcal{G}] = E[X | \mathcal{G}]$ by uniqueness.

Appendix C: Uniform Integrability and Martingale Convergence

We give a brief summary of the definitions and results which are the background for the applications in this book. For proofs and more information we refer to Doob (1984), Liptser and Shirayev (1977), Meyer (1966) or Williams (1979).

DEFINITION C.1. Let (Ω, \mathcal{F}, P) be a probability space. A family $\{f_j\}_{j \in J}$ of real, measurable functions f_j on Ω is called *uniformly integrable* if

$$\lim_{M \rightarrow \infty} \left(\sup_{j \in J} \left\{ \int_{\{|f_j| > M\}} |f_j| dP \right\} \right) = 0.$$

One of the most useful tests for uniform integrability is obtained by using the following concept:

DEFINITION C.2. A function $\psi : [0, \infty) \rightarrow [0, \infty)$ is called a *u.i. (uniform integrability) test function* if ψ is increasing, convex (i.e. $\psi(\lambda x + (1 - \lambda)y) \leq \lambda\psi(x) + (1 - \lambda)\psi(y)$ for all $x, y \in [0, \infty)$, $\lambda \in [0, 1]$) and

$$\lim_{x \rightarrow \infty} \frac{\psi(x)}{x} = \infty.$$

So for example $\psi(x) = x^p$ is a u.i. test function if $p > 1$, but not if $p = 1$.

The justification for the name in Definition C.2 is the following:

THEOREM C.3. The family $\{f_j\}_{j \in J}$ is uniformly integrable if and only if there is a u.i. test function ψ such that

$$\sup_{j \in J} \left\{ \int \psi(|f_j|) dP \right\} < \infty$$

One major reason for the usefulness of uniform integrability is the following result, which may be regarded as the ultimate generalization of the various convergence theorems in integration theory:

THEOREM C.4. Suppose $\{f_k\}_{k=1}^\infty$ is a sequence of real measurable functions on Ω such that

$$\lim_{k \rightarrow \infty} f_k(\omega) = f(\omega) \text{ for a.a.}\omega.$$

Then the following are equivalent:

- 1) $\{f_k\}$ is uniformly integrable
- 2) $f \in L^1(P)$ and $f_k \rightarrow f$ in $L^1(P)$, i.e. $\int |f_k - f| dP \rightarrow 0$ as $k \rightarrow \infty$.

An important application of uniform integrability is within the convergence theorems for martingales:

Let (Ω, \mathcal{N}, P) be a probability space and let $\{\mathcal{N}_t\}_{t \geq 0}$ be an increasing family of σ -algebras, $\mathcal{N}_t \subset \mathcal{N}$ for all t . A stochastic process $N_t : \Omega \rightarrow \mathbf{R}$ is called a *supermartingale* (w.r.t. $\{\mathcal{N}_t\}$) if N_t is \mathcal{N}_t -adapted, $E[|N_t|] < \infty$ for all t and

$$N_t \geq E[N_s | \mathcal{N}_t] \text{ for all } s > t. \quad (C.1)$$

Similarly, if (C.1) holds with the inequality reversed for all $s > t$, then N_t is called a *submartingale*. And if (C.1) holds with equality then N_t is called a *martingale*.

As is customary we will assume that each \mathcal{N}_t contains all the null sets of \mathcal{N} , that $t \rightarrow N_t(\omega)$ is right continuous for a.a. ω and that $\{\mathcal{N}_t\}$ is right continuous, in the sense that $\mathcal{N}_t = \bigcap_{s > t} \mathcal{N}_s$ for all $t \geq 0$.

THEOREM C.5. (Doob's martingale convergence theorem I)

Let N_t be a right continuous supermartingale with the property that

$$\sup_{t > 0} E[N_t^-] < \infty,$$

where $N_t^- = \max(-N_t, 0)$. Then the pointwise limit

$$N(\omega) = \lim_{t \rightarrow \infty} N_t(\omega)$$

exists for a.a. ω and $E[N^-] < \infty$.

Note, however, that the convergence need not be in $L^1(P)$. In order to obtain this we need uniform integrability:

THEOREM C.6. (Doob's martingale convergence theorem II).

Let N_t be a right-continuous supermartingale. Then the following are equivalent:

- 1) $\{N_t\}_{t \geq 0}$ is uniformly integrable
- 2) There exists $N \in L^1(P)$ such that $N_t \rightarrow N$ a.e. (P) and $N_t \rightarrow N$ in $L^1(P)$, i.e. $\int |N_t - N| dP \rightarrow 0$ as $t \rightarrow \infty$.

Combining Theorems C.6 and C.3 (with $\psi(x) = x^p$) we get

COROLLARY C.7. Let M_t be a continuous martingale such that

$$\sup_{t>0} E[|M_t|^p] < \infty \text{ for some } p > 1.$$

Then there exists $M \in L^1(P)$ such that $M_t \rightarrow M$ a.e. (P) and

$$\int |M_t - M| dP \rightarrow 0 \text{ as } t \rightarrow \infty.$$

Finally, we mention that similar results can be obtained for the analogous discrete time super/sub-martingales $\{N_k, \mathcal{N}_k\}$, $k = 1, 2, \dots$. Of course, no continuity assumptions are needed in this case. For example, we have the following result, which is used in Chapter IX:

COROLLARY C.8. Let $M_k; k = 1, 2, \dots$ be a discrete time martingale and assume that

$$\sup_k E[|M_k|^p] < \infty \text{ for some } p > 1.$$

Then there exists $M \in L^1(P)$ such that $M_k \rightarrow M$ a.e. (P) and

$$\int |M_k - M| dP \rightarrow 0 \text{ as } k \rightarrow \infty.$$

COROLLARY C.9. Let $X \in L^1(P)$, let $\{\mathcal{N}_k\}_{k=1}^\infty$ be an increasing family of σ -algebras, $\mathcal{N}_k \subset \mathcal{F}$ and define \mathcal{N}_∞ to be the σ -algebra generated by $\{\mathcal{N}_k\}_{k=1}^\infty$. Then

$$E[X|\mathcal{N}_k] \rightarrow E[X|\mathcal{N}_\infty] \text{ as } k \rightarrow \infty,$$

a.e. P and in $L^1(P)$.

Proof. $M_k := E[X|\mathcal{N}_k]$ is a u.i. martingale, so there exists $M \in L^1(P)$ such that $M_k \rightarrow M$ a.e. P and in $L^1(P)$, as $k \rightarrow \infty$. It remains to prove that $M = E[X|\mathcal{N}_\infty]$: Note that

$$\begin{aligned} \|M_k - E[M|\mathcal{N}_k]\|_{L^1(P)} &= \|E[M_k|\mathcal{N}_k] - E[M|\mathcal{N}_k]\|_{L^1(P)} \\ &\leq \|M_k - M\|_{L^1(P)} \rightarrow 0 \text{ as } k \rightarrow \infty. \end{aligned}$$

Hence if $F \in \mathcal{N}_{k_0}$ and $k \geq k_0$ we have

$$\int_F (X - M) dP = \int_F E[X - M|\mathcal{N}_k] dP = \int_F (M_k - E[M|\mathcal{N}_k]) dP \rightarrow 0 \text{ as } k \rightarrow \infty.$$

Therefore

$$\int_F (X - M) dP = 0 \text{ for all } F \in \bigcup_{k=1}^\infty \mathcal{N}_k$$

and hence

$$E[X|\mathcal{N}_\infty] = E[M|\mathcal{N}_\infty] = M.$$

□

Solutions and Additional Hints to Some of the Exercises

2.12.

$$P^0[B_t \in D_\rho] = 1 - e^{-\frac{\rho^2}{2t}}.$$

2.13.

$$E^x \left[\int_0^\infty \chi_H(B_t) dt \right] = \int_0^\infty P^x[B_t \in H] dt = \int_0^\infty (2\pi t)^{-n/2} \left(\int_H e^{-\frac{(x-y)^2}{2t}} dy \right) dt = 0$$

for all $x \in \mathbf{R}^n$, if $H \subset \mathbf{R}^n$ has n-dimensional Lebesgue measure equal to 0.

2.14.

$$\begin{aligned} P[\tilde{B}_{t_1} \in F_1, \dots, \tilde{B}_{t_k} \in F_k] &= P[B_{t_1} \in U^{-1}F_1, \dots, B_{t_k} \in U^{-1}F_k] \\ &= \int_{U^{-1}F_1 \times \dots \times U^{-1}F_k} p(t_1, 0, x_1)p(t_2 - t_1, x_1, x_2) \cdots p(t_k - t_{k-1}, x_{k-1}, x_k) dx_1 \cdots dx_k \\ &= \int_{F_1 \times \dots \times F_k} p(t_1, 0, y_1)p(t_2 - t_1, y_1, y_2) \cdots p(t_k - t_{k-1}, y_{k-1}, y_k) dy_1 \cdots dy_k \\ &= P[B_{t_1} \in F_1, \dots, B_{t_k} \in F_k], \end{aligned}$$

by (2.1), using the substitutions $y_j = Ux_j$ and the fact that $|Ux_j - Ux_{j-1}|^2 = |x_j - x_{j-1}|^2$.

2.15.

$$\begin{aligned} \text{a)} \quad E[(Y_n(t, \cdot) - t)^2] &= E \left[\left(\sum_{k=0}^{2^n-1} (\Delta B_k)^2 - \sum_{k=0}^{2^n-1} 2^{-n}t \right)^2 \right] \\ &= E \left[\left\{ \sum_{k=0}^{2^n-1} ((\Delta B_k)^2 - 2^{-n}t) \right\}^2 \right] = E \left[\left(\sum_{j,k=0}^{2^n-1} ((\Delta B_j)^2 - 2^{-n}t)((\Delta B_k)^2 - 2^{-n}t) \right) \right] \\ &= \sum_{k=0}^{2^n-1} E[((\Delta B_k)^2 - 2^{-n}t)^2] = \sum_{k=0}^{2^n-1} E[(\Delta B_k)^4 - 2 \cdot 2^{-2n}t^2 - 2^{-2n}t^2] \\ &= \sum_{k=0}^{2^n-1} 2 \cdot 2^{-2n}t^2 = 2 \cdot 2^{-n}t^2 \rightarrow 0 \quad \text{as } n \rightarrow \infty. \end{aligned}$$

- b) This follows from the following general result: If the quadratic variation of a real function over an interval is positive, then the total variation of the function over that interval is infinite.

3.1.

$$\begin{aligned} tB_t &= \sum_{j=0}^{n-1} \Delta(s_j B_j) = \sum_{j=0}^{n-1} s_j \Delta B_j + \sum_{j=0}^{n-1} B_{j+1} \Delta s_j \\ &\rightarrow \int_0^t s dB_s + \int_0^t B_s ds \quad \text{as } n \rightarrow \infty. \end{aligned}$$

3.4.

The processes in (iii) and (iv) are martingales, but those in (i) and (ii) are not.

3.9.

$$\int_0^T B_t \circ dB_t = \frac{1}{2} B_T^2 \quad \text{if } B_0 = 0.$$

3.12.

(i) a) $dX_t = (\gamma + \frac{1}{2}\alpha^2)X_t dt + \alpha X_t dB_t.$

b) $dX_t = \frac{1}{2} \sin X_t [\cos X_t - t^2] dt + (t^2 + \cos X_t) dB_t$

(ii) a) $dX_t = (r - \frac{1}{2}\alpha^2)X_t dt + \alpha X_t \circ dB_t.$

b) $dX_t = (2e^{-X_t} - X_t^3)dt + X_t^2 \circ dB_t.$

4.1.

a) $dX_t = 2B_t dB_t + dt.$

b) $dX_t = (1 + \frac{1}{2}eB_t)dt + e^{B_t} dB_t.$

c) $dX_t = 2dt + 2B_1 dB_1(t) + 2B_2 dB_2(t).$

d) $dx_t = \begin{bmatrix} dt \\ dB_t \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} dt + \begin{bmatrix} 0 \\ 1 \end{bmatrix} dB_t.$

e) $dX_1(t) = dB_1(t) + dB_2(t) + dB_3(t)$

$$dX_2(t) = dt - B_3(t)dB_1(t) + 2B_2(t)dB_2(t) - B_1(t)dB_3(t)$$

or

$$dX_t = \begin{bmatrix} dX_1(t) \\ dX_2(t) \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} dt + \begin{bmatrix} 1 & 1 & 1 \\ -B_3(t) & 2B_2(t) & -B_1(t) \end{bmatrix} \begin{bmatrix} dB_1(t) \\ dB_2(t) \\ dB_3(t) \end{bmatrix}$$

4.5.

$$E[B_t^6] = 15t^3 \quad \text{if } B_0 = 0.$$

5.3.

$$X_t = X_0 \cdot \exp((r - \frac{1}{2} \sum_{k=1}^n \alpha_k^2)t + \sum_{k=1}^n \alpha_k B_k(t)) \quad (\text{if } B(0) = 0).$$

5.4.

(i) $X_1(t) = X_1(0) + t + B_1(t),$

$$X_2(t) = X_2(0) + X_1(0)B_2(t) + \int_0^t s dB_2(s) + \int_0^t B_1(s) dB_2(s), \quad \text{assuming (as usual) that } B(0) = 0.$$

(ii) $X_t = e^t X_0 + \int_0^t e^{t-s} dB_t.$

(iii) $X_t = e^{-t} X_0 + e^{-t} B_t \quad (\text{assuming } B_0 = 0).$

5.6.

$$Y_t = \exp(\alpha B_t - \frac{1}{2} \alpha^2 t)[Y_0 + r \int_0^t \exp(-\alpha B_s + \frac{1}{2} \alpha^2 s) ds] \quad (B_0 = 0).$$

5.7.

a) $X_t = m + (X_0 - m)e^{-t} + \sigma \int_0^t e^{s-t} dB_s.$

b) $E[X_t] = m + (X_0 - m)e^{-t}.$

$$\text{Var}[X_t] = \frac{\sigma^2}{2}[1 - e^{-2t}].$$

5.8.

$$X(t) = \begin{bmatrix} X_1(t) \\ X_2(t) \end{bmatrix} = \exp(tJ)X(0) + \exp(tJ) \int_0^t \exp(-sJ)M dB(s),$$

where

$$J = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad M = \begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix}, \quad dB(s) = \begin{bmatrix} dB_1(s) \\ dB_2(s) \end{bmatrix}$$

and

$$\exp(tJ) = I + tJ + \frac{t^2}{2} J^2 + \cdots + \frac{t^n}{n!} J^n + \cdots \in \mathbf{R}^{2 \times 2}.$$

Using that $J^2 = I$ this can be rewritten as

$$X_1(t) = X_1(0) \cosh(t) + X_2(0) \sinh(t) + \int_0^t \alpha \cosh(t-s) dB_1(s)$$

$$\begin{aligned}
& - \int_0^t \beta \sinh(t-s) dB_2(s), \\
X_2(t) &= X_1(0) \sinh(t) + X_2(0) \cosh(t) - \int_0^t \alpha \sinh(t-s) dB_1(s) \\
& + \beta \int_0^t \cosh(t-s) dB_2(s). \\
\text{where } \cosh \xi &= \frac{1}{2}(e^\xi + e^{-\xi}), \sinh \xi = \frac{1}{2}(e^\xi - e^{-\xi}).
\end{aligned}$$

5.11.

Hint: To prove that $\lim_{t \rightarrow 1^-} (1-t) \int_0^t \frac{dB_s}{1-s} = 0$ a.s., put $M_t = \int_0^t \frac{dB_s}{1-s}$ for $0 \leq t < 1$ and apply the martingale inequality to prove that

$$P[\sup\{(1-t)|M_t|; t \in [1-2^{-n}, 1-2^{-n-1}]\} > \epsilon] \leq 2\epsilon^{-2} \cdot 2^{-n}.$$

Hence by the Borel-Cantelli lemma we obtain that for a.a. ω there exists $n(\omega) < \infty$ such that

$$n \geq n(\omega) \Rightarrow \omega \notin A_n,$$

where

$$A_n = \{\omega; \sup\{(1-t)|M_t|; t \in [1-2^{-n}, 1-2^{-n-1}]\} > 2^{-\frac{n}{4}}\}.$$

7.1.

- a) $Af(x) = \mu xf'(x) + \frac{1}{2}\sigma^2 f''(x); \quad f \in C_0^2(\mathbf{R}).$
- b) $Af(x) = rxf'(x) + \frac{1}{2}\alpha^2 x^2 f''(x); \quad f \in C_0^2(\mathbf{R}).$
- c) $Af(y) = rf'(y) + \frac{1}{2}\alpha^2 y^2 f''(y); \quad f \in C_0^2(\mathbf{R}).$
- d) $Af(t, x) = \frac{\partial f}{\partial t} + \mu x \frac{\partial f}{\partial x} + \frac{1}{2}\sigma^2 \frac{\partial^2 f}{\partial x^2}; \quad f \in C_0^2(\mathbf{R}^2).$
- e) $Af(x_1, x_2) = \frac{\partial f}{\partial x_1} + x_2 \frac{\partial f}{\partial x_2} + \frac{1}{2} e^{2x_1} \frac{\partial^2 f}{\partial x_2^2}; \quad f \in C_0^2(\mathbf{R}^2).$
- f) $Af(x_1, x_2) = \frac{\partial f}{\partial x_1} + \frac{1}{2} \frac{\partial^2 f}{\partial x_1^2} + \frac{1}{2} x_1^2 \frac{\partial^2 f}{\partial x_2^2}; \quad f \in C_0^2(\mathbf{R}^2).$
- g) $Af(x_1, \dots, x_n) = \sum_{k=1}^n r_k x_k \frac{\partial f}{\partial x_k} + \frac{1}{2} \sum_{i,j=1}^n x_i x_j \left(\sum_{k=1}^n \alpha_{ik} \alpha_{jk} \right) \frac{\partial^2 f}{\partial x_i \partial x_j};$
 $f \in C_0^2(\mathbf{R}^n).$

7.2.

- a) $dX_t = dt + \sqrt{2} dB_t.$
- b) $dX(t) = \begin{bmatrix} dX_1(t) \\ dX_2(t) \end{bmatrix} = \begin{bmatrix} 1 \\ cX_2(t) \end{bmatrix} dt + \begin{bmatrix} 0 \\ \alpha X_2(t) \end{bmatrix} dB_t.$

$$c) dX(t) = \begin{bmatrix} dX_1(t) \\ dX_2(t) \end{bmatrix} = \begin{bmatrix} 2X_2(t) \\ \ln(1 + X_1^2(t) + X_2^2(t)) \end{bmatrix} dt + \begin{bmatrix} X_1(t) & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} dB_1(t) \\ dB_2(t) \end{bmatrix}.$$

(Several other diffusion coefficients are possible.)

7.4.

a), b). Let $\tau_k = \inf\{t > 0; B_t^x = 0 \text{ or } B_t^x = k\}$; $k > x > 0$ and put

$$\rho_k = P^x[B_{\tau_k} = k].$$

Then by Dynkin's formula applied to $f(y) = y^2$ for $0 \leq y \leq k$ we get

$$E^x[\tau_k] = k^2 p_k - x^2. \quad (S1)$$

On the other hand, Dynkin's formula applied to $f(y) = y$ for $0 \leq y \leq k$ gives

$$kp_k = x. \quad (S2)$$

Combining these two identities we get that

$$E^x[\tau] = \lim_{k \rightarrow \infty} E^x[\tau_k] = \lim_{k \rightarrow \infty} x(k - x) = \infty. \quad (S3)$$

Moreover, from (S2) we get

$$P^x[\exists t < \infty \text{ with } B_t = 0] = \lim_{k \rightarrow \infty} P^x[B_{\tau_k} = 0] = \lim_{k \rightarrow \infty} (1 - p_k) = 1, \quad (S4)$$

so $\tau < \infty$ a.s. P^x .

c) Hence if we choose $f(x) = x$ and $X_t = B_t$ and τ as above we get

$$E^x[f(X_\tau)] = E^x[B_\tau] = 0$$

while

$$f(x) + E^x\left[\int_0^\tau \frac{1}{2} \Delta f(B_t) dt\right] = f(x) = x,$$

so (7.23) does not hold in this case.

8.1.

a) $g(t, x) = E^x[\phi(B_t)]$

b) $u(x) = E^x\left[\int_0^\infty e^{-\alpha t} \psi(B_t) dt\right]$

9.1.

a) $dX_t = \begin{bmatrix} \alpha \\ 0 \end{bmatrix} dt + \begin{bmatrix} 0 \\ \beta \end{bmatrix} dB_t$

b) $dX_t = \begin{bmatrix} a \\ b \end{bmatrix} dt + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} dB_t$

c) $dX_t = \alpha X_t dt + \beta dB_t$

d) $dX_t = \alpha dt + \beta X_t dB_t$

e) $dX_t = \begin{bmatrix} dX_1(t) \\ dX_2(t) \end{bmatrix} = \begin{bmatrix} \ln(1 + X_1^2(t)) \\ X_2(t) \end{bmatrix} dt + \sqrt{2} \begin{bmatrix} X_2(t) & 0 \\ X_1(t) & X_1(t) \end{bmatrix} \begin{bmatrix} dB_1(t) \\ dB_2(t) \end{bmatrix}$

9.3.

a) $u(t, x) = E^x[\phi(B_{T-t})]$.

b) $u(t, x) = E^x[\psi(B_t)]$.

9.8.

- a) Let $X_t \in \mathbf{R}^2$ be uniform motion to the right, as described in Example 9.7. Then each one-point set $\{(x_1, x_2)\}$ is thin (and hence semipolar) but not polar.
- b) With X_t as in a) let $H_k = \{(a_k, 1)\}; k = 1, 2, \dots$ where $\{a_k\}_{k=1}^\infty$ is the set of rational numbers. Then each H_k is thin but $Q^{(x_1, 1)}[T_H = 0] = 1$ for all $x_1 \in \mathbf{R}$.

9.10.

Let $Y_t = Y_t^{s,x} = (s + t, X_t^x)$ for $t \geq 0$, where $X_t = X_t^x$ satisfies

$$dX_t = \alpha X_t dt + \beta X_t dB_t; \quad t \geq 0, \quad X_0 = x > 0.$$

Then the generator \hat{A} of Y_t is given by

$$\hat{A}f(s, x) = \frac{\partial f}{\partial s} + \alpha x \frac{\partial f}{\partial x} + \frac{1}{2} \beta^2 x^2 \frac{\partial^2 f}{\partial x^2}; \quad f \in C_0^2(\mathbf{R}^2).$$

Moreover, with $D = \{(t, x); x > 0 \text{ and } t < T\}$ we have

$$\tau_D := \inf\{t > 0; Y_t \notin D\} = \inf\{t > 0; s + t < T\} = T - s.$$

Hence

$$Y_{\tau_D} = (T, X_{T-s}).$$

Therefore, by Theorem 9.18 the solution is

$$f(s, x) = E[e^{-\rho T} \phi(X_{T-s}^x) - \int_0^{T-s} e^{-\rho(s+t)} K(X_t^x) dt].$$

10.1.

- a) $g^*(x) = \infty$, τ^* does not exist.
- b) $g^*(x) = \infty$, τ^* does not exist.
- c) $g^*(x) = 1$, $\tau^* = \inf\{t > 0; B_t = 0\}$.
- d) If $\rho < \frac{1}{2}$ then $g^*(s, x) = \infty$ and τ^* does not exist.

If $\rho \geq \frac{1}{2}$ then $g^*(s, x) = g(s, x) = e^{-\rho s} \cosh x$ and $\tau^* = 0$.

10.3.

$x_0 > 0$ is given implicitly by the equation

$$x_0 = \sqrt{\frac{2}{\rho}} \cdot \frac{e^{2\sqrt{2\rho}x_0} + 1}{e^{2\sqrt{2\rho}x_0} - 1},$$

and $g^*(s, x) = e^{-\rho s} x_0^2 \frac{\cosh(\sqrt{2\rho}x)}{\cosh(\sqrt{2\rho}x_0)}$ for $-x_0 \leq x \leq x_0$, where $\cosh \xi = \frac{1}{2}(e^\xi + e^{-\xi})$.

10.9.

If $0 < \rho \leq 1$ then $\gamma(x) = \frac{1}{\rho}x^2$ but τ^* does not exist. If $\rho > 1$ then

$$\gamma(x) = (1 - \frac{1}{\rho})g^*(0, x) + \frac{1}{\rho}x^2,$$

with $g^*(s, x)$ as in 10.3 and continuation region $D = (-x_0, x_0)$ with x_0 as in (10.3).

10.12.

If $\rho > r$ then $g^*(s, x) = e^{-\rho s}(x_0 - 1)^+(\frac{x}{x_0})^\gamma$ and $\tau^* = \inf\{t > 0; X_t \geq x_0\}$, where

$$\gamma = \alpha^{-2}[\frac{1}{2}\alpha^2 - r + \sqrt{(\frac{1}{2}\alpha^2 - r)^2 + 2\alpha^2\rho}]$$

and

$$x_0 = \frac{\gamma}{\gamma - 1} \quad (\gamma > 1 \Leftrightarrow \rho > r).$$

10.13.

If $\alpha \leq \rho$ then $\tau^* = 0$.

If $\rho < \alpha < \rho + \lambda$ then

$$G^*(s, p, q) = \begin{cases} e^{-\rho s} pq & ; \\ C_1(pq)^{\gamma_1} + \frac{\lambda}{\rho + \lambda - \alpha} \cdot pq - \frac{K}{\rho} & ; \end{cases} \begin{array}{l} \text{if } 0 < pq < y_0 \\ \text{if } pq \geq y_0 \end{array}$$

where

$$\begin{aligned}\gamma_1 &= \beta^{-2} \left[\frac{1}{2} \beta^2 + \lambda - \alpha - \sqrt{\left(\frac{1}{2} \beta^2 + \lambda - \alpha \right)^2 + 2\rho\beta^2} \right] < 0, \\ y_0 &= \frac{(-\gamma_1)K(\rho + \lambda - \alpha)}{(1 - \gamma_1)\rho(\alpha - \rho)} > 0\end{aligned}$$

and

$$C_1 = \frac{(\alpha - \rho)y_0^{1-\gamma_1}}{(-\gamma_1)(\rho + \lambda - \alpha)}.$$

The continuation region is

$$D = \{(s, p, q); pq > y_0\}.$$

If $\rho + \lambda \leq \alpha$ then $G^* = \infty$.

11.6.

$$u^* = \frac{2(a_1 - a_2) - \sigma_2^2}{\sigma_1^2 - \sigma_2^2} \text{ (constant),}$$

$$\Phi(s, x) = e^{\gamma(t-t_1)} \sqrt{x} \text{ for } t < t_1, x > 0$$

where

$$\gamma = \frac{1}{2}[a_1 u^* + a_2(1 - u^*)] - \frac{1}{2}[\sigma_1^2(u^*)^2 + \sigma_2^2(1 - u^*)^2].$$

11.11. Additional hints:

For the solution of the unconstrained problem try a function $\phi_\lambda(s, x)$ of the form

$$\phi_\lambda(s, x) = a_\lambda(s)x^2 + b_\lambda(s),$$

for suitable functions $a_\lambda(s), b_\lambda(s)$ with $\lambda \in \mathbf{R}$ fixed. By substituting this into the HJB equation we arrive at the equations

$$\begin{aligned}a'_\lambda(s) &= \frac{1}{\theta} a_\lambda^2(s) - 1 \quad \text{for } s < t_1 \\ a_\lambda(t_1) &= \lambda\end{aligned}$$

and

$$\begin{aligned}b'_\lambda(s) &= -\sigma^2 a_\lambda(s) \quad \text{for } s < t_1 \\ b_\lambda(t_1) &= -\lambda m,\end{aligned}$$

with optimal control $u^*(s, x) = -\frac{1}{\theta} a_\lambda(s)x$.

Now substitute this into the equation for $X_t^{u^*}$ and use the terminal condition to determine λ_0 .

Bibliography

K. K. AASE (1982):

Stochastic continuous-time model reference adaptive systems with decreasing gain. *Advances in Appl. Prop.* **14**, 763–788.

K. K. AASE (1984):

Optimum portfolio diversification in a general continuous time model. *Stoch. Proc. and their Applications* **18**, 81–98.

R. J. ADLER (1981):

The Geometry of Random Fields. Wiley & Sons.

E. S. ANDERSEN and B. JESSEN (1948):

Some limit theorems on set-functions. *Danske Vid. Selsk. Mat.-Fys. Medd.* **25**, #5, 1–8.

L. ARNOLD (1973):

Stochastische Differentialgleichungen. Theorie und Anwendung. Oldenbourg Verlag.

J. A. BATHER (1970):

Optimal stopping problems for Brownian motion. *Advances in Appl. Prob.* **2**, 259–286.

V. E. BENEŠ (1974):

Girsanov functionals and optimal bang-bang laws for final-value stochastic control. *Stoch. Proc. and Their Appl.* **2**, 127–140.

A. BENOUSSAN and J. L. LIONS (1978):

Applications des inéquations variationnelles en contrôle stochastique. Dunod. (Applications of Variational Inequalities in Stochastic Control. North-Holland.)

A. BERNARD, E. A. CAMPBELL and A.M. DAVIE (1979):

Brownian motion and generalized analytic and inner functions. *Ann. Inst. Fourier* **729**, 207–228.

- L. BERS, F. JOHN and M. SCHECHTER (1964):
Partial Differential Equations. Interscience.
- R. M. BLUMENTHAL and R. K. GETOOR (1968):
Markov Processes and Potential Theory. Academic Press.
- L. BREIMAN (1968):
Probability. Addison-Wesley.
- K. A. BREKKE and B. ØKSENDAL (1991):
The high contact principle as a sufficiency condition for optimal stopping. To appear in D. Lund and B. Øksendal (editors): Stochastic Models and Option Values. North-Holland.
- B. M. BROWN and J. I. HEWITT (1975):
Asymptotic likelihood theory for diffusion processes. *J. Appl. Prob.* **12**, 228–238.
- R. S. BUCY and P. D. JOSEPH (1968):
Filtering for Stochastic Processes with Applications to Guidance. Interscience.
- Y. S. CHOW, H. ROBBINS, and D. SIEGMUND (1971):
Great Expectations: The Theory of Optimal Stopping. Houghton Mifflin Co.
- K. L. CHUNG (1982):
Lectures from Markov Processes to Brownian Motion. Springer-Verlag.
- K. L. CHUNG and R. WILLIAMS (1990):
Introduction to Stochastic Integration. Second Edition. Birkhäuser.
- J. M. CLARK (1970, 1971):
The representation of functionals of Brownian motion by stochastic integrals. *Ann. Math. Stat.* **41**, 1282–1291 and **42**, 1778.
- L. CSINK and B. ØKSENDAL (1983):
Stochastic harmonic morphisms: Functions mapping the paths of one diffusion into the paths of another. *Ann. Inst. Fourier* **33**, 219–240.
- L. CSINK, P. FITZSIMMONS and B. ØKSENDAL (1990):
A stochastic characterization of harmonic morphisms. *Math. Ann.* **287**, 1–18.
- M. H. A. DAVIS (1977):
Linear Estimation and Stochastic Control. Chapman and Hall.
- M. H. A. DAVIS (1984):
Lectures on Stochastic Control and Nonlinear Filtering. Tata Institute of Fundamental Research 75. Springer-Verlag.

M. H. A. DAVIS (1993):

Markov Models and Optimization. Chapman & Hall, London.

M. H. A. DAVIS and R. B. VINTER (1985):

Stochastic Modelling and Control. Chapman and Hall.

A. K. DIXIT and R. S. PINDYCK (1994):

Investment under Uncertainty. Princeton University Press.

J. L. DOOB (1984):

Classical Potential Theory and Its Probabilistic Counterpart. Springer-Verlag.

D. DUFFIE (1992):

Dynamic Asset Pricing Theory. Princeton University Press.

D. DUFFIE (1994):

Martingales, arbitrage, and portfolio choice. First European Congress of Mathematics, Vol. II, Birkhäuser, pp. 3–21.

R. DURRETT (1984):

Brownian Motion and Martingales in Analysis. Wadsworth Inc.

E. B. DYNKIN (1963):

The optimum choice of the instant for stopping a Markov process. Soviet Mathematics **4**, 627–629.

E. B. DYNKIN (1965 I):

Markov Processes, vol. I. Springer-Verlag.

E. B. DYNKIN (1965 II):

Markov Processes, vol. II. Springer-Verlag.

E. B. DYNKIN and A. A. YUSHKEVICH (1979):

Controlled Markov Processes. Springer-Verlag.

R. J. ELLIOTT (1982):

Stochastic Calculus and Applications. Springer-Verlag.

K. D. ELWORTHY (1982):

Stochastic Differential Equations on manifolds. Cambridge University Press.

M. EMERY (1989):

Stochastic Calculus in Manifolds. Springer-Verlag.

W. H. FLEMING and R. W. RISHEL (1975):

Deterministic and Stochastic Optimal Control. Springer-Verlag.

- W. H. FLEMING and H. M. SONER (1993):
 Controlled Markov Processes and Viscosity Solutions. Springer-Verlag.
- G. B. FOLLAND (1984):
 Real Analysis. J. Wiley & Sons.
- M. FREIDLIN (1985):
 Functional Integration and Partial Differential Equations. Princeton University Press.
- A. FRIEDMAN (1975):
 Stochastic Differential Equations and Applications, vol. I. Academic Press.
- A. FRIEDMAN (1976):
 Stochastic Differential Equations and Applications, vol. II. Academic Press.
- M. FUKUSHIMA (1980):
 Dirichlet Forms and Markov Processes. North-Holland/Kodansha.
- T. C. GARD (1988):
 Introduction to Stochastic Differential Equations. Dekker.
- A. GELB (1974):
 Applied Optimal Estimation. MIT.
- I. I. GIHMAN and A. V. SKOROHOD (1974a):
 Stochastic Differential Equations. Springer-Verlag.
- I. I. GIHMAN and A. V. SKOROHOD (1974b):
 The Theory of Stochastic Processes, vol. I. Springer-Verlag.
- I. I. GIHMAN and A. V. SKOROHOD (1975):
 The Theory of Stochastic Processes, vol. II. Springer-Verlag.
- I. I. GIHMAN and A. V. SKOROHOD (1979):
 Controlled Stochastic Processes. Springer-Verlag.
- J. GRUE (1989):
 Wave drift damping of the motions of moored platforms by the method of stochastic differential equations. Manuscript, University of Oslo.
- J. M. HARRISON and D. KREPS (1979):
 Martingales and arbitrage in multiperiod securities markets. *J. Economic Theory* **20**, 381–408.
- J. M. HARRISON and S. PLISKA (1981):
 Martingales and stochastic integrals in the theory of continuous trading. *Stoch. Proc. and Their Applications* **11**, 215–260.

T. HIDA (1980):

Brownian Motion. Springer-Verlag.

T. HIDA, H.-H. KUO, J. POTTHOFF and L. STREIT (1993):

White Noise. An Infinite Dimensional Approach. Kluwer.

P. G. HOEL, S. C. PORT and C. J. STONE (1972):

Introduction to Stochastic Processes. Waveland Press, Illinois 60070.

K. HOFFMAN (1962):

Banach Spaces of Analytic Functions. Prentice Hall.

N. IKEDA and S. WATANABE (1989):

Stochastic Differential Equations and Diffusion Processes. Second Edition.
North-Holland/Kodansha.

K. ITO (1951):

Multiple Wiener integral. J. Math. Soc. Japan **3**, 157–169.

K. ITO and H. P. McKEAN (1965):

Diffusion Processes and Their Sample Paths. Springer-Verlag.

A. H. JASWINSKI (1970):

Stochastic Processes and Filtering Theory. Academic Press.

G. KALLIANPUR (1980):

Stochastic Filtering Theory. Springer-Verlag.

I. KARATZAS, J. LEHOCZKY and S. SHREVE (1987):

Optimal portfolio and consumption decisions for a 'Small Investor' on a finite horizon. SIAM J. Control and Optimization **25**, 1157–1186.

I. KARATZAS and S. E. SHREVE (1988):

Brownian Motion and Stochastic Calculus. Springer-Verlag.

S. KARLIN and H. TAYLOR (1975):

A First Course in Stochastic Processes. Second Edition. Academic Press.

P. E. KLOEDEN and E. PLATEN (1992):

Numerical Solution of Stochastic Differential Equations. Springer-Verlag.

F. B. KNIGHT (1981):

Essentials of Brownian Motion. American Math. Soc.

P. KOPP (1984):

Martingales and Stochastic Integrals. Cambridge University Press.

- V. KRISHNAN (1984):
Nonlinear Filtering and Smoothing: An Introduction to Martingales, Stochastic Integrals and Estimation. J. Wiley & Sons.
- N. V. KRYLOV (1980):
Controlled Diffusion Processes. Springer-Verlag.
- N. V. KRYLOV and A. K. ZVONKIN (1981):
On strong solutions of stochastic differential equations. Sel. Math. Sov. **I**, 19–61.
- H. J. KUSHNER (1967):
Stochastic Stability and Control. Academic Press.
- J. LAMPERTI (1977):
Stochastic Processes. Springer-Verlag.
- R. S. LIPTSER and A. N. SHIRYAYEV (1977):
Statistics of Random Processes, vol. I. Springer-Verlag.
- R. S. LIPTSER and A. N. SHIRYAYEV (1978):
Statistics of Random Processes, vol. II. Springer-Verlag.
- R. McDONALD and D. SIEGEL (1986):
The value of waiting to invest. Quarterly J. of Economics **101**, 707–727.
- T. P. McGARTY (1974):
Stochastic Systems and State Estimation. J. Wiley & Sons.
- H. P. McKEAN (1965):
A free boundary problem for the heat equation arising from a problem of mathematical economics. Industrial managem. review **60**, 32–39.
- H. P. McKEAN (1969):
Stochastic Integrals. Academic Press.
- A. G. MALLIARIS (1983):
Ito's calculus in financial decision making. SIAM Review **25**, 481–496.
- A. G. MALLIARIS and W. A. BROCK (1982):
Stochastic Methods in Economics and Finance. North-Holland.
- H. M. MARKOWITZ (1976):
Portfolio Selection. Efficient Diversification of Investments. Yale University Press.
- P. S. MAYBECK (1979):
Stochastic Models, Estimation, and Control. Vols. 1–3. Academic Press.

R. C. MERTON (1971):

Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory* **3**, 373–413.

R. C. MERTON (1990):

Continuous-Time Finance. Blackwell Publishers.

M. METIVIER and J. PELLAUMAIL (1980):

Stochastic Integration. Academic Press.

P. A. MEYER (1966):

Probability and Potentials. Blaisdell.

P. A. MEYER (1976):

Un cours sur les intégrales stochastiques. Sem. de Prob. X. Lecture Notes in Mathematics, vol. 511. Springer-Verlag, pp. 245–400.

D. OCONE (1984):

Malliavin's calculus and stochastic integral: representation of functionals of diffusion processes. *Stochastics* **12**, 161–185.

B. ØKSENDAL (1984):

Finely harmonic morphisms, Brownian path preserving functions and conformal martingales. *Inventiones math.* **750**, 179–187.

B. ØKSENDAL (1990):

When is a stochastic integral a time change of a diffusion? *Journal of Theoretical Probability* **3**, 207–226.

T. E. OLSEN and G. STENSLAND (1987):

A note on the value of waiting to invest. Manuscript CMI, N–5036 Fantoft, Norway.

E. PARDOUX (1979):

Stochastic partial differential equations and filtering of diffusion processes. *Stochastics* **3**, 127–167.

S. PORT and C. STONE (1979):

Brownian Motion and Classical Potential Theory. Academic Press.

P. PROTTER (1990):

Stochastic Integration and Differential Equations. Springer-Verlag.

F. P. RAMSEY (1928):

A mathematical theory of saving. *Economic J.* **38**, 543–549.

M. RAO (1977):

Brownian Motion and Classical Potential Theory. Aarhus Univ. Lecture Notes in Mathematics 47.

M. M. RAO (1984):

Probability Theory with Applications. Academic Press.

D. REVUZ and M. YOR (1991):

Continuous Martingales and Brownian Motion. Springer-Verlag.

L. C. G. ROGERS and D. WILLIAMS (1987):

Diffusions, Markov Processes, and Martingales. Vol. 2. J. Wiley & Sons.

YU. A. ROZANOV (1982):

Markov Random Fields. Springer-Verlag.

P. A. SAMUELSON (1965):

Rational theory of warrant pricing. Industrial managem. review **6**, 13–32.

A. N. SHIRYAYEV (1978):

Optimal Stopping Rules. Springer-Verlag.

B. SIMON (1979):

Functional Integration and Quantum Physics. Academic Press.

J. L. SNELL (1952):

Applications of martingale system theorems. Trans. Amer. Math. Soc. **73**, 293–312.

R. L. STRATONOVICH (1966):

A new representation for stochastic integrals and equations. J. Siam Control **4**, 362–371.

D. W. STROOCK (1971):

On the growth of stochastic integrals. Z. Wahr. verw. Geb. **18**, 340–344.

D. W. STROOCK (1981):

Topics in Stochastic Differential Equations. Tata Institute of Fundamental Research. Springer-Verlag.

D. W. STROOCK (1993):

Probability Theory, An Analytic View. Cambridge University Press.

D. W. STROOCK and S. R. S. VARADHAN (1979):

Multidimensional Diffusion Processes. Springer-Verlag.

H. J. SUSSMANN (1978):

On the gap between deterministic and stochastic ordinary differential equations. *The Annals of Prob.* **60**, 19–41.

A. TARASKIN (1974):

On the asymptotic normality of vectorvalued stochastic integrals and estimates of drift parameters of a multidimensional diffusion process. *Theory Prob. Math. Statist.* **2**, 209–224.

THE OPEN UNIVERSITY (1981):

Mathematical models and methods, unit 11. The Open University Press.

F. TOPSØE (1978):

An information theoretical game in connection with the maximum entropy principle (Danish). *Nordisk Matematisk Tidsskrift* **25/26**, 157–172.

M. TURELLI (1977):

Random environments and stochastic calculus. *Theor. Pop. Biology* **12**, 140–178.

J. UBØE (1987):

Conformal martingales and analytic functions. *Math. Scand.* **60**, 292–309.

P. VAN MOERBEKE (1974):

An optimal stopping problem with linear reward. *Acta Mathematica* **132**, 111–151.

D. WILLIAMS (1979):

Diffusions, Markov Processes and Martingales. J. Wiley & Sons.

D. WILLIAMS (1981) (editor):

Stochastic Integrals. Lecture Notes in Mathematics, vol. 851. Springer-Verlag.

D. WILLIAMS (1991):

Probability with Martingales. Cambridge University Press.

E. WONG (1971):

Stochastic Processes in Information and Dynamical Systems. McGraw-Hill.

E. WONG and M. ZAKAI (1969):

Riemann-Stieltjes approximations of stochastic integrals. *Z. Wahr. verw. Geb.* **120**, 87–97.

M. YOR (1992):

Some Aspects of Brownian Motion, Part 1: Some Special Functionals. Birkhäuser.

List of Frequently Used Notation and Symbols

\mathbf{R}^n	n-dimensional Euclidean space
\mathbf{R}^+	the non-negative real numbers
\mathbf{Q}	the rational numbers
\mathbf{Z}	the integers
$\mathbf{Z}^+ = \mathbf{N}$	the natural numbers
\mathbf{C}	the complex plane
$\mathbf{R}^{n \times m}$	the $n \times m$ matrices (real entries)
$ C $	the determinant of the $n \times n$ matrix C
$\mathbf{R}^n \simeq \mathbf{R}^{n \times 1}$	i.e. vectors in \mathbf{R}^n are regarded as $n \times 1$ -matrices.
$\mathbf{C}^n = \mathbf{C} \times \cdots \times \mathbf{C}$	the n-dimensional complex space
$ x $	$(\sum_{i=1}^n x_i^2)^{\frac{1}{2}}$ if $x \in \mathbf{R}^n$
$C(U, V)$	the continuous functions from U into V
$C(U)$	the same as $C(U, \mathbf{R})$
$C_0(U)$	the functions in $C(U)$ with compact support
$C^k = C^k(U)$	the functions in $C(U, \mathbf{R})$ with continuous derivatives up to order k .
$C_0^k = C_0^k(U)$	the functions in $C^k(U)$ with compact support in U .

$C^{k+\alpha}$	the functions in C^k whose $k'th$ derivatives are Lipschitz continuous with exponent α
$C^{1,2}(\mathbf{R} \times \mathbf{R}^n)$	the functions $f(t, x) : \mathbf{R} \times \mathbf{R}^n \rightarrow \mathbf{R}$ which are C^1 w.r.t. $t \in \mathbf{R}$ and C^2 w.r.t. $x \in \mathbf{R}^n$
$C_b(U)$	the bounded continuous functions on U
$f K$	the restriction of the function f to the set K
$A = A_X$	the generator of an Ito diffusion X
$\mathcal{A} = \mathcal{A}_X$	the characteristic operator of an Ito diffusion X
$L = L_X$	the second order partial differential operator which coincides with A_X on C_0^2 and with \mathcal{A}_X on C^2
B_t (or $(B_t, \mathcal{F}, \Omega, P^x)$)	Brownian motion
\mathcal{D}_A	the domain of definition of the operator A
∇	the gradient: $\nabla f = (\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n})$
Δ	the Laplace operator: $\Delta f = \sum_i \frac{\partial^2 f}{\partial x_i^2}$
L	a semielliptic second order partial differential operator of the form $L = \sum_i b_i \frac{\partial}{\partial x_i} + \sum_{i,j} a_{ij} \frac{\partial^2}{\partial x_i \partial x_j}$
R_α	the resolvent operator
iff	if and only if
a.a., a.e., a.s.	almost all, almost everywhere, almost surely
w.r.t.	with respect to
s.t.	such that
\sim	coincides in law with
$E[Y] = E^\mu[Y] = \int Y d\mu$	the expectation of the random variable Y w.r.t. the measure μ

$E[Y \mathcal{N}]$	the conditional expectation of Y w.r.t. \mathcal{N}
\mathcal{F}_∞	the σ -algebra generated by $\bigcup_{t>0} \mathcal{F}_t$
\mathcal{B}	the Borel σ -algebra
\mathcal{F}_t	the σ -algebra generated by $\{B_s; s \leq t\}$
\mathcal{F}_τ	the σ -algebra generated by $\{B_{s \wedge \tau}; s \geq 0\}$ (τ is a stopping time)
\perp	orthogonal to (in a Hilbert space)
\mathcal{M}_t	the σ -algebra generated by $\{X_s; s \leq t\}$ (X_t is an Ito diffusion)
\mathcal{M}_τ	the σ -algebra generated by $\{X_{s \wedge \tau}; s \geq 0\}$ (τ is a stopping time)
∂G	the boundary of the set G
\overline{G}	the closure of the set G
$G \subset\subset H$	\overline{G} is compact and $\overline{G} \subset H$
$d(y, K)$	the distance from the point $y \in \mathbf{R}^n$ to the set $K \subset \mathbf{R}^n$
τ_G	the first exit time from the set G of a process $X_t : \tau_G = \inf\{t > 0; X_t \notin G\}$
$\mathcal{V}(S, T)$	Definition 3.4
\mathcal{W}	Definition 3.15
(H)	Hunt's condition (Chapter IX)
HJB	the Hamilton-Jacobi-Bellman equation (Chapter XI)
I_n	the $n \times n$ identity matrix
χ_G	the indicator function of the set G $\chi_G(x) = 1$ if $x \in G$, $\chi_G(x) = 0$ if $x \notin G$
P^x	the probability law of B_t starting at x

$P = P^0$

Q^x the probability law of X_t starting at x ($X_0 = x$)

$R^{(s,x)}$ the probability law of $Y_t = (s + t), X_t^x)_{t \leq 0}$
with $Y_0 = (s, x)$ (Chapter X)

$Q^{s,x}$ the probability law of $Y_t = (s + t, X_{s+t}^{s,x})_{t \geq 0}$
with $Y_0 = (s, x)$ (Chapter XI)

$E^x, E^{(s,x)}, E^{s,x}$ the expectation operator w.r.t. the measures Q^x ,
 $R^{(s,x)}$ and $Q^{s,x}$, respectively

E the expectation w.r.t. a measure which is clear
from the context (usually P^0)

$s \wedge t$ the minimum of s and t ($= \min(s, t)$)

$s \vee t$ the maximum of s and t ($= \max(s, t)$)

σ^T the transposed of the matrix σ

δ_x the unit point mass at x

δ_{ij} $\delta_{ij} = 1$ if $i = j$, $\delta_{ij} = 0$ if $i \neq j$

θ_t the shift operator:
 $\theta_t(f(X_s)) = f(X_{t+s})$ (Chapter VII)

$:=$ equal to by definition

$\underline{\lim}, \overline{\lim}$ the same as \liminf, \limsup

\square end of proof

“increasing” is used with the same meaning as “nondecreasing”, “decreasing” with the same meaning as “nonincreasing”. In the strict cases “strictly increasing/strictly decreasing” are used.

Index

adapted process	22
analytic functions (and Brownian motion)	143
Bellman principle	231
Bessel process	46, 133
bequest function	213
Borel sets, Borel σ -algebra	6
Borel-Cantelli lemma	14
Brownian motion, in \mathbf{R}^n	3, 9–12
Brownian motion, complex	74
Brownian motion, on the unit circle	63, 117
Brownian motion, on the ellipse	70
Brownian motion, on the unit sphere	142
Brownian motion, on a Riemannian manifold	143
Brownian motion, the graph of	114
Brownian motion, w.r.t. an increasing family \mathcal{H}_t of σ -algebras	68
Brownian bridge	72
capacity	157
characteristic function	237
characteristic operator	116
change of time	138
change of variable in an Ito integral	141
Chebychev's inequality	13
coincide in law	134
complete probability space	6
complex Brownian motion	74
conditional expectation	239

conditioned Brownian motion	123
continuation region	189
continuous in mean square	37
control, deterministic (open loop)	214
control, feedback (closed loop)	214
control, Markov	214
control, optimal	214
cost function	220
covariance matrix	236
cross-variation processes	145
density (of a random variable)	13
diffusion, Ito	103
diffusion, Dynkin	117
diffusion coefficient	103
Dirichlet problem	2, 160
Dirichlet problem (stochastic version)	166
distribution (of a random variable)	7
distribution (of a process)	8
distribution function (of a random variable)	13
Dobb-Dynkin lemma	7
drift coefficient	103
Dynkin's formula	114
eigenvalues (of the Laplacian)	178
elementary function/process	23
elliptic partial differential operator	160
estimate (linear/measurable)	81
estimation of a parameter	93
estimation, exact asymptotic	98
events	6
excessive function	187
expectation	7
explosion (of a diffusion)	64
exponential martingale	52
Feller-continuity	126

- Feynman-Kac formula 128
filtering problem, general 2, 75–77
filtering problem, linear 77–97
filtration 27, 34
finite-dimensional distributions (of a stochastic process) 8
first exit time 107, 110
- Gaussian process 10, 64
generalized (distribution valued) process 18
generator (of an Ito diffusion) 111
geometric Brownian motion 60
Girsanov's theorem 57, 146–151
Girsanov transformation 147
Green measure 17, 175
Green formula 176
Green function 157
Green function (classical) 176
Green operator 157
- Hamilton-Jacobi-Bellman (HJB) equation 214–218
harmonic extension (w.r.t. an Ito diffusion) 118
harmonic function (and Brownian motion) 143
harmonic function (w.r.t. a diffusion) 165
harmonic measure (of a diffusion) 110, 111
harmonic measure (of Brownian motion) 120
Hermite polynomials 35
high contact (smooth fit) principle 209
hitting distribution 110, 111
h-transform (of Brownian motion) 122
Hunt's condition (H) 167
- independent 7
independent increments 11, 19
innovation process 78, 82, 83, 86
integration by parts (stochastic) 43, 52
interpolation (smoothing) 100
irregular point 162

iterated Ito integrals.....	35
iterated logarithm (law of).....	62
Ito diffusion	103, 104
Ito integral.....	21–34
Ito integral; multidimensional.....	31
Ito interpretation (of a stochastic differential equation).....	32, 58, 75
Ito isometry	23, 26
Ito process	41
Ito representation theorem.....	48
Ito's formula.....	41, 45
Kalman-Bucy filter.....	2, 91, 97
Kazamaki condition	52
kernel function	122
killing (a diffusion).....	130
killing rate	131, 158
Kolmogorov's backward equation	124
Kolmogorov's forward equation	152
Kolmogorov's continuity theorem	12
Kolmogorov's extension theorem	9
Langevin equation.....	71
Laplace operator Δ	2, 54
Laplace-Beltrami operator	143
least superharmonic majorant	187
least supermeanvalued majorant	186
Levy's theorem	145
L-harmonic extension	160
linear regulator problem	220
local time.....	55, 138
Lyapunov equation	99
Markov control	214
Markov process	106
Markov property	105
martingale	27, 30, 242
martingale convergence theorem	242

martingale inequality	28
martingale problem.....	131
martingale representation theorem	46, 50
maximum likelihood	94
maximum principle.....	180
mean-reverting Ornstein-Uhlenbeck process	71
mean value property (for a diffusion).....	110, 111
mean value property, classical	120
measurable sets (w.r.t. a σ -algebra)	6
measurable space.....	5
moving average, exponentially weighted.....	76
noise	1–3, 18–19, 59
normal distribution.....	10, 236
Novikov condition	52
optimal control.....	214
optimal performance	214
optimal portfolio selection	223
optimal stopping	3, 183–206
optimal stopping existence theorem.....	189
optimal stopping uniqueness theorem	193
Ornstein-Uhlenbeck equation/process	71
orthogonal increments	78
path (of a stochastic process)	8
performance criterion.....	213
Perron-Wiener-Brelot solution	171
Poisson formula	181
Poisson kernel	181
Poisson problem	172
Poisson problem (stochastic version).....	172
polar set	167
population growth.....	1, 59, 94
prediction	99
probability measure.....	5
probability space	6

p th variation process.....	17
quadratic variation process.....	17, 54
random time change.....	138
random variable.....	5, 7
recurrent.....	116
regular point	162
resolvent operator.....	126
reward function.....	183, 195
reward rate function.....	184
Riccati equation	91, 97, 222
semi-elliptic partial differential operator	160
semi-polar set	167
separation principle	222
shift operator	109
σ -algebra	5, 6
smoothing (interpretation).....	100
stationary process.....	16, 18, 19
stochastic control.....	3, 212–229
stochastic differential equation; definition	59
stochastic differential equation; existence and uniqueness of solution	64
stochastic differential equation; weak and strong solution.....	68
stochastic Dirichlet problem.....	166
stochastic integral	41
stochastic Poisson problem	172
stochastic process.....	8
stopping time	54, 106, 107
Stratonovich integral.....	21, 33–34, 36
Stratonovich interpretation (of a stochastic differential equation) ..	33, 60, 75
strong Feller process.....	169
strong Markov property	107–110
strong solution (of a stochastic differential equation)	68
submartingale	242
superharmonic function.....	184
superharmonic majorant	186

supermartingale	186, 242
supermeanvalued function	184
supermeanvalued majorant	186
support (of a diffusion)	101
Tanaka's equation	155
Tanaka's formula	55
terminal conditions (in stochastic control)	228
thin set	167
time-homogeneous	104
time change formula Ito integrals	141
transient	116
transition measure	176
transition operator	159
trap	116
uniformly elliptic partial differential operator	168
uniformly integrable	241
utility function	3, 213
variational inequalities (and optimal stopping)	202–206
version (of a process)	12, 28
Volterra equation, stochastic	72
Volterra equation, deterministic	85
weak solution (of a stochastic differential equation)	68
weak uniqueness	68
well posed (martingale problem)	133
white noise	18, 59
Wiener criterion	164
X-harmonic	165
zero-one law	161