Probability and Its Applications

Samuel N. Cohen
Robert J. Elliott

Stochastic Calculus and Applications

Second Edition

Birkhäuser

Birkhäuser

Probability and Its Applications

More information about this series at http://www.springer.com/series/4893

Samuel N. Cohen • Robert J. Elliott

Stochastic Calculus and Applications

Second Edition

Birkhäuser

Samuel N. Cohen
Mathematical Institute
University of Oxford
Oxford, UK

Robert J. Elliott
School of Mathematics
University of Adelaide
Adelaide, Australia

Haskayne School of Business
University of Calgary
Calgary, Canada

ISSN 2297-0371 ISSN 2297-0398 (electronic)
Probability and Its Applications
ISBN 978-1-4939-2866-8 ISBN 978-1-4939-2867-5 (eBook)
DOI 10.1007/978-1-4939-2867-5

Library of Congress Control Number: 2015060429

Mathematics Subject Classification (2010): 60-01, 49-01, 93E11, 93E20

Springer New York Heidelberg Dordrecht London

Printed on acid-free paper

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

To Juli and Ann

I returned, and saw vnder the Sunne, That the race is not to the swift, nor the battell to the strong, neither yet bread to the wise, nor yet riches to men of vnderstanding, nor yet fauour to men of skil; but time and chance happeneth to them all.

— Ecclesiastes 9:11 (AV, 1611)

The Queue of Fortune from John Lydgate's *The Siege of Troy*, mid-fifteenth century.

Preface to the Second Edition (2015)

The theory of probability and stochastic calculus has grown significantly since the publication of the first edition of this book. The theory of stochastic integration and semimartingales, a relatively recent development at the time of the first edition, is now a standard and significant part of the working mathematician's toolkit. Concepts such as Backward SDEs, which were unheard of in 1982 (apart from one paper of Bismut), are now understood to be fundamental to the theory of stochastic control and mathematical finance.

Applications of stochastic processes arise particularly in finance and engineering. This book presents a rigorous mathematical framework for these problems in a comprehensive and inclusive way. The general theory of processes was developed in the 1970s by Paul-André Meyer and Claude Dellacherie, but for some years it was little known or appreciated in the 'anglo saxon' world ("sauf que les ingenieurs anglais", as Meyer referred to the remarkable group at Berkeley led by Gene Wong and Pravin Varaiya.). The first edition was an attempt to fill this gap in the English literature.

To describe this volume as a second edition is an understatement. The original volume was 300 pages; this has over 650. Consequently, the book contains a large amount of additional material, including much new material.

The growth in the discipline over the past 30 years has the consequence that it is even less possible now to attempt to give a comprehensive view of the subject. Our aim in preparing this second edition is nevertheless to give a broad overview of the theory, with enough rigour to provide a firm foundation for further developments. We do not pretend that this is the most introductory text to stochastic calculus, as we wish to provide the reader with the full power of the general theory of stochastic processes, rather than restricting attention *ab initio* to the case of Brownian motion or Markov processes.

A difficult consequence of this perspective is that it, therefore, takes some time before we reach the 'action' of stochastic integration theory. However, when we get there, we find that we already have all the desired tools at our disposal.

In writing such a book, one naturally compares with and is informed by other works on the topic, and it is difficult to know how to cite such works. These have included, in no particular order, the books of Revuz and Yor [155], Protter [152], Jacod and Shiryaev [110], Jacod [107], Dellacherie and Meyer [54], Dellacherie [53], Karatzas and Shreve [117], Øksendal [142], Rogers and Williams [159], Williams [183], He, Wang and Yan [94], Touzi [177], Ethier and Kurtz [77], Ikeda and Watanabe [98], Stroock and Varadhan [174], Pham [149], Föllmer and Schied [81] and the blog of George Lowther [127].

Numerous people deserve thanks for their support, input and comments on this text. In particular, Steve Clark, whose notes evolved into an early version of Chapter 1. Thanks also to Victor Fedyashov, Michael Monoyios, Gonçalo Simões, Hendrick Brackmann, Gechun Liang, Dmitry Kramkov and Łukasz Szpruch, and to groups in both Oxford and Calgary, who read various sections of the text and made useful comments, and particularly to Johannes Ruf, who read through the first half of the text in an early version. Also thanks are due to three anonymous reviewers, whose attention has resulted in a much improved text. Finally, thanks to Vivian Spak for her assistance in preparing a LaTeXversion of the first edition from which to work.

We now review the content of this edition, emphasizing the new material. Even though there is significant content in common, the names and content of chapters differ significantly from the first edition.

Chapter 1 is new and presents a rigorous treatment of measure theory. Chapter 2 discusses Probabilities and Expectation, Chapter 3 Filtrations, Stopping Times and Stochastic Processes and Chapter 4 Martingales in discrete time. These chapters have all been largely rewritten. The presentation in Chapter 5, Martingales in Continuous Time, is largely new, particularly the section giving examples of martingales. Chapters 6 to 10 contain much new material and are mostly rewritten. They discuss The Classification of Stopping Times, The Progressive, Optional and Predictable σ-Algebras, Processes of Finite Variation and the Doob–Meyer Decomposition and The Structure of Square Integrable Martingales. Chapter 11 on Quadratic Variation and Semimartingales is mostly rewritten and the Burkholder–Davis–Gundy inequality is included. Stochastic integrals are constructed in Chapter 12 and Émery's Semimartingale Topology introduced.

The treatment of random measures in Chapter 13 is clearer and Chapter 14 on the Itô Differential Rule gives a cleaner treatment. Chapter 15 discusses The Exponential Formula and Girsanov's Theorem. There is an extensive presentation of the Novikov and Kazamaki Criteria. The treatment of Lipschitz Stochastic Differential Equations in Chapter 16 is new and their Markov properties presented in a rewritten Chapter 16. Weak solutions of stochastic differential equations are presented in a completely new Chapter 18. As mentioned above, Backward stochastic differential equations were largely unknown in the 1980s but now play a central role in financial modelling and control. They are discussed in a new Chapter 19.

Applications are treated in Chapters 20, 21 and 22. The single jump process is discussed in a rewritten Chapter 20. Chapter 21 is largely new and uses backward stochastic differential equations to discuss the control of diffusions and jump processes. Chapter 22 discusses filtering. The Appendices are new and include topics such as Outer Measure and Carathéodory's Extension Theorem and Kolmogorov's Extension Theorems.

Oxford, UK Samuel Cohen

Adelaide, Australia Robert Elliott
Calgary, Canada

April 2015

Preface to the First Edition (1982)

The object of this book is to take a reader, who has been exposed to only the usual courses in probability, measure theory and stochastic processes, through the modern French general theory of random processes, to the point where it is being applied by systems theorists and electronic engineers. It is surprising and unfortunate that, although this general theory is found so useful by theoretical engineers, it is not (with a few significant exceptions) widely taught or appreciated in the English-speaking world. Such natural and basic concepts as the stochastic integral with respect to semimartingales, the general differentiation rule and the dual predictable projection should be familiar to a larger audience, so that still more applications and results might be found.

This book is, therefore, at a first-year graduate level. The first part is, of course, largely drawn from the original works of the French school, particularly those of Dellacherie, Jacod and Meyer, but the development is hopefully almost self-contained. Most proofs are carefully given in full (an exception, for example, being the proof of the section theorem). However, the aim is to reach the results of the stochastic calculus in as direct a manner as possible, so embellishments and extensions of the theory are not usually given. Also the original approach and definitions of the French authors are followed when these appear more intuitive than the even more abstract (although beautiful) recent treatments in, for example, the second editions of Dellacherie and Meyer's *Probabilités et Potentiel*. (So a predictable stopping time is a stopping time which is announced by a sequence of earlier stopping times, rather than a stopping time T for which $[\![T, \infty[\![$ belongs to the σ-field generated by processes adapted to the filtration $\{\mathcal{F}_{t-}\}$.) In its treatment of strong Markov solutions of stochastic differential equations and Girsanov's theorem, this book combines the approaches of Kallianpur, Liptser and Shiryayev, and Neveu.

The use of martingale methods in stochastic control was first developed by Benes, Davis, Duncan, Haussmann and Varaiya, *inter alia*. The chapters of this book, dealing with the stochastic control of continuous and jump processes, are based on the formulation of this approach due to Davis and the author. The chapter on filtering uses the canonical decomposition of a special semimartingale and an idea of Wong to obtain the general nonlinear filtering equation and Zakai's equation for the unnormalized distribution. This technique appears to be new. The book is more elementary than those of Dellacherie and Meyer, and unlike the treatments of Kallianpur and Liptser and Shiryayev, it presents the general theory of processes and stochastic calculus in full, including discontinuous processes. The martingale approach to optimal control has not yet been described in any text. Such a self-contained treatment of stochastic calculus and its applications does not, so far, exist, and hopefully this book fills a gap in the literature.

Acknowledgements

This book has grown out of graduate courses I gave at the University of Alberta and the University of Kentucky during the academic year 1977/78. I wish to thank Professor Ghurye and Professor A. Al-Hussaini of the University of Alberta and Professor R. Rishel and Professor R. Wets of the University of Kentucky for arranging my visits and, in addition, the audiences of my lectures for mathematical stimulation and encouragement. Dr. E. Kopp and Dr. W. Kendall of the University of Hull have read sections of the manuscript and suggested many improvements. I am particularly indebted to Dr. M.H.A. Davis of Imperial College, London, for invaluable discussions and advice over the years. Gill Turpin of the Department of Pure Mathematics of the University of Hull produced a beautiful typed version (which I, nevertheless, chopped and changed). Finally, I wish to thank my family for their constant support.

Hull, UK R.J. Elliott

Contents

Part II Stochastic Processes

Part III Stochastic Integration

Part V Applications

Introduction

This book aims to take a reader, with a basis in classical real analysis, through the theory of stochastic processes, the stochastic calculus, applications in control and filtering. The aim is to present a largely self-contained theory, setting out the foundations before proceeding to build upon them. The broad structure of this book is as follows.

Part I of this book deals with the basics of measure theory and probability. In Chapter 1, we give a quick summary of the key pertinent results from classical measure theory and real analysis, covering measures and signed measures, Lebesgue integration, spaces of functions, the monotone class theorem and the Radon–Nikodym theorem. In Chapter 2 we apply this theory to modelling probability, defining expectations and conditional expectations (with respect to σ-algebras), and connections with the theory of uniform integrability. Appendix A.1 sits with Part I as well, giving Carathéodory's extension theorem, which allows us to construct measures on various spaces.

Part II addresses stochastic processes, that is, families of random variables indexed by time. Chapter 3 explores the concept of a filtration, which is a formal way of modelling the information available at different times. It also presents the fundamental idea of stopping times, their basic properties and the σ-algebra $\mathcal{F}_T$, where T is a stopping time. Chapter 4 introduces 'martingales', which are a key class of stochastic processes with the property that their expected value in the future is the current value. Their basic properties in discrete time, including Doob's optional stopping theorem, inequalities for the maximum value attained and proofs of convergence are derived. Chapter 5 extends these results to continuous time, and also gives constructions for two of the basic martingales which are often encountered – Brownian motion and compensated Poisson processes.

Chapter 6 delves more deeply into the behaviour of stopping times, defining predictable, accessible and totally inaccessible times, and characterizing general stopping times in terms of these. It also explores the σ-algebra $\mathcal{F}_{T-}$, which describes the information available prior to a stopping time T. Chapter 7 uses these classifications to give a fine characterization of different processes, in terms of the progressive, optional and predictable σ-algebras on the product space of outcomes and time. These technical results give a general structure in which to perform stochastic integration in continuous time.

Appendices A.2, A.3, A.4 and A.5 supplement the material in Part II. Appendix A.2 proves the Kolmogorov extension theorem, which is used in one of the presented constructions of Brownian motion. Appendix A.4 gives the Kolmogorov–Čentsov theorem, which is used to establish when a process is (Hölder) continuous. Appendix A.5 considers the set of zeros of a Brownian motion, and gives an example of a set which is progressive, but not optional.

Part III builds the theory of the stochastic integral. Chapter 8 begins with the simple case where our processes are of finite variation, and so the theory of integration follows the classical Stieltjes construction. It also explores the projection of a finite variation process onto the predictable and optional processes, which provides us with the notion of a 'compensator' of a process. Chapter 9 presents the Doob–Meyer decomposition, which allows us to break many processes into the sum of a finite variation process and a martingale.

Chapter 10 defines an analogue of the L^p spaces for martingales (the $\mathcal{H}^p$ spaces), and explores their properties. It particularly focusses on the space of pure jump martingales in $\mathcal{H}^2$. Chapter 11 defines the 'quadratic variation' processes associated with a martingale, and explores how these can be used to simplify our analysis. It also proves some fundamental inequalities regarding the quadratic variations, and introduces the general class of semimartingales, as the sum of a finite variation process and a local martingale.

Chapter 12, finally, gives the general form of the stochastic integral, through Itô's isometry. It also introduces Émery's topology on the space of semimartingales, which can be seen as the operator topology when semimartingales are considered as integrators. Chapter 13 gives an extension of the theory of stochastic integration, with the theory of random measures. It begins with a presentation of the simple case of the random measure associated with a single jump, and then proceeds to the general case. As an application, we briefly introduce Lévy processes, and give a direct proof of the martingale representation theorem with respect to a random measure (with deterministic or finite activity compensator).

Appendix A.6 complements the material of Part III, proving two main results. The first shows that the integrands allowed in the stochastic integral are exhaustive, given some natural restrictions on the behaviour of the integral. The second is the Bichteler–Dellacherie–Mokobodzki theorem, which shows that the integrators allowed in the stochastic integral (the semimartingales) are exhaustive, given some weak continuity assumptions on the integral. Appendix A.8 also extends the material of Part III, discussing the class of bounded mean oscillation (BMO) semimartingales, and their basic properties.

Part IV moves from the basic stochastic integral to consider stochastic differential equations (SDEs). It begins, in Chapter 14, with the famous Itô differential rule, and its extension to the Tanaka–Meyer–Itô rule. In this chapter we also present Lévy's characterization of Brownian motion, and a construction of the Stratonovich integral. In Chapter 15, we consider a particularly simple SDE, which is satisfied by the stochastic, or Doléans-Dade, exponential. The connections of this with changes of measure are also discussed via Girsanov's theorem, along with the Novikov and Kazamaki criteria for uniform integrability in the continuous case. Appendix A.7 gives two versions, due to Lépingle and Mémin, of the Novikov condition in the presence of jumps.

Chapter 16 proves that SDEs are well posed in a general setting, with Lipschitz continuous coefficients. This is done by introducing the spaces S^p and $\mathcal{H}^p_{\mathcal{S}}$ (the semimartingale analogue of the $\mathcal{H}^p$ space). Various other basic properties, including stability and approximation schemes, and a closed form for general linear equations are presented. Chapter 17 restricts our attention to SDEs driven by a Brownian motion and a Poisson random measure, and considers their basic properties as Markov processes. The key result is the general Feynman–Kac theorem, which connects solutions of SDEs with solutions of certain partial integro-differential equations. Chapter 18 pushes this connection further and outlines how measure change techniques and solutions of martingale problems can be used to construct solutions to SDEs in a non-Lipschitz continuous setting.

Chapter 19 explores the theory of Backward SDEs, which appear in various settings in control problems. It gives a general approach to these equations, in a setting with a sequence of Brownian motions and a Poisson random measure. The comparison theorem is proven, in the presence of jumps, and connections to semilinear PIDEs are also discussed. Appendix A.9 extends these results to allow BSDEs with coefficients which are not uniformly Lipschitz to be considered. We give a presentation of Tevzadze's construction for quadratic-growth BSDEs (with jumps), and also an extension of Hamadène and Lepeltier's approach to BSDEs with stochastic Lipschitz coefficients.

Part V considers applications of this theory to problems in control and filtering. Chapter 20 presents the simple case where a controller determines the rates associated with a single jump process. Chapter 21 gives the general setting of a controller who can determine the drift and jump rates of an SDE, by first considering the connection between BSDEs and the martingale optimality principle. Appendix A.10 supplements these chapters, providing the proof of Beneš' extension of Filippov's implicit function theorem, which allows us to select measurable controls in a general way.

Chapter 22 concludes by considering a classical filtering problem, where a Markov process X is observed only through the drift of continuous process Y. The filtering equation and Zakai equation are derived, as is the Kalman filter as a special case. We also outline the case when X is a finite-state Markov chain, and so the finite-dimensional Wonham filter appears for the state process. The calculation of various associated quantities, which are important for statistical calibration, is also presented.

Part I

Measure Theoretic Probability

1

Measure and Integral

In the first two chapters, we outline definitions and results from basic real analysis and measure theory, and their application to probability. These concepts form the foundation for all that follows.

The results presented here are intended as a revision of the relevant theory, with some extensions beyond what is typically covered in a first course on measure theory. For thorough treatments, with more extensive discussion, examples, and motivation, we recommend the books by Capiński and Kopp [29], Billingsley [16], and Shiryaev [166] for a treatment of measure theory as it pertains to probability, or to the classic works by Royden and Fitzpatrick [160] and Rudin [163] for a general approach.

Remark 1.0.1. In this book, we adopt the convention $\mathbb{N} = \{1, 2, \dots\}$, that is, 0 is not considered a natural number. We write $\mathbb{Z}^+ = \mathbb{N} \cup \{0\}$ and $\emptyset$ for the empty set.

We denote by $\overline{\mathbb{R}}$ the set of extended real numbers, $[-\infty, \infty]$. This has a natural topology, where intervals of the form $[-\infty, \infty]$, $[-\infty, a[,]a, \infty]$ and $]a, b[$, for $a, b \in \overline{\mathbb{R}}$, generate the open sets (that is, the open sets are arbitrary unions of these intervals). In keeping with the French style of notation, we denote by $[a, b[$ the interval $\{x : a \leq x < b\}$, and similarly for $]a, b]$, $]a, b[$, etc…

1.1 Boolean Algebras and σ-Algebras

Underlying the mathematical theory of probability is the theory of sets. For much of analysis and probability, a key structure is given by collections of sets, in particular, by collections of subsets of some set S.

S.N. Cohen, R.J. Elliott, *Stochastic Calculus and Applications*,
Probability and Its Applications, DOI 10.1007/978-1-4939-2867-5_1

The basic aim of measure theory is to assign a 'size' to a large class of sets, extending our intuitive notions of the size of a finite set or an interval. The problem is that one can find sets for which the notion of 'size' is poorly defined, so we need to proceed carefully.

Definition 1.1.1. *Let S be a set. A collection of subsets Σ of S is called a* (Boolean) algebra *of S (or* field *of subsets of S) provided*

(i) $\emptyset \in \Sigma$,
(ii) if $A \in \Sigma$ *then* $A^c := S \setminus A \in \Sigma$,
(iii) if $m \in \mathbb{N}$ *and* $A_n \in \Sigma$ *for* $n = 1, 2, \ldots, m$ *then* $\bigcup_{n=1}^{m} A_n \in \Sigma$.

If, furthermore, (iii) can be strengthened to

(iii') if $A_n \in \Sigma$ *for* $n = 1, 2, \ldots$, *then* $\bigcup_{n=1}^{\infty} A_n \in \Sigma$.

*then Σ is called a σ-*algebra *(or σ-field) on S. If Σ is a σ-algebra on S, then the pair (S, Σ) is called a* measurable space.

Remark 1.1.2. The difference between an algebra on S and a σ-algebra on S is that an algebra is assumed to be closed only under finite unions (that is, the union of a finite number of elements of Σ will also be an element of Σ), whereas a σ-algebra is assumed to be closed under countable unions.

Neither an algebra nor a σ-algebra is assumed to be closed under uncountable unions.

Remark 1.1.3. Clearly, (i) and (ii) imply that $S \in \Sigma$. It is easy to show that (ii) and (iii) imply: if $A_n \in \Sigma$ for $n = 1, 2, \ldots$ then $\bigcap_{n=1}^{\infty} A_n \in \Sigma$.

Example 1.1.4. A few classic examples of algebras and σ-algebras:

(i) For any set S, the trivial σ-algebra $\Sigma = \{\emptyset, S\}$ and the power set 2^S (that is, the set of all subsets of S) are both σ-algebras on S.
(ii) If $A \subseteq S$, then $\Sigma = \{\emptyset, A, A^c, S\}$ is a σ-algebra on S.
(iii) Let $\mathfrak{I}$ consist of all sets of the form $\{]a, b] : -\infty \leq a \leq b < \infty\}$ or $\{]a, \infty[: -\infty \leq a < \infty\}$ and suppose $\Sigma_{\mathfrak{I}}$ is the collection of all finite (disjoint) unions of sets in $\mathfrak{I}$. Then $\Sigma_{\mathfrak{I}}$ is an algebra of subsets of $\mathbb{R}$ (but not a σ-algebra).

Remark 1.1.5. In many circumstances, the set S may only be implicitly considered, as our attention will be on the algebra Σ. For any algebra Σ on a set S, it is true that $S = \bigcup_{A \in \Sigma} A$, so this does not lead to confusion.

Theorem 1.1.6. *Let $\mathcal{G}$ be a collection of subsets of a set S. Then there exists a smallest σ-algebra on S which contains $\mathcal{G}$. This is denoted $\sigma(\mathcal{G})$ and is called the σ-algebra generated by $\mathcal{G}$.*

Proof. Let $\{\Sigma_\alpha\}_{\alpha \in \mathcal{A}}$ be the collection of all σ-algebras on S such that $\mathcal{G} \subset \Sigma_\alpha$ for every $\alpha \in \mathcal{A}$. This is not empty, as it contains 2^S. Set $\Sigma = \bigcap_{\alpha \in \mathcal{A}} \Sigma_\alpha$. By Exercise 1.8.1, Σ is a σ-algebra and any other σ-algebra containing $\mathcal{G}$ also contains Σ. $\square$

It is often important to see the interaction between σ-algebras and the topology of a set.

Definition 1.1.7. *A* topology *on a set S is a collection $\mathcal{T}$ of subsets of S satisfying*

(i) $\emptyset \in \mathcal{T}$ *and* $S \in \mathcal{T}$*,*
(ii) if $O_\alpha \in \mathcal{T}$*, for all* α *in some (possibly uncountable) index set* $\mathcal{A}$*, then* $\bigcup_{\alpha \in \mathcal{A}} O_\alpha \in \mathcal{T}$*,*
(iii) if $m \in \mathbb{N}$ *and* $\{O_n\}_{n=1}^m \subset \mathcal{T}$*, then* $\bigcap_{n=1}^m O_n \in \mathcal{T}$*.*

The pair $(S, \mathcal{T})$ *is called a* topological space *and the elements of* $\mathcal{T}$ *are called the open subsets in* $(S, \mathcal{T})$*.*

Example 1.1.8. For $S = \mathbb{R}$, the classical 'open' sets, that is, sets that can be written as an arbitrary union of open intervals $]a, b[$, form a topology of $\mathbb{R}$.

Definition 1.1.9. *Let S be a topological space with topology $\mathcal{T}$. The* Borel σ-algebra*, denoted $\mathcal{B}(S)$, is the σ-algebra generated by the open sets in $\mathcal{T}$; that is, $\mathcal{B}(S)$ is the smallest σ-algebra that contains $\mathcal{T}$. The elements of $\mathcal{B}(S)$ are called the* Borel sets *of S.*

We use $\mathcal{B}(\mathbb{R})$ to denote the Borel σ-algebra generated by the topology consisting of all unions of open intervals in $\mathbb{R}$.

Remark 1.1.10. We note that, following the notation of Example 1.1.4, $\sigma(\Sigma_{\mathfrak{I}}) = \mathcal{B}(\mathbb{R})$. It is left to the reader to fill in the details (Exercise 1.8.2).

It is important to note that there are usually many possible σ-algebras on a set S.

Definition 1.1.11. *Let Σ_n be a collection of σ-algebras on a set S. Then we define $\bigvee_n \Sigma_n := \sigma(\bigcup_n \Sigma_n)$, that is, $\bigvee_n \Sigma_n$ is the smallest σ-algebra with $\Sigma_m \subseteq \bigvee_n \Sigma_n$ for all m. As $\bigcup_n \Sigma_n$ is simply a collection of subsets of S, $\bigvee_n \Sigma_n$ exists by Theorem 1.1.6.*

1.1.1 The Monotone Class Theorem

We now prove a fundamental result known as the *monotone class theorem.* This technical result will simplify some proofs considerably, as it allows us to take any desired property, prove it holds for a 'monotone class' and then conclude that it holds for any σ-algebra within that class.

Definition 1.1.12. *A family of sets $\mathcal{M}$ is said to be a* monotone class *if $A \in \mathcal{M}$ whenever $\{A_n\}_{n \in \mathbb{N}}$ is a sequence of sets in $\mathcal{M}$ with either*

(i) $A_n \subseteq A_{n+1}$ *and* $\bigcup_{n=0}^\infty A_n = A$*, or*
(ii) $A_{n+1} \subseteq A_n$ *and* $\bigcap_{n=0}^\infty A_n = A$*.*

Lemma 1.1.13. *Let $\mathcal{K}$ be a family of sets. Then there is a smallest monotone class containing $\mathcal{K}$.*

Proof. Let $\mathfrak{M}$ denote the collection of all monotone classes containing $\mathcal{K}$. As $\mathfrak{M}$ contains the power set of $S = \bigcup \mathcal{K}$, we know that $\mathcal{K} \subseteq \bigcap \mathfrak{M}$, and it is easy to verify that $\bigcap \mathfrak{M}$ is a monotone class. Hence $\bigcap \mathfrak{M}$ is the smallest monotone class containing $\mathcal{K}$. □

Theorem 1.1.14 (Monotone Class Theorem). *Let S be a set, and $\mathcal{N}$ an algebra of subsets of S (but not necessarily a σ-algebra). Suppose $\mathcal{M}$ is a monotone class of subsets of S which contains $\mathcal{N}$. Then $\mathcal{M}$ contains the σ-algebra $\sigma(\mathcal{N})$. Furthermore, $\sigma(\mathcal{N})$ is the smallest monotone class containing $\mathcal{N}$.*

Proof. Let $\mathfrak{m}(\mathcal{N})$ denote the smallest monotone class containing $\mathcal{N}$. It is enough to check that $\sigma(\mathcal{N}) = \mathfrak{m}(\mathcal{N})$. As $\sigma(\mathcal{N})$ is a σ-algebra, it is a monotone class, so $\mathfrak{m}(\mathcal{N}) \subseteq \sigma(\mathcal{N})$.

For a set A, let

$$\mathfrak{M}_A = \big\{B \in \mathfrak{m}(\mathcal{N}) : A \cap B, A \cup B \text{ and } A \setminus B \in \mathfrak{m}(\mathcal{N})\big\} \subseteq \mathfrak{m}(\mathcal{N}).$$

By direct calculation, we can see that $\mathfrak{M}_A$ is a monotone class for any A. As $\mathcal{N}$ is a Boolean algebra, $\mathcal{N} \subseteq \mathfrak{M}_A$ for any $A \in \mathcal{N}$. As $\mathfrak{m}(\mathcal{N})$ is the smallest monotone class containing $\mathcal{N}$, it follows that $\mathfrak{M}_A = \mathfrak{m}(\mathcal{N})$ for any $A \in \mathcal{N}$.

Therefore, we know that for any $A \in \mathcal{N}$ and any $B \in \mathfrak{m}(\mathcal{N})$, the sets $A \cap B$, $A \cup B$ and $A \setminus B$ are all in $\mathfrak{m}(\mathcal{N})$. This implies $\mathcal{N} \subseteq \mathfrak{M}_B$, and again by minimality of $\mathfrak{m}(\mathcal{N})$ we know $\mathfrak{m}(\mathcal{N}) = \mathfrak{M}_B$ for all $B \in \mathfrak{m}(\mathcal{N})$. It follows that $\mathfrak{m}(\mathcal{N})$ is a Boolean algebra, and is closed under countable unions (as it is a monotone class), and is hence a σ-algebra. By minimality of $\sigma(\mathcal{N})$, it follows that $\sigma(\mathcal{N}) \subseteq \mathfrak{m}(\mathcal{N})$, as desired. □

Remark 1.1.15. A typical application of this result is to consider a simple algebra Σ (for example, the intervals of $\mathbb{R}$), and to define $\mathcal{M}$ to be the collection of sets in $\sigma(\Sigma)$ where some property holds. If we show that

(i) the algebra Σ lies in $\mathcal{M}$,
(ii) limits of monotone sequences in $\mathcal{M}$ lie in $\mathcal{M}$,

then the monotone class theorem allows us to conclude that $\mathcal{M}$ contains all of $\sigma(\Sigma)$. See the proof of Theorem 1.4.5 for an example of such an argument.

A closely related result, sometimes also referred to as the monotone class theorem, is due to Dynkin.

Definition 1.1.16. *A collection $\mathcal{N}$ of sets is called a λ-system (or d-system) on S if*

(i) $S \in \mathcal{N}$,
(ii) For any $A, B \in \mathcal{N}$ with $A \subseteq B$, $B \setminus A \in \mathcal{N}$,

(iii) if $\{A_n\}_{n\in\mathbb{N}} \subseteq \mathcal{N}$ *and* $A_n \subseteq A_{n+1}$ *for all* n*, then* $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{N}$*.*

A collection $\mathcal{N}$ *of sets is called a* π*-system if it is stable under finite intersections, that is* $A \cap B \in \mathcal{N}$ *whenever* $A, B \in \mathcal{N}$*.*

Lemma 1.1.17. *A collection* $\mathcal{K}$ *of subsets of* S *is a* σ*-algebra if and only if it is both a* π*-system and a* λ*-system.*

Proof. Clearly every σ-algebra is both a π-system and a λ-system. To prove the converse, we only need to prove that $\mathcal{K}$ is closed under countable unions (of not necessarily increasing sets) whenever $\mathcal{K}$ is a π-system and a λ-system. For any $A, B \in \mathcal{K}$, we know $A \cup B = S \setminus (A^c \cap B^c) \in \mathcal{K}$. Hence for any $\{A_n\}_{n\in\mathbb{N}} \subseteq \mathcal{K}$ the sequence $B_n := \bigcup_{k\leq n} A_k$ satisfies $B_n \subseteq B_{n+1}$, and so

$$\bigcup_{n\in\mathbb{N}} A_n = \bigcup_{n\in\mathbb{N}} B_n \in \mathcal{K},$$

so $\mathcal{K}$ is a σ-algebra. □

Theorem 1.1.18 (Dynkin's π-λ-Systems Lemma). *If* $\mathcal{N}$ *is a* π*-system, then any* λ*-system containing* $\mathcal{N}$ *contains* $\sigma(\mathcal{N})$*.*

Proof. As for the monotone class argument, we first define $\lambda(\mathcal{N})$ to be the intersection of all λ-systems containing $\mathcal{N}$, and it is easy to check that $\lambda(\mathcal{N})$ is a λ-system. It remains to prove that $\lambda(\mathcal{N})$ is a π-system. The proof is similar to that for the monotone class theorem, so we provide only a sketch: consider the set $\Lambda_1 = \{B \in \lambda(\mathcal{N}) : B \cap A \in \lambda(\mathcal{N}) \text{ for all } A \in \mathcal{N}\}$. Then Λ_1 is a λ-system, and as $\mathcal{N}$ is a π-system we can check that $\mathcal{N} \subseteq \Lambda_1$. However this implies $\Lambda_1 = \lambda(\mathcal{N})$. Then let $\Lambda_2 = \{B \in \lambda(\mathcal{N}) : B \cap A \in \lambda(\mathcal{N}) \text{ for all } A \in \lambda(\mathcal{N})\}$. Similarly as for Λ_1, we observe Λ_2 is a λ-system and $\Lambda_2 = \lambda(\mathcal{N})$. This implies $\lambda(\mathcal{N})$ is a π-system, as required. □

1.2 Set Functions and Measures

In many situations, we wish to generalize the notion of the 'size' of a set. We are used to this idea when thinking about discrete sets – where the size is simply the number of elements – or for intervals on the real line – where the size is the length of the interval. It is not clear, however, how this would rigorously generalize to other spaces. Measure theory allows us to do this in a general way.

Definition 1.2.1. *By a* set function *we mean a map*

$$\mu : \Sigma \to \overline{\mathbb{R}},$$

where Σ *is a collection of sets. For simplicity, we shall hereafter assume that a set function takes at most one of the values* $-\infty$ *and* ∞*.*

Definition 1.2.2. *A set function μ defined on an algebra of sets Σ is said to be* finitely additive *if $\mu(\emptyset) = 0$ and, for $m \in \mathbb{N}$,*

$$\mu\Big(\bigcup_{n=1}^{m} A_n\Big) = \sum_{n=1}^{m} \mu(A_n),$$

whenever $A_i \cap A_j = \emptyset$ for all $i \neq j$.

If

$$\mu\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} \mu(A_n),$$

whenever $A_i \cap A_j = \emptyset$ for all $i \neq j$ and $\bigcup_{n=1}^{\infty} A_n \in \Sigma$, then μ is said to be countably additive*.*

Definition 1.2.3. *Given a σ-algebra of sets Σ, a* measure *on Σ is a countably additive set function $\mu : \Sigma \to [0, \infty]$.*

If $\mu(S) < \infty$, then μ is called a finite measure*. If there exists a sequence $\{A_n\}_{n\in\mathbb{N}}$, of sets in Σ such that $\mu(A_n) < \infty$ for all n and $\cup_n A_n = S$, then μ is said to be a σ-*finite *measure. If we need to clarify which σ-algebra we are working with, then we will write Σ-σ-finite.*

Definition 1.2.4. *If μ is a measure on the measurable space (S, Σ), then the triple (S, Σ, μ) is called a* measure space*.*

Remark 1.2.5. Most of the spaces we shall consider will be σ-finite, and many will be finite (where $\mu(S) < \infty$).

Example 1.2.6. A few classic measure spaces.

(i) Suppose $S = \{H, T\}$, $\Sigma = 2^S = \{\emptyset, \{H\}, \{T\}, \{H, T\}\}$ and

$$\mu(A) = \begin{cases} 0 & A = \emptyset \\ 1/2 & A = \{H\} \text{ or } \{T\} \\ 1 & A = \{H, T\} \end{cases}.$$

Then (S, Σ, μ) is a measure space. (This is, of course, a standard model for the outcomes of a toss of a fair coin, where $\mu(A)$ gives the probability of an outcome in A, H corresponds to observing a head, and T to a tail.)

(ii) Suppose Σ is a σ-algebra of subsets of an arbitrary set S. Then the set function

$$\mu(A) := \text{number of elements in } A$$

defines a measure called the *counting measure*. Clearly μ is σ-finite if and only if A is countable.

(iii) Let $S = \mathbb{R}^n$, $n \in \mathbb{N}$ and $\Sigma = \mathcal{B}(\mathbb{R}^n)$. (Here, $\mathcal{B}(\mathbb{R}^n)$ is the smallest σ-algebra containing all the open rectangles $R =]a_1, b_1[\times]a_2, b_2[\times \cdots \times]a_n, b_n[$.) There exists a unique measure λ on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ which assigns to each rectangle R the measure

$$\lambda(R) = \prod_{i=1}^{n} (b_i - a_i).$$

This measure is commonly known as the *Lebesgue measure* on $\mathbb{R}^n$. We construct this measure explicitly for $\mathbb{R}$ in Appendix A.1.

At times, constructing measures on general spaces can be difficult. This is made considerably easier by the following result, which is the key consequence of Carathéodory's extension theorem (Theorem A.1.17), and is proven in Appendix A.1. While this construction is important, in that without it we could define very few interesting examples of measures, the details are usually not the main focus in applications.

Theorem 1.2.7. *Let Σ be an algebra of sets (but not necessarily a σ-algebra), and let $\mu : \Sigma \to \mathbb{R}$ be*

- *countably additive, that is, for any sequence $\{A_n\}_{n\in\mathbb{N}} \subset \Sigma$ with $A_n \cap A_m = \emptyset$ for $n \neq m$, and $\cup_n A_n \in \Sigma$, we have $\mu\big(\bigcup_{n\in\mathbb{N}} A_n\big) = \sum_{n\in\mathbb{N}} \mu(A_n)$, and*
- *σ-finite, in the sense that there exists a sequence A_n with $\bigcup_n A_n = S$ and $\mu(A_n) < \infty$ for each n.*

Then there exists a unique extension of μ to a measure on the σ-algebra $\sigma(\Sigma)$.

This theorem allows us to construct measures in a simple way, by constructing them on Boolean algebras, and then (given we can verify countable additivity and σ-finiteness) directly generalizing them to measures on the corresponding σ-algebras.

Definition 1.2.8. *A measure space (S, Σ, μ) will be called* complete *if it contains every subset of every set of measure zero. That is, for any $A \subseteq B \in \Sigma$ with $\mu(B) = 0$, we have $A \in \Sigma$.*

Example 1.2.9. For any set S, the measure space $(S, 2^S, \mu)$, where μ is the counting measure, is a complete measure space.

The space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$, where μ is Lebesgue measure, is not a complete measure space – proving this requires the Axiom of Choice, see Gelbaum and Olmsted [86, p.98].

Definition 1.2.10. *If (S, Σ) and (S, Σ') are two measurable spaces with $\Sigma \subseteq \Sigma'$, then, for any measure μ on (S, Σ'), we write $\mu|_\Sigma$ for the restriction of μ to Σ.*

Lemma 1.2.11. *For any measure space* (S, Σ^0, μ^0)*, there exists a complete measure space* (S, Σ, μ)*, where* $\Sigma^0 \subseteq \Sigma$ *and* $\mu|_{\Sigma^0} = \mu^0$*. The space* (S, Σ, μ) *is called the* completion *of* (S, Σ^0, μ^0)*.*

Proof. Define Σ as follows. Let $A' \in \Sigma$ if there exists $A, B^0, C^0 \in \Sigma^0$ with $\mu(B^0) = \mu(C^0) = 0$, and $A' = (A \cup B) \setminus C$ for some $C \subseteq C^0, B \subseteq B^0$. Note $A' \in \Sigma$ if and only if A' differs from A by a subset of a set of measure zero, for some $A \in \Sigma^0$. It is straightforward to show that this is a σ-algebra.

For A' of the above form, let $\mu(A') := \mu(A)$. If $A' = (A \cup B) \setminus C = (\tilde{A} \cup \tilde{B}) \setminus \tilde{C}$, then A and $\tilde{A}$ differ only on a set of measure zero, and so $\mu(A') = \mu(A) = \mu(\tilde{A})$ is well defined.

Clearly $\mu|_{\Sigma} = \mu^0$, as we let $B = C = \emptyset$. Also, (S, Σ, μ) is complete, as if $\mu(A') = 0$, then $A' \subseteq A \cup B$ with $\mu^0(A) = 0$. Hence any subset $\tilde{A}$ of A' is a subset of $A \cup B$, which is a set of measure zero. Hence $\tilde{A}$ differs from A by a subset of a set of measure zero, and so is in Σ. □

Example 1.2.12. If $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$ denotes the measure space on the Borel sets of $\mathbb{R}$ under Lebesgue measure, then we define $(\mathbb{R}, \overline{\mathcal{B}}(\mathbb{R}), \overline{\mu})$ to be the completion of $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$. Sets in $\overline{\mathcal{B}}(\mathbb{R})$ are called the *Lebesgue measurable* subsets of $\mathbb{R}$.

Definition 1.2.13. *For* (S, Σ, μ) *a measure space, we say that* μ charges *a set* $A \in \Sigma$ *if* $\mu(A) > 0$*. A set* $A \in \Sigma$ *is called an* atom *if* $\mu(A) > 0$ *and, for all* $B \in \Sigma$ *with* $B \subseteq A$*, either* $\mu(B) = 0$ *or* $\mu(B) = \mu(A)$*.*

1.2.1 Distribution Functions & Lebesgue–Stieltjes Measures

We would like to be able to construct interesting measures in a simple way. Distribution functions give a nice way of doing this, which will prove adequate for many simple problems in probability theory. There is a direct link between distribution functions as discussed here and the (cumulative) distribution functions considered in many basic courses on probability and statistics.

Definition 1.2.14. *Recall that a function* F *is right-continuous if for all* t*,* $\lim_{h\downarrow 0} F(t+h) = F(t)$*. A nondecreasing, right-continuous function* $F : \mathbb{R} \to \mathbb{R}$ *is called a* distribution function*.*

Definition 1.2.15. *A measure* μ *on* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *(more generally, on* $(\mathbb{R}, \overline{\mathcal{B}}(\mathbb{R}))$*) is called a* Lebesgue–Stieltjes measure *or* Baire measure *if, for any bounded interval* I *(whether open, half-open or closed),* $\mu(I) < \infty$*.*

Theorem 1.2.16. *There is a one-to-one correspondence between distribution functions (up to addition by a constant) and Lebesgue–Stieltjes measures on* $\mathcal{B}(\mathbb{R})$*, given by*

$$\mu(]a, b]) = F(b) - F(a)$$

and the requirement $F(0) = 0$*.*

Proof. See Appendix A.1 (Theorem A.1.20). ☐

Remark 1.2.17. Let μ be Lebesgue measure on $\mathbb{R}$. Then μ is a Lebesgue–Stieltjes measure, corresponding to the distribution function

$$F(t) = t.$$

Remark 1.2.18. In light of the preceding theorem, we may regard all Lebesgue–Stieltjes measures as arising from distribution functions and vice versa.

Lemma 1.2.19. *Let μ be a Lebesgue–Stieltjes measure on $\mathcal{B}(\mathbb{R})$. A set $A = \{t\}$ is an atom under μ if and only if the distribution function F is discontinuous at t.*

Proof. Simply note that

$$\mu(\{t\}) = \lim_{h \downarrow 0} \mu(]t-h,t]) = \lim_{h \downarrow 0} \big(F(t) - F(t-h)\big) = F(t) - F(t-),$$

where $F(t-)$ is the left-limit $F(t-) = \lim_{s \uparrow t} F(s)$ (which exists as F is non-decreasing). ☐

Example 1.2.20. (i) Take $F(t) = I_{\{t \geq 0\}} - 1$. This yields a measure μ with $\mu(\{0\}) = 1$ and $\mu(A) = 0$ for $A \not\ni 0$.

(ii) Take $F(t) = \lfloor t \rfloor$, where $\lfloor t \rfloor$ denotes the greatest integer less than or equal to t. Then we have a measure μ with $\mu(\{n\}) = 1$ for every integer n and $\mu(\mathbb{R} \setminus \mathbb{N}) = 0$.

(iii) Take $F(t) = \int_{]-\infty,t]} f(x)dx$ for some integrable function $f \geq 0$. Then F is a continuous distribution function.

(iv) Take $F(t)$ to be the Cantor–Lebesgue function, which increases only on the points of the Cantor set (see [160, p.51]). Then μ is a measure which has $\mu(A) = 0$ for all sets A not intersecting the Cantor set. The function F is also continuous, so does not charge single points (that is, there are no atoms under this measure).

1.3 The Lebesgue Integral

We now seek to use this theory of measure to define a theory of integration. This will generalize the more familiar concept of Riemann integration, and allow us to take integrals over more general spaces. For the purposes of probability theory, we particularly wish to be able to calculate integrals over abstract spaces of ‘outcomes,’ which are often considerably larger than the real line.

1.3.1 Measurable Functions

We begin by defining a space of functions for which integration is possible – the measurable functions. These are those functions which are 'well behaved enough' that they can be considered using the tools of measure theory.

Definition 1.3.1. *Suppose (S, Σ) and $(E, \mathcal{E})$ are both measurable spaces. A function $f : S \to E$ is called $\Sigma/\mathcal{E}$-* measurable *if $f^{-1}(B) \in \Sigma$ for every $B \in \mathcal{E}$.*

If Σ is a Borel σ-algebra on S, then f is said to be Borel measurable.

Remark 1.3.2. For a function $f : S \to \overline{\mathbb{R}}$, we shall typically take $(E, \mathcal{E}) = (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ as implicit. In this case, we shall often simply say that f is Σ-measurable.

If $f : \overline{\mathbb{R}} \to \overline{\mathbb{R}}$, then we shall often also assume $(S, \Sigma) = (\overline{\mathbb{R}}, \overline{\mathcal{B}}(\overline{\mathbb{R}}))$, and simply say that f is a (Lebesgue) measurable function. If $(S, \Sigma) = (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$, then as above we say that f is a Borel measurable function.

The choice of including the sets of measure zero in Σ but not $\mathcal{E}$ can cause problems, as it implies the composition of measurable functions is not necessarily measurable (see Exercise 1.8.6). At the same time, taking $\mathcal{E} = \mathcal{B}(\overline{\mathbb{R}})$ is needed to ensure that all continuous functions are measurable, and taking Σ complete is needed to ensure that if $f = g$ except on (a subset of) a set of measure zero and f is measurable, then g is measurable.

Remark 1.3.3. As the sets $]a, \infty[$ generate the Borel σ-algebra, we could equivalently define $f : S \to \mathbb{R}$ as measurable when $\{s : f(s) \leq a\} = \{s : f(s) > a\}^c \in \Sigma$, for all $a \in \mathbb{R}$. (This is easily seen using the monotone class theorem, but can be shown directly without much difficulty.)

Remark 1.3.4. Our definition of measurability is similar to the definition of the continuous functions in a general topological space, i.e. a function f is continuous if $f^{-1}(B)$ is open for every open set B. An immediate corollary of this is, if S and E are topological spaces, $\mathcal{E} = \mathcal{B}(E)$ and $\Sigma \supseteq \mathcal{B}(S)$, then any continuous function is measurable.

Remark 1.3.5. Various other properties of measurable functions are given in Exercises 1.8.4 and 1.8.7, in particular the measurable functions are closed under the operations of addition, subtraction, multiplication and division, as well as taking countable limits, suprema, infima, maxima and minima. As mentioned before, the composition of measurable functions is not generally measurable, but if f is a *Borel* measurable function and g is (Lebesgue) measurable, then the composition $f \circ g$ is (Lebesgue) measurable.

Definition 1.3.6. *Two measurable functions f and g on a measure space (S, Σ, μ) will be said to be equal* almost everywhere *(a.e.) if*

$$\mu(\{s : f(s) \neq g(s)\}) = 0.$$

Remark 1.3.7. In general, a statement will be true 'almost everywhere', or 'for almost all (a.a.) s', if the set of values where it is false has measure zero.

Definition 1.3.8. *If (S, Σ) is a measurable space, the function*

$$I_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

is a measurable function if and only if $A \in \Sigma$. I_A is called the indicator function *(or sometimes characteristic function) of A.*

Definition 1.3.9. *If (S, Σ) is a measurable space, a function $\phi : S \to \mathbb{R}$ is called* simple *if ϕ is measurable and it takes only a finite number of values.*

Remark 1.3.10. It is easy to see that a function ϕ is simple if and only if it can be written as a finite sum

$$\phi(x) = \sum_{n=1}^{m} x_n I_{A_n}(x),$$

where $\{A_i\}_{i \le m} \subseteq \Sigma$ and $A_i \cap A_j = \emptyset$ for $i \neq j$.

Definition 1.3.11. *Let f be a function from a set S to a measurable space $(E, \mathcal{E})$. Then the σ-algebra on S given by $\{f^{-1}(A)\}_{A \in \mathcal{E}}$, is the smallest σ-algebra such that f is measurable, and is called the σ-algebra generated by f. We denote this σ-algebra $\sigma(f)$.*

This clearly extends to collections of functions $\{f_a\}_{a \in \mathcal{A}}$.

$$\begin{tikzcd}
(S, \sigma(f)) \arrow[r, "f"] \arrow[dr, "g"'] & (E, \mathcal{E}) \arrow[d, "h"] \\
 & (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))
\end{tikzcd}$$

Fig. 1.1. The Doob–Dynkin Lemma as a commuting diagram

Theorem 1.3.12 (Doob–Dynkin Lemma). *Let f be a function from S to a measurable space $(E, \mathcal{E})$, and let $\sigma(f)$ denote the σ-algebra generated by f. Let g be a measurable function $S \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$. Then g is $\sigma(f)$-measurable if and only if there exists a measurable function $h : (E, \mathcal{E}) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ such that*

$$g(s) = h \circ f(s).$$

Proof. It is straightforward to see that if $g = h \circ f$, then g is $\sigma(f)$-measurable.

To see the converse, first assume $g = I_A$ for some set A (see Definition 1.3.8 above). Then g is $\sigma(f)$-measurable if and only if $A \in \sigma(f)$, that is, if $A = f^{-1}(B)$ for some $B \in \mathcal{E}$. Let $h = I_B$. Then $g = h \circ f$.

Next assume is a simple function, that is, we can write $g = \sum_i x_i I_{A_i}$. Then there exist $B_i \in \mathcal{E}$ with $A_i = f^{-1}(B_i)$. Hence let $h = \sum_i x_i I_{B_i}$, and we have $g = h \circ f$.

Now assume g is measurable. Define a sequence of simple functions ϕ_n converging pointwise to g. Then $\phi_n = h_n \circ f$ for all n. For fixed f, as $h_n \circ f$ converges everywhere, we know that h_n converges everywhere within the range of f. Hence

$$h := \begin{cases} \lim_n h_n & \text{when it exists} \\ 0 & \text{otherwise} \end{cases}$$

is a well defined function and satisfies $g = h \circ f$. □

Remark 1.3.13. This theorem helps us to develop an intuition for what it means to be measurable in a more general σ-algebra, as it shows that if g is measurable with respect to the σ-algebra generated by f, then g is a function of the result of f. Hence f can be thought of as containing all relevant information needed to calculate g (Fig. 1.1).

1.3.2 Integration

Given a measure μ on Σ, we first define the *Lebesgue integral* of a simple function ϕ over a set of finite measure.

Definition 1.3.14. *Let $\phi = \sum_{n=1}^m x_n I_{A_n}$ be a simple function on a measure space (S, Σ, μ), and suppose ϕ is zero outside a set of finite measure, that is, $\mu(\{s : \phi(s) \neq 0\}) < \infty$. Then the (Lebesgue) integral of ϕ over S with respect to μ is*

$$\int_S \phi \, d\mu := \sum_{n=1}^m x_n \mu(A_n).$$

One can easily verify that the integral is independent of the representation of ϕ.

Remark 1.3.15. This terminology can be slightly confusing, as the "Lebesgue integral" of ϕ with respect to μ does not assume that μ is the "Lebesgue measure" on $\mathbb{R}$.

Definition 1.3.16. *If $f : (S, \Sigma) \to ([0,\infty], \mathcal{B}([0,\infty]))$ is a measurable function, we define the Lebesgue integral of f over S with respect to μ by*

$$\int_S f \, d\mu = \sup\Big\{ \int_S \phi \, d\mu : \phi \text{ simple}, \phi \leq f \ \mu\text{-a.e.} \Big\}.$$

Remark 1.3.17. We shall see in a moment that this supremum is achieved as the limit of *any* sequence of simple functions ϕ_n increasing pointwise to f almost everywhere (Theorem 1.3.29). Therefore, an explicit definition of this integral can be given, as in Corollary 1.3.30. This will also allow us to prove that the integral we have defined is linear in f (Theorem 1.3.31).

Remark 1.3.18. This definition allows us to integrate all reasonable *nonnegative* functions on (S, Σ). Using this, we can construct the integral of a generic function.

Definition 1.3.19. *For simplicity of notation, write* $\min\{a, b\} = a \wedge b$ *and* $\max\{a, b\} = a \vee b$

Definition 1.3.20. *For a function f, we define f^+ and f^-, the positive and negative parts of f, by*

$$\begin{aligned} f^+ &= f \vee 0 = \max\{f, 0\} = f I_{\{f \geq 0\}}, \\ f^- &= (-f) \vee 0 = \max\{-f, 0\} = -f I_{\{f \leq 0\}}. \end{aligned}$$

It follows that $f = f^+ - f^-$ and $|f| = f^+ + f^-$. It is easy to show that f^+ and f^- are measurable if and only if f is measurable.

Definition 1.3.21. *If $f : (S, \Sigma) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ we define the Lebesgue integral of f over S with respect to μ by*

$$\int_S f \, d\mu = \int_S f^+ \, d\mu - \int_S f^- \, d\mu$$

provided at least one of the integrals on the right is finite.

Remark 1.3.22. To avoid confusion over the variable of integration, we write $\int f(s) d\mu(s)$ when needed. Note that this definition naturally fits with the linearity of the integral (which we have yet to prove).

Definition 1.3.23. *If $A \in \Sigma$, we define*

$$\int_A f \, d\mu = \int_S I_A f \, d\mu.$$

Definition 1.3.24. *A measurable function f is said to be* integrable *over S with respect to μ if*

$$\int_S |f| \, d\mu < +\infty.$$

Remark 1.3.25. Note that the integral is well defined (but infinite) for some functions which we do not call ‘integrable’.

Remark 1.3.26. When μ is a Lebesgue–Stieltjes measure, it is natural to write

$$\int_A f \, d\mu = \int_A f \, dF$$

where F is the distribution function associated with μ. When μ is Lebesgue measure on $\mathbb{R}$, this becomes the classic notation $\int_A f(x) dx$.

1.3.3 Convergence Theorems and Properties of Integrals

This section presents key results for working with integrals. In practice, these results tell us "when we can take a limit through the integral/expectation sign", and allow us to give a more explicit construction of the integral.

Definition 1.3.27. *A sequence of measurable functions $\{f_n\}_{n\in\mathbb{N}}$ will be said to converge almost everywhere if there exists a function f such that*

$$\mu\Big(\big\{s : \lim_{n\to\infty} f_n(s) \neq f(s)\big\}\Big) = 0.$$

That is, except possibly on some set $A \in \Sigma$ with $\mu(A) = 0$, we have $f_n(s)$ converges to $f(s)$ for all $s \notin A$. We then write $f_n \to f$ a.e.

Lemma 1.3.28. *Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of measurable functions. If there exists a function f such that $\{f_n\}_{n\in\mathbb{N}}$ converges pointwise to f, (that is, $f_n(s) \to f(s)$ for all s) then f is measurable. Similarly, if $f_n \to f$ almost everywhere, then there exists a* measurable *function $\tilde{f}$ with $f_n \to \tilde{f}$ a.e.*

Proof. We know from Exercise 1.8.4 that $\liminf f_n$ and $\limsup f_n$ are both measurable. In the first case, the fact $\lim f_n = \liminf f_n = \limsup f_n$, gives the result. In the second case, except possibly on some set A with $\mu(A) = 0$, we know $\lim f_n = \liminf f_n = \limsup f_n$ exists and is measurable. Set $f = \limsup f_n$, and the result is obtained. □

Theorem 1.3.29 (Monotone Convergence Theorem). *Let $\{f_n\}_{n\in\mathbb{N}}$ be a nondecreasing sequence of nonnegative measurable functions (that is, $f_n(s) \leq f_{n+1}(s)$ for all $s \in S$ and all $n \in \mathbb{N}$). Then*

$$\int_S \lim_{n\to\infty} f_n \, d\mu = \lim_{n\to\infty} \int_S f_n \, d\mu.$$

Proof. Let $f := \lim_n f_n$, the pointwise limit of the sequence f_n, which exists as f_n is monotone and $\overline{\mathbb{R}}$ is compact (see Definition 1.5.5). Then f is a nonnegative measurable function, and so has a well-defined integral (see Exercise 1.8.4).

By Exercise 1.8.9, it is easy to see that $\int_S f_n \, d\mu \leq \int_S f d\mu$ for all n. As $\int_S f_n \, d\mu$ is nondecreasing in n, it has a well-defined limit in $\overline{\mathbb{R}}$, and hence

$$\lim_{n\to\infty} \int_S f_n \, d\mu \leq \int_S f d\mu.$$

To show the reverse inequality, we note that, from the definition of the integral, there exists a non-decreasing sequence $\{\phi_k\}_{k\in\mathbb{N}}$ of simple functions, each vanishing outside a set of finite measure, with $\phi_k \leq f$ and

$$\lim_{k\to\infty} \int_S \phi_k \, d\mu = \int_S f \, d\mu.$$

For every $\epsilon > 0$ we can define the sets $A_{\epsilon,n} = \{s : f_n(s) \geq \phi_k(s) - \epsilon\}$. If $B_k = \{s : \phi_k \geq 0\}$, so $\mu(B_k) < \infty$ for all k, then

$$\int_{A_{\epsilon,n}} \phi_k \, d\mu - \epsilon\mu(B_k) \leq \int_{A_{\epsilon,n}} f_n \, d\mu \leq \int_S f_n \, d\mu.$$

As $\lim_n f_n = f \geq \phi_k$ for all k, we allow $n \to \infty$, and hence, by Exercise 1.8.8,

$$\int_S \phi_k \, d\mu - \epsilon\mu(B_k) \leq \lim_n \int_S f_n \, d\mu.$$

Letting $\epsilon \to 0$ implies

$$\int_S \phi_k(s) \, d\mu \leq \lim_n \int_S f_n \, d\mu.$$

Therefore

$$\int_S f \, d\mu = \lim_{k\to\infty} \int_S \phi_k \, d\mu \leq \lim_n \int_S f_n \, d\mu$$

as desired. □

Corollary 1.3.30. *Consider a nonnegative function f. For $0 \leq i \leq 2^{2n}$, let $A_i^n = \{s : i2^{-n} \leq f(s) < (i+1)2^{-n}\}$. Define*

$$\phi_n(s) := \sum_{i=0}^{2^{2n}} i2^{-n} I_{A_i^n}(s).$$

Then ϕ_n is a nondecreasing sequence of measurable functions converging pointwise to f. Hence

$$\int_S f \, d\mu = \lim_{n\to\infty} \int_S \phi_n \, d\mu = \lim_{n\to\infty} \sum_{i=0}^{2^{2n}} i2^{-n}\mu(A_i^n).$$

Theorem 1.3.31. *For integrable functions, the integral is linear in the integrand. That is, for any integrable f and g and any constant λ,*

$$\int_S (\lambda f + g) \, d\mu = \lambda \int_S f \, d\mu + \int_S g \, d\mu.$$

Proof. Considering Definition 1.3.21, it is clear that it is enough to prove the result under the assumption that λ, f and g are nonnegative. By direct calculation, it is easy to check that the statement holds for f and g simple functions. Applying the monotone convergence theorem, we can approximate the integral from below by simple functions f_n and g_n. Therefore,

$$\begin{aligned}
\int_S (\lambda f + g) \, d\mu &= \lim_{n\to\infty} \Big(\int_S (\lambda f_n + g_n) \, d\mu \Big) \\
&= \lambda \lim_{n\to\infty} \int_S f_n \, d\mu + \lim_{n\to\infty} \int_S g_n \, d\mu \\
&= \lambda \int_S f \, d\mu + \int_S g \, d\mu
\end{aligned}$$

□

Corollary 1.3.32. *Let $\{f_n\}_{n\in\mathbb{N}}$ be a nondecreasing or nonincreasing sequence of measurable functions with f_1 integrable. Then*

$$\int_S \lim_{n\to\infty} f_n \, d\mu = \lim_{n\to\infty} \int_S f_n \, d\mu.$$

Proof. In the nondecreasing case, the sequence $\{f_n - f_1\}_{n\in\mathbb{N}}$ is a nonnegative, nondecreasing sequence of measurable functions. Hence

$$\int_S \lim_{n\to\infty} (f_n - f_1) \, d\mu = \lim_{n\to\infty} \int_S (f_n - f_1) \, d\mu.$$

As $\int_S f_1 \, d\mu$ is well defined, the desired result follows by linearity of the integral. For the nonincreasing case, the same argument is applied to $\{f_1 - f_n\}_{n\in\mathbb{N}}$. □

Theorem 1.3.33 (Fatou's Inequality). *Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of nonnegative measurable functions. Then*

$$\int_S \liminf_n f_n \, d\mu \leq \liminf_n \int_S f_n \, d\mu$$

Proof. Let $g_k(s) := \inf_{n\geq k} f_n(s)$. Then $\{g_k\}_{k\in\mathbb{N}}$ is a nondecreasing sequence of measurable functions, and by the monotone convergence theorem,

$$\int_S \lim_k g_k \, d\mu = \lim_k \int_S g_k \, d\mu.$$

It is clear that $\lim_k g_k = \liminf_n f_n$. It is also clear that $g_k \leq f_k$ for all k, and therefore, by Exercise 1.8.9,

$$\lim_n \int_S g_n \, d\mu \leq \liminf_n \int_S f_n \, d\mu.$$

Combining these gives the desired result,

$$\int_S \liminf_n f_n \, d\mu = \int_S \lim_k g_k \, d\mu = \lim_k \int_S g_k \, d\mu \leq \liminf_n \int_S f_n \, d\mu.$$

□

Theorem 1.3.34 (Lebesgue's Dominated Convergence Theorem). *Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of measurable functions such that $f_n \to f$ a.e. Suppose there exists a nonnegative* integrable *function g with $|f_n| \leq g$ for all n. Then*

$$\lim_n \int_S f_n \, d\mu = \int_S f \, d\mu.$$

Proof. Clearly $\{g + f_n\}_{n\in\mathbb{N}}$ and $\{g - f_n\}_{n\in\mathbb{N}}$ are two sequences of nonnegative functions. An application of Fatou's inequality gives the desired result. □

Definition 1.3.35. *For any set $A \subset \overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$, the* least upper bound, *or* supremum *of A is the smallest real number $b \in \overline{\mathbb{R}}$ such that $a \leq b$ for all $a \in A$. As $a \leq +\infty$ for all $a \in \overline{\mathbb{R}}$, such a number always exists.*

In the context of measure spaces, an analogous concept is the essential supremum *of a set $A \in \mathcal{B}(\overline{\mathbb{R}})$. This is the smallest value b satisfying*

$$\mu(x \in A : x > b) = \mu(A \cap]b, \infty]) = 0,$$

and is denoted $\operatorname{ess\,sup}(A)$. *For f a measurable function on S, we define the quantity* $\operatorname{ess\,sup} f$ *to be the essential supremum of the set $\{f(s)|s \in S\}$.*

Similarly, we can define the essential infimum *of a set A to be the largest $b \in \overline{\mathbb{R}}$ such that $\mu(x \in A : x < b) = 0$. The essential infimum of a function is defined analogously and satisfies* $\operatorname{ess\,inf}(f) = -\operatorname{ess\,sup}(-f)$.

The following theorem shows that, for finite measure spaces, a sequence of functions which converges almost everywhere must converge uniformly, except on a set with small measure.

Theorem 1.3.36 (Egorov's Theorem). *Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of measurable functions, and assume $\mu(S) < \infty$. If $f_n \to f$ a.e., then, for any $\epsilon > 0$, there exists a set A such that $\mu(A) < \epsilon$ and* $\operatorname{ess\,sup}_{s\in S\setminus A} |f_n(s) - f(s)| \to 0$ *as $n \to \infty$.*

Proof. First suppose $\mu(S) < \infty$. For $n, k \in \mathbb{N}$, define the sets

$$A_{n,k} = \bigcup_{m\geq n} \{s : |f_m(s) - f(s)| \geq k^{-1}\}.$$

As $f_m \to f$ almost everywhere, we know that $\mu(\cap_n A_{n,k}) = 0$. As $\mu(S) < \infty$, dominated convergence implies

$$\begin{aligned} 0 = \mu(\cap_n A_{n,k}) &= \int_S I_{\{\cap_n A_{n,k}\}} d\mu = \lim_{m\to\infty} \int_S I_{\{\cap_{\{n\leq m\}} A_{n,k}\}} d\mu \\ &= \lim_{m\to\infty} \mu(\cap_{\{n\leq m\}} A_{n,k}), \end{aligned}$$

and so, for each k, there exists some N_k such that $\mu(A_{N_k,k}) \leq \epsilon 2^{-k}$. Taking $A = \cup_{k\in\mathbb{N}} A_{N_k,k}$, we see

$$\mu(A) \leq \sum_k \mu(A_{N_k,k}) \leq \epsilon \sum_{k=1}^{\infty} 2^{-k} = \epsilon$$

and for $s \in S \setminus A$ we have $|f_n(s) - f(s)| \leq k^{-1}$ for all $n > N_k$, so the convergence is uniform. □

Definition 1.3.37. *A sequence of measurable functions $\{f_n\}_{n\in\mathbb{N}}$ will be said to converge* in measure *to a measurable function f if, for all $\epsilon > 0$, there is an N such that for all $n \geq N$*

$$\mu\big(\{s : |f_n(s) - f(s)| \geq \epsilon\}\big) < \epsilon.$$

Lemma 1.3.38. *Suppose $\{f_n\}_{n\in\mathbb{N}}$ is a sequence of measurable functions.*

(i) If $\mu(S) < \infty$ and $\{f_n\}_{n\in\mathbb{N}}$ converges almost everywhere, then $\{f_n\}_{n\in\mathbb{N}}$ converges in measure.
(ii) If $\{f_n\}_{n\in\mathbb{N}}$ converges in measure, then there exists a sub-*sequence $\{f_{n_k}\}_{k\in\mathbb{N}}$ which converges almost everywhere.*

Proof. (i) Let $B = \{s : f_n(s) \to f(s)\}$, so $\mu(S \setminus B) = 0$. Applying Egorov's theorem, there exists a set A such that $\mu(B\setminus A) < \epsilon$ and $f_n \to f$ uniformly on A. That is, given $\epsilon > 0$, there exists N such that $|f(s) - f_n(s)| < \epsilon$ for all $n \geq N$ and all $s \in A$. Therefore, for any $n \geq N$,

$$\{s : |f(s) - f_n(s)| > \epsilon\} \subseteq (B \setminus A) \cup (S \setminus B),$$

and the set on the right has measure at most ϵ. Therefore, $\{f_n\}_{n\in\mathbb{N}}$ converges in measure.

(ii) For any $k \in \mathbb{N}$, there exists N_k such that, for $n \geq N_k$, we know

$$\mu\big(\{s : |f(s) - f_n(s)| > 2^{-k}\}\big) < 2^{-k}.$$

Let $B_k = \{s : |f(s) - f_{N_k}(s)| > 2^{-k}\}$ and $A = \cap_{k=1}^{\infty} \cup_{i=k}^{\infty} B_k$. Then, if $s \notin A$, it follows that

$$|f(s) - f_{N_i}(s)| < 2^{-i} \quad \text{for all } i \text{ sufficiently large,}$$

so $f_n \to f$ on $S \setminus A$. Finally, we note that for any k,

$$\mu(A) \leq \mu(\cup_{i=k}^{\infty} E_i) \leq \sum_{i=k}^{\infty} 2^{-i} = 2^{-k+1},$$

so $\mu(A) = 0$.

□

Note that the limit of a sequence converging in measure is uniquely defined, up to equality almost everywhere.

Lemma 1.3.39. *The dominated convergence theorem also holds true if we only assume $f_n \to f$ in measure (along with $|f_n| \leq g$ for some integrable function g).*

Proof. For any sequence converging in measure, every subsequence also converges in measure. Hence, by Lemma 1.3.38, every subsequence contains a subsequence which converges almost surely. By our earlier dominated convergence theorem, this implies there cannot exist a subsequence $\{f_{n_k}\}_{k\in\mathbb{N}}$ such that $|\int f_{n_k} d\mu - \int f d\mu| > \epsilon$ for all n (as it contains an a.e. convergent subsequence). Therefore, $\int f_n d\mu \to \int f d\mu$, as desired. □

There is another notion of essential supremum, which is less often covered in courses on measure theory, but will be of use in various problems, particularly when we consider optimal control. This is where the supremum is not of the values of a single function, but of a collection of functions, and so the result is not a number, but another function.

Theorem 1.3.40. *Let (S, Σ, μ) be a σ-finite measure space. Let $\mathcal{F}$ be a (possibly uncountable) collection of Σ-measurable functions. Then there exists a Σ-measurable function f^* such that*

(i) $f^* \geq f$ μ-a.e. for all $f \in \mathcal{F}$,
(ii) $f^* \leq g$ μ-a.e. for all measurable g satisfying '$g \geq f$ μ-a.e. for all $f \in \mathcal{F}$'.

Suppose in addition that $\mathcal{F}$ is directed upwards, that is, for $f, f' \in \mathcal{F}$ there exists $\tilde{f} \in \mathcal{F}$ with $\tilde{f} \geq f \vee f'$ μ-a.e. Then there exists an increasing sequence $\{f_n\}_{n\in\mathbb{N}} \subseteq \mathcal{F}$ such that $f^* = \lim_n f_n$ μ-a.e.

We call the function f^* the essential supremum of $\mathcal{F}$, and write $f^* = \operatorname{ess\,sup} \mathcal{F}$. Similarly $\operatorname{ess\,inf} \mathcal{F} = -\operatorname{ess\,sup}\{-\mathcal{F}\}$. If we need to specify the sets involved, we will say that the essential infimum is taken over $\mathcal{F}$, in the Σ-measurable functions, and defined μ-a.e.

Proof. First assume that the functions in $\mathcal{F}$ are uniformly bounded above and μ is finite. If $\mathcal{F}$ is countable, then $f^*(x) := \sup_{f\in\mathcal{F}} f(x)$ is measurable (Exercise 1.8.4) and satisfies the requirements. Now consider the quantity

$$c := \sup\Big\{ \int_S \Big(\sup_{f\in\mathcal{G}} f(x)\Big) d\mu \,\Big|\, \mathcal{G} \subset \mathcal{F} \text{ countable}\Big\} < \infty.$$

Let $\mathcal{G}_n$ be a sequence of countable subsets of $\mathcal{F}$ approaching the outer supremum, that is, $\int \big(\sup_{f\in\mathcal{G}_n} f(x)\big) d\mu \uparrow c$. Then $\mathcal{G}^* = \cup_n \mathcal{G}_n$ is a countable subset of $\mathcal{F}$ which attains the supremum, that is, $\int \big(\sup_{f\in\mathcal{G}^*} f(x)\big) d\mu = c$. Now let $f^*(x) := \sup_{f\in\mathcal{G}^*}\{f(x)\}$ for every x, and note that f^* is Σ-measurable.

To show this f^* satisfies the requirements of the theorem, observe that if we have $f' \in \mathcal{F}$ with $\mu(\{f' > f^*\}) > 0$ then $\{f'\} \cup \mathcal{G}^*$ is a countable subset of $\mathcal{F}$ and

$$\int_S \Big(\sup_{f\in\{f'\}\cup\mathcal{G}} f(x)\Big) d\mu = \int_S \big(f'(x) \vee f^*(x)\big) d\mu > c$$

giving a contradiction. Furthermore, if g satisfies $g \geq f$ μ-a.e. for all $f \in \mathcal{F}$, then $g(x) \geq \sup_{f\in\mathcal{G}^*} f(x) = f^*$. Finally, if $\mathcal{F}$ is upward directed, then $\mathcal{G}^*$ can be replaced by an increasing sequence of functions, and the result follows.

If the functions are not uniformly bounded, then the monotonic transformation $f(x) \mapsto \arctan(f(x))$ gives a uniformly bounded family. Using this,

$$f^* = \tan(\operatorname{ess\,sup}_{f\in\mathcal{F}}\{\arctan \circ f\})\}$$

gives the essential supremum of the original unbounded family. If μ is not finite but σ-finite, then decomposing into finite sections and constructing the essential supremum on each gives the result. $\square$

1.3.4 Integration for Lebesgue–Stieltjes Measures

For Lebesgue–Stieltjes measures, we can prove further useful properties. Denote by $\int \cdot\, dF_u$ and $\int \cdot\, dG_u$ the integrals with respect to the measures induced by considering F and G as distribution functions.

Lemma 1.3.41. *Let G be a nondecreasing right-continuous function. Then for any dG-integrable function f, the function*

$$F_t = \int_{]0,t]} f(s) dG_s$$

is everywhere right-continuous and has left-limits.

Proof. By linearity of the integral, we can suppose that f is nonnegative. Then F is certainly nondecreasing, and so has left limits. By dominated convergence, for any t,

$$F_{t+\epsilon} = \int_{]0,\infty[} I_{]0,t+\epsilon]} f(s) dG_s \qquad \downarrow \qquad \int_{]0,\infty[} I_{]0,t]} f(s) dG_s = F_t,$$

so F is right-continuous at t. □

We now present a generalization of the well-known method of integration by parts, for the Lebesgue–Stieltjes integral. This result applies only to integrals with respect to Lebesgue–Stieltjes measures, that is, where the measures can be considered through their distribution functions.

We shall see that the presence of jumps in the distribution functions (or equivalently, of atoms in the measures), results in an additional 'quadratic variation' term appearing in the integration formula. This result forms a deterministic precursor to the more general 'differentiation rule' for stochastic integrals that we shall consider in Chapter 14.

Lemma 1.3.42. *Let $F : \mathbb{R} \to \mathbb{R}$ be a right-continuous function. Then F has at most countably many points of discontinuity.*

Proof. Fix $n \in \mathbb{N}$. By right continuity of F, for any point t, there exists an interval $]t, t+\epsilon]$ such that

$$\limsup_{s\to u} F_s - \liminf_{s\to u} F_s < n^{-1} \qquad \text{for } u \in]t, t+\epsilon[.$$

That is, following t, there is an interval within which F has no discontinuities of size n^{-1} or larger. Hence, every discontinuity of F of size n^{-1} or larger can be identified by a rational in the interval following it, and therefore F has at most countably many discontinuities of size n^{-1} or larger. A countable union of countable sets is countable, so taking the union over $n \in \mathbb{N}$ we have that F has at most countably many discontinuities of any size. □

Theorem 1.3.43. *Let $F, G : \mathbb{R} \to \mathbb{R}$, be two nondecreasing right-continuous functions. Then, for any $s < t \in \mathbb{R}$,*

$$F_t G_t - F_s G_s = \int_{]s,t]} F_{u-} dG_u + \int_{]s,t]} G_{u-} dF_u + \sum_{u \in]s,t]} \Delta F_u \Delta G_u.$$

where F_{u-} is the left-limit of F, $\Delta F_u = F_u - F_{u-}$ is the jump of F at u, and similarly for G.

Proof. We first note that, as F and G are right-continuous, F and G can have at most countably many jumps. Hence the nonnegative sum $\sum_{u \in]s,t]} \Delta F_u \Delta G_u$ is well defined.

We define a sequence of finite partitions $\big\{\mathcal{P}^n = \{x_0 = s < x_1 < \ldots < x_{N(n)} = t\}\big\}_{n \in \mathbb{N}}$ such that $\mathcal{P}^n \subseteq \mathcal{P}^{n+1}$, and, for any n, $\max_i \{x_{i+1} - x_i : x_i, x_{i+1} \in \mathcal{P}^n\} < cn^{-1}$, for some fixed $c \in \mathbb{R}$.

Consider an approximation of F, namely

$$F^n(u) = F(x_i), \qquad \text{for} \quad x_i \le u < x_{i+1},$$

and similarly for G. This approximates F with a right-continuous step function. As F is increasing, it is easy to show that $F^n_{u-} \uparrow F_{u-}$ pointwise and $F(x_i) = F^n(x_i)$ for all $x_i \in \mathcal{P}^n$. We also define the incremental process

$$F^{\Delta,n}(u) = F(x_{i+1}) - F(x_i), \qquad \text{for} \quad x_i < u \le x_{i+1}$$

and notice that $F^{\Delta,n}(u) \downarrow \Delta F(u)$ pointwise. A similar argument holds for G.

Now use a telescoping sum to write, for any n,

$$\begin{aligned}
G_t F_t - G_s F_s &= \sum_{i=0}^{N(n)} \big(G(x_{i+1})F(x_{i+1}) - G(x_i)F(x_i)\big) \\
&= \sum_{i=0}^{N(n)} \Big(G(x_i)\big(F(x_{i+1}) - F(x_i)\big) + F(x_i)\big(G(x_{i+1}) - G(x_i)\big) \\
&\qquad + \big(G(x_{i+1}) - G(x_i)\big)\big(F(x_{i+1}) - F(x_i)\big)\Big) \\
&= \sum_{i=0}^{N(n)} G^n(x_i)\big(F(x_{i+1}) - F(x_i)\big) \\
&\qquad + \sum_{i=0}^{N(n)} F^n(x_i)\big(G(x_{i+1}) - G(x_i)\big) \\
&\qquad + \sum_{i=0}^{N(n)} G^{\Delta,n}(x_{i+1})\big(F(x_{i+1}) - F(x_i)\big) \\
&= \int_{]s,t]} G^n_{u-} dF_u + \int_{]s,t]} F^n_{u-} dG_u + \int_{]s,t]} G^{\Delta,n}_u dF_u.
\end{aligned}$$

By the dominated convergence theorem, we then let $n \to \infty$ to obtain

$$G_t F_t - G_s F_s = \int_{]s,t]} G_{u-} dF_u + \int_{]s,t]} F_{u-} dG_u + \int_{]s,t]} \Delta G_u dF_u.$$

As ΔG_u is zero except on a countable set, we can write $\int_{]s,t]} \Delta G_u dF_u = \sum_{u \in]s,t]} \Delta G_u \Delta F_u$ to obtain the desired result. □

Corollary 1.3.44. *For F a nondecreasing right-continuous function and any $s < t$,*

$$\begin{aligned} F_t^2 &= F_s^2 + 2 \int_{]s,t]} F_{s-} dF_s + \sum_{u \in]s,t]} (\Delta F_s)^2 \\ &= F_s^2 + 2 \int_{]s,t]} F_s dF_s - \sum_{u \in]s,t]} (\Delta F_s)^2. \end{aligned}$$

The following result shows how Stieltjes integrals can be defined in terms of a related Lebesgue integral.

Theorem 1.3.45. *Suppose $F : [0,\infty] \to [0,\infty]$ is a nondecreasing, right-continuous function. For $t \in [0,\infty[$ define $\gamma(s) = \inf\{t : F(t) > s\}$. If F is finite, then, for every nonnegative Borel measurable function $f : [0,\infty] \to \overline{\mathbb{R}}$,*

$$\int_{[0,\infty]} f(t)\, dF(t) = \int_{[F(0),F(\infty)]} f(\gamma(t))\, dt = \int_{[F(0),\infty]} I_{\{\gamma<\infty\}}(t) f(\gamma(t))\, dt.$$

Proof. The relationship between F and γ is most easily verified by observing that the graph of γ is obtained from the graph of F by reflection in the diagonal. (If F is continuous and strictly increasing, γ is simply the inverse of F.) Preserving right-continuity, intervals where F is constant correspond to jumps of γ, and jumps of F correspond to intervals where γ is constant.

To establish the identity of the integrals, first consider f of the form $f(t) = I_{[0,s]}(t)$ where $s \in [0,\infty]$. Then

$$\int_{[0,\infty]} I_{[0,s]}(t)\, dF(t) = F(s) - F(0),$$

and $\int_{[F(0),F(\infty)]} I_{[0,s]}(\gamma(t))\, dt$ is the length of the interval $\{t : t \geq F(0)$ and $F(t) \leq s\}$. However, $\sup\{t : \gamma(t) \leq s\} = \inf\{t : \gamma(t) > s\} = F(s)$, so the integral equals $F(s) - F(0)$.

By linearity and the monotone convergence theorem we see that the identity holds for all nonnegative f. □

Note that $\gamma(F(t)) = t$ if and only if γ has no jump at t. Similarly, $F(\gamma(t)) = t$ if and only if F has no jump at t. If F is continuous, and so has no jumps, applying Theorem 1.3.45 to $f(t) = g(F(t))$ we have the following result.

Corollary 1.3.46. *Suppose $F : [0, \infty] \to [0, \infty]$ is a nondecreasing* continuous *function. Then for every nonnegative Borel function g*

$$\int_{[0,\infty]} g(F(t))\, dF(t) = \int_{[F(0),F(\infty)]} g(t)\, dt.$$

Remark 1.3.47. By taking g to be the difference of two nonnegative functions, these results can clearly be extended to every Borel measurable function such that the integral is finite.

1.4 Product Measures

Just as we are used to extending the theory of integration on the real line to integration on $\mathbb{R}^n$, we wish to be able to combine measure spaces and perform integration on them simultaneously.

For this to be practically useful, we need to know when the integral over the product space can be performed iteratively, just as we reduce integration over $\mathbb{R}^n$ to a sequence of integrals, one in each coordinate. For this to work, the order in which we take each of the one-coordinate integrals must not matter. The question of when this can be done is addressed in this section.

Definition 1.4.1. *Let (S_1, Σ_1, μ) and (S_2, Σ_2, ν) be two measure spaces. Then these define a measure space $(S_1 \times S_2, \Sigma_1 \otimes \Sigma_2, \mu \times \nu)$, called the* product space*, where $\Sigma_1 \otimes \Sigma_2$, is the σ-algebra given by*

$$\Sigma_1 \otimes \Sigma_2 := \sigma\{A \times B : A \in \Sigma_1 \text{ and } B \in \Sigma_2\}$$

and $\mu \times \nu$ is the extension (by Theorem 1.2.7 or Theorem A.1.17) to $\Sigma_1 \otimes \Sigma_2$ of the set function

$$(\mu \times \nu)(A \times B) := \mu(A) \cdot \nu(B).$$

Remark 1.4.2. Even if (S_1, Σ_1, μ) and (S_2, Σ_2, ν) are complete measure spaces, this does not guarantee that $(S_1 \times S_2, \Sigma_1 \otimes \Sigma_2, \mu \times \nu)$ is a complete measure space.

We can, of course, still invoke Lemma 1.2.11 to extend $(S_1 \times S_2, \Sigma_1 \otimes \Sigma_2, \mu \times \nu)$ to a complete measure space.

Remark 1.4.3. If μ and ν are σ-finite, then one can show that the product measure $\xi = \mu \times \nu$ defined here (by reference to Theorem 1.2.7) is the only measure on $\Sigma_1 \otimes \Sigma_2$ with the property that $\xi(A \times B) = \mu(A) \cdot \nu(B)$ for all $A \in \Sigma_1$, $B \in \Sigma_2$. If the measures are not σ-finite, then this uniqueness is lost, and it is common to call measures with this property *product measures*. In such a setting, the more general Theorem A.1.17 defines a particular choice of product measure, which is often convenient to work with. This shall typically not concern us, as very few not-σ-finite spaces arise in probability theory (the only example of significance is the counting measure). This and related issues are discussed in detail in Fremlin [82, Chapter 25].

Lemma 1.4.4. *Let $\mathcal{N}$ be the collection of finite disjoint unions of 'measurable rectangles', that is, finite unions of disjoint sets of the form $A \times B$ for $A \in \Sigma_1$, $B \in \Sigma_2$. Then $\mathcal{N}$ is an algebra of sets and $\Sigma_1 \otimes \Sigma_2 = \sigma(\mathcal{N})$.*

Proof. For any sets $\left(\bigcup_i \left(A_i^1 \times B_i^1\right)\right), \left(\bigcup_j \left(A_j^2 \times B_j^2\right)\right) \in \mathcal{N}$, we know

$$\left(\bigcup_i (A_i^1 \times B_i^1)\right) \cap \left(\bigcup_j (A_j^2 \times B_j^2)\right) = \bigcup_{i,j} \left((A_i^1 \cap A_j^2) \times (B_i^1 \cap B_j^2)\right)$$

and

$$\begin{aligned}&\left(\bigcup_i (A_i^1 \times B_i^1)\right) \cup \left(\bigcup_j (A_j^2 \times B_j^2)\right) \\ &\quad = \bigcup_{i,j} \Big(((A_i^1 \setminus A_j^2) \times B_i^1) \cup ((A_i^1 \cap A_j^2) \times (B_i^1 \cup B_j^2)) \cup ((A_j^2 \setminus A_i^1) \times B_j^2)\Big)\end{aligned}$$

which are disjoint unions of rectangles, and so $\mathcal{N}$ is closed under finite intersections and unions. As

$$\left(\bigcup_i A_i \times B_i\right)^c = \bigcap_i \left((S_1 \times B_i^c) \cup (A_i^c \times B_i)\right),$$

which is a finite intersection of unions of disjoint rectangles, we see that $\mathcal{N}$ is closed under complementation. That $\Sigma_1 \otimes \Sigma_2 = \sigma(\mathcal{N})$ is direct from the definition of $\Sigma_1 \otimes \Sigma_2$. □

Theorem 1.4.5 (Fubini's Theorem). *Let f be an integrable function on $(S_1 \times S_2, \Sigma_1 \otimes \Sigma_2, \mu \times \nu)$. Then*

(i) For ν-almost all s_2, the function $f_{s_2}(s_1) := f(s_1, s_2)$ is an integrable function on (S_1, Σ_1, μ),
(ii) The function

$$F(s_2) := \int_{S_1} f_{s_2}(s_1)\, d\mu(s_1)$$

is an integrable function on (S_2, Σ_2, ν),
(iii)

$$\int_{S_2} F\, d\nu = \int_{S_1 \times S_2} f\, d(\mu \times \nu),$$

and conversely with the roles of s_1 and s_2 reversed.

Proof. We shall prove this using the monotone class theorem.

As f is integrable, and the integral is defined by the difference of nonnegative functions, it is sufficient to prove this under the further assumption that f is nonnegative. Furthermore, as f is integrable, by the monotone convergence theorem it is sufficient to prove this under the assumption that f is simple, that is, $f = \sum_i x_i I_{C_i}$, and that $\mu(C_i) < \infty$ for all i. By linearity, it is then sufficient to prove this under the assumption $f = I_C$ for some $C \in \Sigma_1 \otimes \Sigma_2$.

If C is a measurable rectangle $C = A \times B$, then $F(s_2) = \mu(A) I_B(s_2)$, and the result is straightforward. If C is a finite disjoint union of measurable rectangles, then $F(s_2) = \sum_i \mu(A_i) I_{B_i}(s_2)$, and the result follows from linearity.

Let $\mathcal{M}$ be the class of sets C such that the theorem holds with $f = I_C$. We have just shown that $\mathcal{M}$ contains the collection $\mathcal{N}$ of all finite disjoint unions of measurable rectangles, which is an algebra by Lemma 1.4.4.

If $C = \bigcup_{i=0}^{\infty} C_i$ for a nondecreasing sequence $C_i \in \mathcal{M}$, then we can verify that $C \in \mathcal{M}$ by the monotone convergence theorem. Similarly if $C = \bigcap_{i=0}^{\infty} C_i$ for a nonincreasing sequence $C_i \in \mathcal{N}$, then $C_1 \setminus C_i$ is a nondecreasing sequence in $\mathcal{M}$, hence, again by the monotone convergence theorem, $C_1 \setminus C \in \mathcal{M}$ and it follows $C \in \mathcal{M}$. Hence $\mathcal{M}$ is a monotone class containing $\mathcal{N}$.

Therefore, by the monotone class theorem, $\mathcal{M}$ contains $\sigma(\mathcal{N}) = \Sigma_1 \otimes \Sigma_2$, and the result holds in general. □

Theorem 1.4.6 (Tonelli's Theorem). *Let f be a nonnegative function on $(S_1 \times S_2, \Sigma_1 \otimes \Sigma_2, \mu \times \nu)$. Assume (S_1, Σ_1, μ) and (S_2, Σ_2, ν) are both σ-finite measure spaces. Then*

(i) For ν-almost all s_2, the function $f_{s_2}(s_1) := f(s_1, s_2)$ is a measurable function on (S_1, Σ_1, μ),

(ii) The function

$$F(s_2) := \int_{S_1} f_{s_2}(s_1)\, d\mu$$

is a measurable function on (S_2, Σ_2, ν),

(iii)

$$\int_{S_2} F\, d\nu = \int_{S_1 \times S_2} f\, d(\mu \times \nu),$$

and conversely with the roles of s_1 and s_2 reversed.

Proof. The proof follows almost exactly as for Fubini's theorem. As before, we can assume f is simple, that is $f = \sum_i x_i I_{C_i}$. However, we do not know f is integrable. Therefore we must use the fact that our spaces are σ-finite, and hence the product space is σ-finite, to show that $\mu(C_i) < \infty$, without loss of generality (simply decompose the space into finite sections, and take C_i to lie within a single section for each i). The remainder of the proof is as before. □

Remark 1.4.7. The key differences between these theorems is that Fubini's theorem requires us to check that f is integrable *before* attempting to perform the integral iteratively. Tonelli's theorem on the other hand only assumes f is nonnegative and the spaces are σ-finite, which will often allow us to integrate $|f|$, thereby checking the required assumptions for Fubini's theorem.

The assumption of σ-finiteness is crucial to this theorem, and, particularly when dealing with certain stochastic processes with jumps, may not hold.

Remark 1.4.8. We have given these results on the product space $(S_1 \times S_2, \Sigma_1 \otimes \Sigma_2, \mu \times \nu)$. However, one can clearly iterate this result, giving a product measure on $S_1 \times S_2 \times \ldots \times S_N$ for any finite N.

For probability spaces, we shall see that $\mu(A) \leq 1$ for all A. Hence, given a sequence of spaces $\{(S_i, \Sigma_i, \mu_i)\}_{i \in \mathbb{N}}$, the countable product

$$\mu(A_1 \times A_2 \times \ldots) := \prod_{i=1}^{\infty} \mu_i(A_i)$$

will converge, and so a similar method can be used to construct a measure $\mu = \prod_i \mu_i$ on the countable product space $\left(\prod_i S_i, \bigotimes_i \Sigma_i\right)$.

1.5 Linear, Banach, Hilbert and L^p Spaces

The previous discussion of integration theory, particularly the discussion of pointwise limits of functions, motivates a discussion of spaces of functions and limits of functions in a consistent and general way. For this reason, we wish to define topologies on the space of functions, and to study relations between this topology and the integral.

The most common way to do this is through the study of L^p spaces. These are spaces of functions with certain boundedness properties, and the study of limits of these functions is fairly natural.

We begin with a review of common definitions from real analysis.

Definition 1.5.1. *A* real vector space *is a collection X of objects where for any $a, b \in \mathbb{R}$, any $f, g, h \in X$,*

(i) $(X, +)$ is an Abelian group, that is, we can perform addition in X and this satisfies the usual closure, associativity and commutativity properties, and an additive identity (denoted 0) and additive inverses (denoted $-x$) all exist,
(ii) scalar multiplication by real numbers is well defined and satisfies the usual distributive properties.

A vector space with a topology is called a topological vector space*, provided addition and scalar multiplication are continuous.*

A norm *on X is a function $\|\cdot\| : X \to [0, \infty[$ such that for $f, g \in X$ and $a \in \mathbb{R}$,*

(i) $\|f + g\| \leq \|f\| + \|g\|$ (Subadditivity),
(ii) $\|af\| = |a|\|f\|$ (Homogeneity),
(iii) $\|f\| = 0$ implies $f = 0$ (Faithfulness).

The pair $(X, \|\cdot\|)$ is called a normed vector space*. A function $\|\cdot\|$ satisfying at least properties (i) and (ii) above is called a* seminorm*. Any seminorm on a vector space X defines a vector space $\hat{X}$ of equivalence classes, where $x, y \in X$*

correspond to the same element of $\hat{X}$ *whenever* $\|x-y\|=0$. *On the space* $\hat{X}$, *the function* $\|\cdot\|$ *is a true norm.*

A metric *on a space (which may or may not be a vector space) is a function* $d : X \times X \to [0,\infty[$ *where*

(i) $d(f,g) = d(g,f)$,
(ii) $d(f,g) = 0$ *implies* $f = g$,
(iii) $d(f,g) \le d(f,h) + d(h,g)$ *(the Triangle inequality).*

Every normed vector space $(X,\|\cdot\|)$ *is a* metric space (X,d) *where* $d(x,y) = \|x-y\|$. *A metric over a space induces a topology on that space, with open sets given by unions of sets of the form* $\{f \in X : d(f,g) < \epsilon\}$, *the radius* ϵ *balls around points* $g \in X$.

Definition 1.5.2. *A sequence of elements* $\{f_n\}_{n\in\mathbb{N}}$ *is called a* Cauchy sequence *if, for every* $\epsilon > 0$, *there exists* $N \in \mathbb{N}$ *such that* $d(f_m, f_n) < \epsilon$ *for all* $m, n > N$.

A metric space is called complete *if every* Cauchy sequence *converges, that is, for each Cauchy sequence* $\{f_n\}_{n\in\mathbb{N}}$ *there exists an element* f *such that, for all* $\epsilon > 0$, *there exists* $N \in \mathbb{N}$ *with* $d(f, f_m) < \epsilon$ *for all* $m > N$. *We write* $f = \lim_n f_n$.

A normed vector space which is complete with respect to the metric arising from the norm is called a Banach space.

Definition 1.5.3. *Let* X *be a vector space and* $\|\cdot\|$ *and* $\|\cdot\|^*$ *two norms on* X. *We say that the norms are equivalent if there exist constants* $c, c' \ge 0$ *such that*

$$c\|x\| \le \|x\|^* \le c'\|x\|.$$

Lemma 1.5.4. *Equivalent norms generate the same topology.*

Proof. It is enough to show that the ball $B(0,\alpha) = \{x : \|x\| \le \alpha\}$ can be written as the union of balls $B^*(y,\beta) = \{x : \|x-y\|^* \le \beta\}$, as linear shifts and exchanging the role of $\|\cdot\|$ and $\|\cdot\|^*$ then shows that the topologies are the same. For any $y \in B(0,\alpha)$, as $c\|x\| \le \|x\|^*$ we have $\{x : \|x-y\|^* < c(\alpha - \|y\|)\} \subseteq B(0,\alpha)$. Therefore, we can write $B(0,\alpha) = \bigcup_{y\in B(0,\alpha)} B^*(y, c(\alpha - \|y\|))$, as desired. □

Definition 1.5.5. *A set* K *in a topological space is*

(i) compact *(or Heine–Borel compact) if every cover of* K *by open sets admits a finite subcover, that is, if for every collection* $\{H_a\}_{a\in\mathcal{A}}$ *of open sets with* $K \subseteq \bigcup_{a\in\mathcal{A}} H_a$, *there is a finite collection* $\{H_{a_i}\}_{i=1}^m$ *with* $K \subseteq \bigcup_{i=1}^m H_{a_i}$,
(ii) sequentially compact *if every sequence in* K *has a convergent subsequence, that is, for any sequence* $\{k_n\}_{n\in\mathbb{N}} \subseteq K$ *there exists a point* $k \in K$ *such that* $\{k_n\}_{n\in\mathbb{N}} \cap N \ne \emptyset$ *for* N *any open set containing* k,
(iii) limit point compact *if every infinite subset of* K *has a limit point in* K, *that is, a point* $k \in K$ *such that every open set containing* k *contains infinitely many points in* K.

A set is said to be relatively compact if its closure is compact, and similarly for sequentially and limit point compactness.

Remark 1.5.6. For a metric space, the three notions of compactness in Definition 1.5.5 agree. The Heine–Borel theorem states that a set in $\mathbb{R}$ is compact if and only if it is closed and bounded. See Royden and Fitzpatrick [160] for details.

Definition 1.5.7. *If X and Y are both normed vector spaces, then a function $T : X \to Y$ is called an* operator*. If $Y = \mathbb{R}$, then T is called a (real)* functional*.*

If T is a linear operator *(that is, $T(af+bg) = aT(f)+bT(g)$ for all $f, g \in X$, $a, b \in \mathbb{R}$) and* bounded *(that is, for some $c \in \mathbb{R}$, $\|T(x)\|_Y \leq c\|x\|_X$ for all $x \in X$), then we say that T is a* bounded linear operator*. If $\|T(x)\|_Y = \|x\|_X$, we say T is an isometry.*

The set of all bounded linear operators from X to Y is denoted $B(X, Y)$. If $Y = \mathbb{R}$, then $X' = B(X, \mathbb{R})$ is the set of all bounded linear functionals and called the (topological) dual space *of X.*

Remark 1.5.8. This definition can be somewhat confusing, as a bounded linear functional does not map X to a bounded subset of Y. The name comes because we can define a norm on the space of linear operators by $\|T\|_{op} = \sup_x\{\|T(x)\|/\|x\|\}$ (called the *operator norm*) and under this norm, a bounded linear functional has $\|T\|_{op} < \infty$.

A slightly less trivial, but very useful, result from the theory of Banach spaces is the following.

Lemma 1.5.9. *Let X be a Banach space, and suppose X can be written $X = Y \oplus Z$, (that is, $X = Y + Z$ and $Y \cap Z = \{0\}$). Then, writing $x = y + z$, the map*

$$x \mapsto \|x\|_\oplus := \|y\|_X + \|z\|_X$$

is a norm on X and is equivalent to $\|\cdot\|_X$.

Proof. It is easy to check that $\|x\|_\oplus$ is a norm. Clearly $\|x\|_X \leq \|x\|_\oplus$, by the triangle inequality. Conversely, we see that the map $T : x \to (y, z)$ is a bounded linear operator and is bijective. Therefore, by the bounded inverse theorem (see, for example, Royden and Fitzpatrick [160, p.265]) we know that T^{-1} is also a bounded linear operator, and so there exists $C > 0$ such that

$$\|x\|_X = \|T^{-1}(y, z)\| \leq C(\|y\|_X + \|z\|_X). \qquad \square$$

Lemma 1.5.10. *Let $F : X \to X$ be a bounded linear operator. Then, on the dual space X' there exists a unique bounded linear operator $F^* : X' \to X'$, called the* adjoint *of F, such that*

$$(g \circ F)(f) = (F^* \circ g)(f)$$

for all $f \in X$ and $g \in X'$. Furthermore, if F is a projection (that is, $(F \circ F)(f)) = F(f)$ for all $f \in X$), then

$$(g \circ F)(f) = (F^* \circ g)(f) = (F^* \circ g)(F(f)).$$

Proof. Simply define $F^* : g \mapsto g \circ F$. Each of the stated properties (boundedness, linearity, uniqueness and the projection property) can then be verified directly. □

The following follows directly from the definition.

Lemma 1.5.11. *A bounded linear functional is (Lipschitz) continuous in the norm topology, that is, $\|T(x) - T(y)\| = \|T(x - y)\| \leq c\|x - y\|$.*

See [160, p.275] for a more in-depth exploration of the following concept.

Definition 1.5.12. *The* weak topology *is the smallest topology on X (that is, collection of subsets of X which we call 'open') such that all bounded linear functionals are continuous.*

If $x_n \to x$ in the norm topology (that is, $\|x_n - x\| \to 0$), then $x_n \to x$ in the weak topology, but not vice versa.

The following results can be found in [160, p.278 and p.292], and are fundamental to much of functional analysis.

Theorem 1.5.13 (Hahn–Banach Theorem). *Let ϕ be a positively homogeneous, subadditive functional (i.e. $\phi(\lambda x) = \lambda\phi(x)$ and $\phi(x+y) \leq \phi(x)+\phi(y)$ for all $\lambda > 0$, $x, y \in X$) on a vector space X, and Y a subspace of X on which there is defined a linear functional ψ for which $\psi \leq \phi$ on Y. Then ψ may be extended to a linear functional ψ on all of X, for which $\psi \leq \phi$ on all of X.*

Corollary 1.5.14. *Let X be a Banach space and let Y be a linear subspace with closure $\bar{Y}$. Then $\bar{Y} = X$ if and only if the only bounded linear functional ψ such that $\psi(y) = 0$ for all $y \in Y$ is $\psi \equiv 0$ (i.e. $\psi(x) = 0$ for all $x \in X$).*

Lemma 1.5.15 (Mazur's Lemma).

(i) Let K be a convex subset of a normed vector space X. Then K is strongly closed (i.e. in the norm topology) if and only if it is weakly closed.
(ii) Let $\{x_n\}_{n\in\mathbb{N}}$ be a weakly convergent sequence in a normed vector space X, with limit $x \in X$. Then there is a sequence $\{z_n\}_{n\in\mathbb{N}}$ which converges strongly to x, and each z_n is a convex combination of $\{x_n, x_{n+1}, \ldots\}$.

Lemma 1.5.16 (Eberlein–Šmulian Theorem). *For the weak topology on a Banach space, the three notions of compactness in Definition 1.5.5 agree.*

For the study of equations, it is often convenient to define objects in terms of fixed points of functionals. The most basic construction is as follows.

Definition 1.5.17. *A functional $F : X \to X$ (not necessarily linear) is called a* contraction *if for any $x, x' \in X$, we know $\|F(x) - F(x')\| \le c\|x - x'\|$ for some $c \in [0, 1[$.*

Lemma 1.5.18. *A contraction in a Banach space has a unique fixed point, that is, there is a unique x such that $F(x) = x$. This fixed point satisfies $\|x\| \le \frac{1}{1-c}\|F(0)\|$.*

Proof. Write F^n for the n-fold application of F, that is $F^3 = F \circ F \circ F$, etc. Then, for any x, we have $\|F^n(x) - F^m(x)\| \le c^{n\vee m}\|x\|$, so $F^n(x)$ is a Cauchy sequence. Therefore, there is a limit $F^n(x) \to y$, and for any $x \in X$, $\|F^n(x) - y\| \to 0$. In particular, this implies $F(y) = y$ and that the fixed point must be unique. To show the stated bound, consider the sequence $F^n(0) \to y$, with $F^0(0) := 0$. Then

$$\begin{aligned}\|y\| &= \Big\|\sum_{i=0}^{\infty}(F^{i+1}(0) - F^i(0))\Big\| \le \sum_{i=0}^{\infty}\|F^{i+1}(0) - F^i(0)\| \\ &\le \sum_{i=0}^{\infty} c^i\|F(0)\| \le \frac{1}{1-c}\|F(0)\|.\end{aligned}$$

□

Definition 1.5.19. *For a real vector space X, an inner product is a mapping $\langle\cdot,\cdot\rangle : X \times X \to \mathbb{R}$ such that for $f, g, h \in X$, $a \in \mathbb{R}$*

(i) $\langle f + g, h\rangle = \langle f, h\rangle + \langle g, h\rangle$
(ii) $\langle af, g\rangle = a\langle f, g\rangle$
(iii) $\langle f, g\rangle = \langle g, f\rangle$
(iv) $\langle f, f\rangle \ge 0$ *with* $\langle f, f\rangle = 0$ *if and only if* $f = 0$.

A vector space X equipped with an inner product is called an inner product space*. It is easy to check that we may define a norm on X by $\|x\| = \sqrt{\langle x, x\rangle}$ for all $x \in X$.*

A complete normed vector space $(X, \|\cdot\|)$ in which the norm is defined by a given inner product is called a Hilbert space*.*

Example 1.5.20. Consider the space of infinite real sequences $x = (x_1, x_2, \ldots)$. Then we can define the inner product $\langle x, y\rangle = \sum_i x_i y_i$, and the sequences x with $\langle x, x\rangle < \infty$ form a Hilbert space, commonly known as ℓ^2.

For $p \ge 1$, we can also define a norm $\|x\|_{\ell^p} = \big(\sum_i |x_i|^p\big)^{1/p}$, and the set of sequences with $\|x\|_{\ell^p} < \infty$ form a Banach space (known as ℓ^p).

We cite the following results without proof (see [160, pp.309–313]):

Lemma 1.5.21. *Let A be a closed convex subset of a Hilbert space H. Then there exists a continuous map $\Pi_A : H \to A$, called the orthogonal projection on A, given by*

$$\Pi_A(x) = \arg\min_{a\in A} \|a - x\|.$$

If A is also a vector subspace of H, then $\Pi_A(x)$ and $x - \Pi_A(x)$ are orthogonal, that is $\langle \Pi_A(x), x - \Pi_A(x)\rangle = 0$. Consequently, we can write $H = A \oplus A^{\perp}$, where $A^{\perp}$ is the set of vectors orthogonal to all elements of A.

Theorem 1.5.22 (Riesz–Fréchet Representation Theorem). *Let F be a bounded linear functional on a Hilbert space H. Then there exists a unique $g \in H$ such that $F(f) = \langle f, g\rangle$ for all $f \in H$.*

1.5.1 Spaces of Functions

We will now consider vector spaces of measurable functions.

Definition 1.5.23. *Let $\mathcal{L}^0(S, \Sigma, \mu)$ denote the space of measurable functions, that is,*

$$\mathcal{L}^0(S, \Sigma, \mu) := \{f : S \to \overline{\mathbb{R}} : f^{-1}(A) \in \Sigma \text{ for all } A \in \mathcal{B}([0, \infty])\}.$$

Definition 1.5.24. *Given $f \in \mathcal{L}^0(S, \Sigma, \mu)$ and $p \in\,]0, \infty[$ we define a functional $\|\cdot\|_p$ by*

$$\|f\|_p := \left(\int_S |f|^p \, d\mu\right)^{1/p}.$$

We define $\|\cdot\|_\infty$ by

$$\|f\|_\infty = \operatorname{ess\,sup}_{s\in S}\{|f(s)|\} := \inf\{M \in [0, \infty] : |f(s)| \leq M \text{ a.e.}\}.$$

Definition 1.5.25. *For $p \in\,]0, \infty]$, we define $\mathcal{L}^p$ by*

$$\mathcal{L}^p(S, \Sigma, \mu) = \{f \in \mathcal{L}^0(S, \Sigma, \mu) : \|f\|_p < \infty\}.$$

For $p \in\,]0, \infty]$, $\mathcal{L}^p(S, \Sigma, \mu)$ is a linear subspace of $\mathcal{L}^0(S, \Sigma, \mu)$. In general, for $p \in\,]0, \infty]$, $\|\cdot\|_p$ is not a norm on $\mathcal{L}^p(S, \Sigma, \mu)$. However for $p \in [1, \infty]$, we shall see that $\|\cdot\|_p$ is a seminorm on $\mathcal{L}^p(S, \Sigma, \mu)$.

Lemma 1.5.26 (Young's Inequality). *For $a, b > 0$, if $p^{-1} + q^{-1} = 1$ then*

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q},$$

with equality if and only if $a^p = b^q$.

Proof. Clearly

$$\ln ab = \frac{\ln(a^p)}{p} + \frac{\ln(b^q)}{q}.$$

Then, as $\exp(x)$ is convex, for any $\lambda \in [0, 1]$,

$$ab = \exp\left(\lambda \ln(a^p) + (1 - \lambda)\ln(b^q)\right) \leq \lambda \exp\ln(a^p) + (1 - \lambda)\exp\ln(b^q).$$

For $\lambda = p^{-1} = 1 - q^{-1}$, this gives the desired result. As $\exp(x)$ is strictly convex, this is an equality if and only if it is independent of λ, that is, if $\ln(a^p) = \ln(b^q)$. $\square$

Theorem 1.5.27 (Hölder's Inequality). *Suppose that $f \in \mathcal{L}^p$, $g \in \mathcal{L}^q$, where $p^{-1} + q^{-1} = 1$ (p and q are 'Hölder conjugates') and $p \in [1, \infty]$. Then*

$$\int_S fg \, d\mu \leq \int_S |fg| \, d\mu \leq \|f\|_p \cdot \|g\|_q.$$

Furthermore, equality holds if and only if $|f|^p = k|g|^q$ a.e. for some $k > 0$.

Proof. The first inequality is clear, so we restrict ourselves to the case where $f \geq 0$, $g \geq 0$. The case $p = 1$ or $p = \infty$ can be shown directly from

$$\int_S fg \, d\mu \leq \int_S f \cdot (\operatorname{ess\,sup} g) \, d\mu = (\operatorname{ess\,sup} g) \cdot \int_S f \, d\mu = \|g\|_\infty \cdot \|f\|_1.$$

For $p \in]1, \infty[$, if $\|f\|_p = 0$ or $\|g\|_q = 0$ then $f = 0$ a.e. or $g = 0$ a.e., and so the inequality is trivial. Define $\tilde{f} = f/\|f\|_p$ and $\tilde{g} = g/\|g\|_q$. Then by Young's inequality,

$$\tilde{f}\tilde{g} \leq p^{-1}\tilde{f}^p + q^{-1}\tilde{g}^q.$$

integrating both sides gives

$$\int_S \tilde{f}\tilde{g} \, d\mu \leq p^{-1} \int_S \tilde{f}^p \, d\mu + q^{-1} \int_S \tilde{g}^q \, d\mu = p^{-1} + q^{-1} = 1$$

and multiplying both sides by $\|f\|_p\|g\|_q$ gives the desired result.

From Young's inequality, equality can only be preserved if $\tilde{f}^p = \tilde{g}^q$ a.e., that is, if $|f|^p = \frac{\|f\|_p}{\|g\|_q}|g|^q$ almost everywhere. □

Remark 1.5.28. Taking $p = q = 2$ in Hölder's inequality gives the Cauchy–Schwarz inequality, $|\langle f, g \rangle| \leq \|f\|_2 \cdot \|g\|_2$, where the inner product is given by the integral of the product of the functions. (This is a true inner product in L^2, which we shall define shortly.)

Theorem 1.5.29 (Minkowski's Inequality). *Suppose that $f, g \in \mathcal{L}^p$ for some $p \geq 1$. Then $f + g \in \mathcal{L}^p$ and*

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p.$$

Proof. The cases $p = 1$ and $p = \infty$ are trivial.

For $p \in]1, \infty[$, let q be the Hölder conjugate of p, so $q = (1-p^{-1})^{-1} = \frac{p}{p-1}$. Then

$$\begin{aligned} \big(\|f + g\|_p\big)^p &\leq \int_S \big(|f| + |g|\big)|f + g|^{p-1} \, d\mu \\ &= \int_S |f||f + g|^{p-1} \, d\mu + \int_S |g||f + g|^{p-1} \, d\mu. \end{aligned}$$

As

$$\left\| |f+g|^{p-1} \right\|_q = \left(\int_S |f+g|^{(p-1)q} \, d\mu \right)^{1/q}$$
$$= \left(\int_S |f+g|^p \, d\mu \right)^{\frac{p-1}{p}} = (\|f+g\|_p)^{p-1},$$

we have $|f+g|^{p-1} \in \mathcal{L}^q$. So by Hölder's inequality,

$$(\|f+g\|_p)^p \leq \|f\|_p \left\| |f+g|^{p-1} \right\|_q + \|g\|_p \left\| |f+g|^{p-1} \right\|_q$$
$$= (\|f+g\|_p)^{p-1}(\|f\|_p + \|g\|_p).$$

Division by $(\|f+g\|_p)^{p-1}$ yields the desired result. $\square$

Remark 1.5.30. Minkowski's inequality shows that $\|\cdot\|_p$ is a seminorm over $\mathcal{L}^p$ (as homogeneity is trivial). It is not a norm, however, as $\|f-g\|_p = 0$ if and only if $f = g$ μ-a.e. (rather than $f = g$ everywhere). This motivates the following definition.

Definition 1.5.31. *For $p \in [1, \infty]$, let L^p denote the space of equivalence classes in $\mathcal{L}^p$, under the equivalence relation $f = g$ a.e.*

If needed to avoid confusion, we shall write $L^p(S, \Sigma, \mu)$, or simply $L^p(\Sigma)$ or $L^p(\mu)$. We can see that the spaces ℓ_p defined earlier agree with the space $L^p(\mathbb{N})$, under the counting measure.

Remark 1.5.32. For functions in $\mathcal{L}^p$, it is clear that they will take the values ∞ and $-\infty$ only on sets of measure zero. For this reason, there is little need to distinguish between the L^p theory for $\mathbb{R}$ and $\overline{\mathbb{R}}$.

It is clear that L^p is a normed vector space, with norm $\|\cdot\|_{L^p(S,\Sigma,\mu)} = \|\cdot\|_p$. The following theorem shows that it is also complete.

Theorem 1.5.33 (Riesz–Fisher Theorem). *For $p \in [1, \infty]$, L^p is a complete vector space, that is, if $\{f_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in the norm $\|\cdot\|_p$, then there is a function $f \in L^p$ with $\|f_n - f\|_p \to 0$, and we say $f_n \to f$ in L^p.*

Proof. The case $p = \infty$ is left as an exercise.

For $p < \infty$, let $\{f_n\}_{n\in\mathbb{N}}$ be a Cauchy sequence in L^p. Then there exists a subsequence $\{f_{n_k}\}_{k\in\mathbb{N}}$ such that $\|f_{n_{k+1}} - f_{n_k}\|_p \leq 2^{-k}$. Define a function g as the pointwise limit

$$g(s) = \sum_{k=1}^{\infty} |f_{n_{k+1}}(s) - f_{n_k}(s)|.$$

As g is the pointwise limit of measurable functions, it is itself measurable and is clearly nonnegative. Furthermore, by Minkowski's inequality,

$$\Big\| \sum_{k=1}^{N} |f_{n_{k+1}} - f_{n_k}| \Big\|_p \leq \sum_{k=1}^{N} 2^{-k} < 1,$$

so by the monotone convergence theorem, $g \in L^p$ and $\|g\|_p \leq 1$. It is then clear that $g \neq \infty$ a.e. Therefore, for almost all s, the sequence $\{f_{n_k}(s)\}_{k\in\mathbb{N}}$ is convergent.

Define

$$f(s) = \begin{cases} f_{n_1}(s) + \sum_{k=1}^{\infty}(f_{n_{k+1}}(s) - f_{n_k}(s)) & \text{when convergent,} \\ 0 & \text{otherwise.} \end{cases}$$

Then $f \in L^p$, and $f(s) - f_{n_N}(s) = \sum_{k=N}^{\infty}(f_{n_{k+1}}(s) - f_{n_k}(s))$. Therefore

$$\|f - f_{n_N}\|_p \leq \sum_{k=N}^{\infty} \|f_{n_{k+1}} - f_{n_k}\|_p \leq \sum_{k=N}^{\infty} 2^{-k} \leq 2^{1-N} \to 0,$$

i.e. f_{n_k} converges to f in L^p.

As a Cauchy sequence converges if a subsequence converges, we have shown that $f_n \to f$ in L^p, hence that L^p is complete. □

Remark 1.5.34. An important special case is L^2. As mentioned earlier, we can define $\|\cdot\|_2$ through an inner product

$$\langle f, g\rangle = \int_S fg\, d\mu.$$

It is clear that L^2 is a Hilbert space. (In fact, it is the only L^p space which is a Hilbert space.)

Theorem 1.5.35. *Let (S, Σ, μ) be a finite measure space, that is $\mu(S) < \infty$. Then $L^p \subseteq L^{p'}$ for all $p \geq p'$.*

Proof. Suppose $f \in L^p$, that is,

$$\int_S |f|^p\, d\mu < \infty$$

As $p \geq p'$ we know $(|f| \wedge 1)^{p'} \leq (|f| \wedge 1)^p \leq (|f| + 1)^p$. As $\mu(S) = \int_S 1\, d\mu < \infty$, we know that $1 \in L^p$ and hence $|f| + 1 \in L^p$. Therefore,

$$\int_S |f|^{p'}\, d\mu \leq \int_S (|f| \wedge 1)^{p'}\, d\mu \leq \int_S (|f| \wedge 1)^p\, d\mu \leq \int_S (|f| + 1)^p\, d\mu < \infty.$$

□

Remark 1.5.36. The requirement in this theorem that (S, Σ, μ) is a *finite* measure space cannot be avoided. For example, on $[1, \infty[$ with Lebesgue measure, if $f(x) = x^{-1}$ then it is easy to see that $f \in L^2$ but $f \notin L^1$.

1.6 The Radon–Nikodym Theorem

When working with measure spaces, it would be nice if there were a simple way of converting from one measure to another. The Radon–Nikodym theorem allows us to do this. It will also allow us to establish the existence of conditional expectations, which will be useful later.

Some examples of the Radon–Nikodym derivative are familiar. A classic example is, for a continuous random variable, the probability density function (pdf), which is the Radon–Nikodym derivative of the probability measure with respect to Lebesgue measure. Using these densities, we are able to make statements about probabilities by doing classical integration over the real line.

Definition 1.6.1. *Let (S, Σ) be a measure space. Let μ and ν be measures on (S, Σ). The measure ν is said to be* absolutely continuous *with respect to μ, denoted $\nu \ll \mu$, if $\mu(A) = 0$ implies $\nu(A) = 0$, for all $A \in \Sigma$.*

If $\mu \ll \nu$ and $\nu \ll \mu$, then μ and ν are said to be equivalent measures.

The motivation for this terminology is given in the following lemma.

Lemma 1.6.2. *Let μ, ν be measures on (S, Σ) with $\nu(S) < \infty$. Then the following statements are equivalent.*

(i) $\nu \ll \mu$
(ii) for every $\epsilon > 0$ there exists $\delta > 0$ such that $\nu(A) < \epsilon$ whenever $\mu(A) < \delta$.

Proof. To show (ii) implies (i) is straightforward. To show (i) implies (ii), assume (ii) is false, then, for some $\epsilon > 0$, there exist sets $B_n \in \Sigma$ with $\mu(B_n) < 2^{-n}$ and $\nu(B_n) \geq \epsilon$ for all n. If $A_m = \bigcup_{n \geq m} B_n$ and $A = \bigcap_m A_m$, we have $\mu(A_m) < 2^{-m+1}$ and $A_{m+1} \subset A_m$, so $\mu(A) = 0$; however, $\nu(A_m) \geq \epsilon$ so $\nu(A) \geq \epsilon$ (by Exercise 1.8.3, as $\nu(S) < \infty$). We then see that ν is not absolutely continuous with respect to μ. □

Theorem 1.6.3 (Radon–Nikodym Theorem). *Let μ and ν be σ-finite measures on a measure space (S, Σ) and $\nu \ll \mu$. Then there exists a non-negative measurable function $f : S \to [0, \infty]$ such that*

$$\nu(A) = \int_A f \, d\mu \quad \text{for all } A \in \Sigma.$$

In this case, f is called the Radon–Nikodym derivative *of ν with respect to μ, and is sometimes written $d\nu/d\mu$. Furthermore f is unique up to equality μ-a.e.*

Proof. Assume first that $\mu(S) < \infty$, that is, μ is a finite measure on S. Then let $\lambda = \mu + \nu$. Note $\nu \ll \mu \ll \lambda$. For any $g \in L^2(S, \Sigma, \lambda)$, we know

$$\int_S |g|^2 \, d\lambda = \int_S |g|^2 \, d\mu + \int_S |g|^2 \, d\nu$$

and so $g \in L^2(S, \Sigma, \mu) \cap L^2(S, \Sigma, \nu)$. By the Cauchy–Schwarz inequality,

$$\int_S g\, d\mu \leq \int_S 1|g|\, d\mu \leq \left(\int_S 1\, d\mu\right)^{1/2} \left(\int_S |g|^2\, d\mu\right)^{1/2}$$
$$= \mu(S)^{1/2}\|g\|_{L^2(\mu)} \leq \mu(S)^{1/2}\|g\|_{L^2(\lambda)}.$$

Therefore, the function $F : L^2(\lambda) \to \mathbb{R}, g \mapsto \int_S g\, d\mu$ is a bounded linear functional on $L^2(\lambda)$. As $L^2(\lambda)$ is a Hilbert space, from Theorem 1.5.22 there exists a function h such that $F(g) = \langle g, h\rangle = \int_S (gh)\, d\lambda$, for all $g \in L^2(\lambda)$.

As $\int_S g\, d\mu = \int_S gh\, d\lambda$, if we let $g = I_A$ for $A \in \Sigma$, we see that $h > 0$ λ-a.e. Also, $0 \leq \nu(A) = \lambda(A) - \mu(A) = \int_A (1-h)\, d\lambda$, and so $0 < h \leq 1$ λ-a.e. Define

$$f = \begin{cases} \frac{1-h}{h} & \text{if } h \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Then f is a measurable function (as h is measurable) and

$$\nu(A) = \int_A (1-h)\, d\lambda = \int_A \frac{1-h}{h} h\, d\lambda = \int_A f\, d\mu$$

as desired.

For the σ-finite case, write $S = \bigcup_{i\in\mathbb{N}} S_i$, where $\mu(S_i) < \infty$ and $\nu(S_i) < \infty$, and $S_i \cap S_j = \emptyset$ for $i \neq j$. Then define f_i as before on each S_i, and take $f = \sum_i f_i$.

Finally, if f and $\tilde{f}$ both satisfy the desired equation, then let $A = \{s : f > \tilde{f}\}$, and note that

$$\int_A f\, d\mu = \int_A \tilde{f}\, d\mu$$

which implies that A is of measure zero. Similarly for $\tilde{f} < f$. Hence $f = \tilde{f}$ μ-a.e. □

Example 1.6.4. A classic simple example of the Radon–Nikodym derivative is the density of a probability distribution. For example, consider a random variable X with a standard normal distribution. Then, to calculate the expected value of $g(X)$, for g a Borel-measurable function, we calculate

$$\int_{\mathbb{R}} g(x)\phi(x)dx$$

where $\phi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$. In this case, ϕ is the Radon–Nikodym derivative of the probability measure with respect to the usual Lebesgue measure.

More generally, if μ is any Stieltjes measure with distribution function F, and F is differentiable, then the Radon–Nikodym derivative of μ with respect to Lebesgue measure is given by the classical derivative dF/dx.

1.7 Signed Measures

It will prove useful to also have some results for *signed measures.* These are countably additive set functions which can take on both positive and negative values, and give a richer theory of integration than classic (nonnegative) measures.

Definition 1.7.1. *A countably additive set function that can take on negative values, but takes only one of the values $\pm\infty$, is called a* signed *measure.*

Definition 1.7.2. *Let μ be a signed measure on a measurable space (S, Σ).*

A set $A \in \Sigma$ will be called a positive set *if every measurable $B \subseteq A$ has $\mu(B) \geq 0$. It will be called a* negative set *if every measurable subset B has $\mu(B) \leq 0$. A set which is both positive and negative will be called a* null set.

Lemma 1.7.3. *Suppose $\mu(A) \neq -\infty$ for all $A \in \Sigma$. Then a set B with $\mu(B) > 0$ contains a positive subset $\tilde{B}$ with $\mu(\tilde{B}) > 0$.*

Proof. We define a sequence of sets recursively, by removing subsets of B with measure less than or equal to $-n^{-1}$. We wish to show that such a method will remove all subsets of B with negative measure – the remainder will then be a positive set.

For every $n \in \mathbb{N}$, let A_n be a measurable subset of $B \setminus (\bigcup_{m<n} A_m)$ with $\mu(A_n) \leq -n^{-1}$, if one exists, or $A_n = \emptyset$ otherwise. Either $\mu(A_n) = 0$ for infinitely many n, or, for some k, we have $\mu(A_n) \leq -n^{-1}$ for all $n \geq k$. In the latter case,

$$\sum_n \mu(A_n) \leq \sum_{n\geq k} \mu(A_n) \leq -\sum_{n\geq k} n^{-1} = -\infty,$$

that is, the sum will diverge. However, the sets A_n form a sequence of disjoint measurable subsets and, as $\mu(A) \neq -\infty$ for all $A \in \Sigma$, it is true that $0 \geq \mu(\bigcup_n A_n) = \sum_n \mu(A_n) > -\infty$.

Therefore, $\mu(A_n) = 0$ for infinitely many n. This implies that, for every n, there exists k such that $B \setminus (\bigcup_{m<k} A_m)$ has no subsets of measure less than or equal to $(-n^{-1})$.

Define $\tilde{B} = B \setminus (\bigcup_m A_m)$. Then

$$\mu(\tilde{B}) = \mu(B) - \mu\Big(\bigcup_m A_m\Big) = \mu(B) - \sum_m \mu(A_m) \geq \mu(B) > 0$$

and we know that $\tilde{B}$ has no subsets of measure less than or equal to $-n^{-1}$ for all n. Hence $\tilde{B}$ is a positive set. □

Lemma 1.7.4 (Hahn Decomposition). *For μ a signed measure on (S, Σ), there exists a positive set P and a negative set N such that $P \cup N = S$ and $P \cap N = \emptyset$. This decomposition is unique up to null sets.*

Proof. Without loss of generality, assume $\mu(A) \neq -\infty$ for all $A \in \Sigma$. Define $\lambda = \sup_{C \in \Sigma} \mu(C)$. Let $\{P_n\}_{n \in \mathbb{N}}$ be a sequence of positive sets such that

$$\lambda = \lim_n \mu(P_n).$$

It is clear that the union of a countable collection of positive sets is a positive set, and hence that $P = \bigcup_n P_n$ is a positive set. As $P \setminus P_n \subseteq P$, this implies that $P \setminus P_n$ is a positive set, and hence that

$$\mu(P) = \mu(P_n) + \mu(P \setminus P_n) \geq \mu(P_n)$$

for all n. Therefore $\mu(P) = \lambda$.

Now define $N = S \setminus P$. Let A be a positive subset of N. Then $\mu(P \cup A) = \mu(P) + \mu(A) \geq \lambda$, however $P \cup A$ is a positive set, therefore $\mu(P) + \mu(A) = \lambda = \mu(P)$, and hence $\mu(A) = 0$. Therefore, N contains no positive subsets of positive measure. Hence N is a negative set by Lemma 1.7.3.

The uniqueness of N and P is straightforward. □

Definition 1.7.5. *Two measures μ and ν on a space (S, Σ) will be called* mutually singular *if there exists a measurable set A with $\mu(A) = \nu(S \setminus A) = 0$, that is, we can divide S into a set A where μ is zero, and a set $S \setminus A$ where ν is zero.*

Lemma 1.7.6 (Jordan–Hahn Decomposition). *Let μ be a signed measure over a measurable space (S, Σ). Then there exist two unique measures μ^+ and μ^- such that $\mu = \mu^+ - \mu^-$ and μ^+ and μ^- are mutually singular.*

Proof. Let P be the positive subset of S given by the Hahn decomposition. Then $\mu^+(A) := \mu(A \cap P)$ and $\mu^-(A) := -\mu(A \setminus P)$. Clearly, μ^+ and μ^- are mutually singular (let $A = P$ in Definition 1.7.5). □

Definition 1.7.7. *Let S be a set and suppose μ is a set function defined on an algebra Σ of subsets of S. Let $\mathcal{D}_\Sigma(A)$ denote the set of all finite collections $\{A_i\}$ of disjoint sets in Σ such that $A_i \subseteq A$. Then for every $A \in \Sigma$ the* total variation *of μ on A, denoted $V_\mu(A)$, is defined by*

$$V_\mu(A) := \sup_{\{A_i\} \in \mathcal{D}_\Sigma(A)} \sum_i |\mu(A_i)|.$$

We say that μ is of bounded variation on $A \subseteq S$ *if $V_\mu(A) < \infty$, or simply of* bounded variation *if $V_\mu(S) < \infty$.*

Theorem 1.7.8. *Let μ be an $\overline{\mathbb{R}}$-valued finitely additive set function defined on an algebra Σ. For $A \in \Sigma$,*

$$V_\mu(A) \leq 2 \sup_{\{E \in \Sigma : E \subseteq A\}} |\mu(E)|.$$

Proof. We first consider the case where μ is $\mathbb{R}$-valued and bounded on A. For any finite disjoint collection $\{A_i\}_{i=1}^m \in \mathcal{D}_\Sigma(A)$, let $\mathcal{A}^+ := \{A_i : \mu(A_i) \geq 0\}$, and $\mathcal{A}^- := \{A_i : \mu(A_i) < 0\}$. Then

$$\sum_{i=1}^m |\mu(A_i)| = \sum_{\mathcal{A}^+} \mu(A_i) - \sum_{\mathcal{A}^-} \mu(A_i) = \mu\Big(\bigcup_{\mathcal{A}^+} A_i\Big) - \mu\Big(\bigcup_{\mathcal{A}^-} A_i\Big).$$

So, as $\bigcup_{\mathcal{A}^+} A_i$ and $\bigcup_{\mathcal{A}^-} A_i$ are both in Σ and are subsets of A,

$$V_\mu(A) = \sup_{\{A_i\}\in\mathcal{D}_\Sigma(A)} \Big\{\mu\Big(\bigcup_{\mathcal{A}^+} A_i\Big) - \mu\Big(\bigcup_{\mathcal{A}^-} A_i\Big)\Big\} \leq 2 \sup_{\{E\in\Sigma : E\subseteq A\}} \sup |\mu(E)|.$$

If μ is unbounded, or takes the value $+\infty$ or $-\infty$ on A, then $V_\mu(A) = \pm\infty$ and $\sup|\mu(E)| = +\infty$. In either case the result holds. $\square$

Corollary 1.7.9. *If an $\mathbb{R}$-valued, finitely additive set function defined on an algebra Σ is bounded, it is of bounded variation.*

Lemma 1.7.10. *For a signed measure μ the* absolute variation measure $|\mu| = \mu^+ + \mu^-$ *satisfies, for all $A \in \Sigma$,*

$$|\mu|(A) = V_\mu(A).$$

Proof. Let P be the positive subset of S from the Hahn decomposition. We know that

$$|\mu(A)| = |\mu^+(A) - \mu^-(A)| \leq \mu^+(A) + \mu^-(A) = |\mu|(A).$$

Therefore, $|\mu|(A)$ is an upper bound on $|\mu(A)|$. Hence, as $|\mu|$ is additive, $V_\mu(A) \leq |\mu|(A)$. By taking the decomposition

$$V_\mu(A) \geq |\mu(A \cap P)| + |\mu(A \setminus P)| = |\mu|(A)$$

we can achieve this bound. Therefore, $V_\mu(A) = |\mu|(A)$. $\square$

Remark 1.7.11. Taking P and N as in the Hahn decomposition 1.7.4, and defining $f = I_P - I_N$, we easily see that $|\mu|(A) = \int_A f d\mu$. For simplicity, we write $f = \operatorname{sign}(d\mu)$, and note that $|f| = 1$.

Definition 1.7.12. *If μ is a signed measure we define the Lebesgue integral of f over S with respect to μ by*

$$\int_S f\, d\mu = \int_S f\, d\mu^+ - \int_S f\, d\mu^-,$$

provided both integrals on the right, and their difference, are defined.

Definition 1.7.13. *For μ a signed measure on a measure space (S, Σ), we shall say μ is finite (resp. σ-finite) if $|\mu|$ is a finite (resp. σ-finite) measure. We shall say ν is absolutely continuous with respect to μ, and write $\nu \ll \mu$, if $|\nu| \ll |\mu|$. We shall say ν is equivalent to μ if $|\nu|$ is equivalent to $|\mu|$.*

Theorem 1.7.14 (Radon–Nikodym Theorem for Signed Measures). *Let μ, ν be σ-finite signed measures on a measure space (S, Σ) such that $\nu \ll \mu$. Then there exists a measurable function $f : S \to \overline{\mathbb{R}}$ such that*

$$\nu(A) = \int_A f \, d\mu.$$

As before, f is called the Radon–Nikodym derivative *of ν with respect to μ, is sometimes written $\frac{d\nu}{d\mu}$ and is unique up to equality except on some null set.*

Proof. Let $\tilde{f}$ be the Radon–Nikodym derivative of $|\nu|$ with respect to $|\mu|$. Let P_μ, P_ν, N_μ, N_ν be the positive and negative sets from the Hahn decomposition with μ, ν respectively.

Define

$$f(s) = \begin{cases} \tilde{f}(s) & \text{for } s \in (P_\mu \cap P_\nu) \cup (N_\mu \cap N_\nu), \\ -\tilde{f}(s) & \text{for } s \in (P_\mu \cap N_\nu) \cup (N_\mu \cap P_\nu). \end{cases}$$

It is easy to check that f is the desired Radon–Nikodym derivative. □

Theorem 1.7.15 (Lebesgue Decomposition). *Let μ, ν be σ-finite signed measures on a measure space (S, Σ). Then there exists a signed measure ν_0 absolutely continuous with respect to μ, and a signed measure ν_1 singular with respect to μ, such that $\nu = \nu_0 + \nu_1$. These signed measures are unique up to equality $|\mu| + |\nu|$-a.e.*

Proof. Clearly $\lambda = |\mu| + |\nu|$ is also a σ-finite measure. Then the Radon–Nikodym theorem yields functions f and g such that $\mu(A) = \int_A f \, d\lambda$ and $\nu(A) = \int_A g \, d\lambda$. Then $h_0 = I_{\{f \neq 0\}} g$ and $h_1 = I_{\{f=0\}} g$ are two measurable functions. Define $\nu_0(A) = \int_A h_0 \, d\lambda$ and $\nu_1 = \int_A h_1 \, d\lambda$. It is straightforward to show the desired properties. □

Theorem 1.7.16 (Riesz Representation Theorem). *Let (S, Σ, μ) be a σ-finite measure space, and let F be a bounded linear functional on L^p (in the sense of Definition 1.5.7), for $p \in [1, \infty[$. Then there exists $g \in L^q$, where $p^{-1} + q^{-1} = 1$, such that*

$$F(f) = \int_S f g \, d\mu.$$

In other words, the topological dual of L^p is isomorphic to L^q.

Proof. By linearity of the integral and Hölder's inequality, it is clear that $F(\cdot) = \int (\cdot) g d\mu$ is in $(L^p)'$ for any $g \in L^q$.

To show the converse, first assume $\mu(S) < \infty$. Then, for any bounded linear functional F we can define a set function $\nu(A) = F(I_A)$. As F is bounded,

$|\nu(A)| < \infty$ for all A. As F is continuous and linear, one can show (see Lemma A.1.1) that ν is a countably additive set function, and therefore is a signed measure. As $\mu(A) = 0 = F(0) = F(I_A) = \nu(A)$ for all μ-null sets A, ν is absolutely continuous with respect to μ.

Hence, by the Radon–Nikodym theorem, we can write

$$F(I_A) = \nu(A) = \int_S I_A g \, d\mu$$

for some (integrable) function g, and more generally, $F(f) = \int_S fg \, d\mu$ for any simple function f. We extend this to the case where μ is σ-finite by finding g on each part of an appropriate partition.

For $p = 1$, suppose that g is unbounded, in particular, that $\operatorname{ess\,sup} |g| = \infty$. Without loss of generality, g is unbounded above, so for any K, there exists a set A such that $g \geq K$ on A and $\mu(A) > 0$. Hence $F(I_A) = \int_A g \, d\mu \geq K\mu(A)$. However, as F is bounded, $F(I_A) \leq c\mu(A)$ for some fixed $c \in \mathbb{R}$, which gives a contradiction. Hence g is essentially bounded, that is, $g \in L^\infty = L^q$.

For $p > 1$, let ϕ_n be a sequence of nonnegative simple functions increasing pointwise to $|g|$. Then we have that $\phi_n^{q/p} \uparrow |g|^{q/p}$. Define $\psi_n := \phi_n \frac{g}{|g|}$, so that

$$\int_S \phi_n^q \, d\mu = \int_S \phi_n^{1+q/p} \, d\mu \leq \int_S |g| \phi_n^{q/p} \, d\mu = \int_S \psi_n g \, d\mu = F(\psi_n^{q/p}).$$

By boundedness of F, for some $c \in \mathbb{R}$,

$$\int_S \phi_n^q \, d\mu = F(\psi_n^{q/p}) \leq c\Big(\int_S |\psi_n|^q\Big)^{1/p} = c\Big(\int_S \phi_n^q \, d\mu\Big)^{1/p}.$$

Therefore,

$$\int_S \phi_n^q \, d\mu \leq c\Big(\int_S \phi_n^q \, d\mu\Big)^{1/p} \qquad \text{and hence} \qquad \Big(\int_S \phi_n^q \, d\mu\Big)^{1/q} \leq c,$$

that is, $\phi_n \in L^q$. By the monotone convergence theorem, this implies that $g \in L^q$.

We know $F(f) = \int_S fg \, d\mu$ for any simple function $f \in L^p$. For general $f \in L^p$ with $f \geq 0$, let $\{f_n\}_{n \in \mathbb{N}}$ be a sequence of simple measurable functions increasing pointwise to f. Then

$$\int (f - f_n)^p \, d\mu \to 0,$$

by the dominated convergence theorem. Therefore, $f_n \to f$ in $\|\cdot\|_p$, so $F(f_n) \to F(f)$ by continuity. By dominated convergence, $\int_S f_n g \, d\mu \to \int_S fg \, d\mu$, and therefore $F(f) = \int_S fg \, d\mu$. Linearity then implies $F(f) = \int_S fg \, d\mu$ for all $f \in L^p$. □

Remark 1.7.17. We note that this result does *not* hold for $p = \infty$, that is, L^1 is not isomorphic to $(L^\infty)'$. (However, it still holds from Hölder's inequality that L^1 is isomorphic to a subset of $(L^\infty)'$.)

Corollary 1.7.18. *For $p \in \,]1, \infty[$, the space L^p is* reflexive*, that is, the dual of the dual of L^p is (isomorphic to) L^p.*

The following theorem, which we state without proof, gives a useful description of weak compactness in terms of L^p boundedness. Note that the important case $p = 1$ is not covered. This will be considered in Theorem 2.5.11.

Theorem 1.7.19 (Riesz Weak Compactness Theorem). *Let (S, Σ, μ) be a σ-finite measure space and $p \in \,]1, \infty[$. Then every L^p-bounded set is weakly relatively compact (that is, its weak closure is a weakly compact set). In particular, any sequence $\{f_n\}_{n\in\mathbb{N}}$ bounded in L^p has a weakly convergent subsequence, that is, if $\|f_n\|_p < K$ for all n (for a fixed $K > 0$), then there is a subsequence $\{f_{n_k}\}_{k\in\mathbb{N}}$ and a function $f \in L^p$ such that*

$$\lim_{k\to\infty} \int f_{n_k} g d\mu = \int f g d\mu$$

for all $g \in L^q$, where $p^{-1} + q^{-1} = 1$.

Proof. See Royden and Fitzpatrick [160, p.408]. □

1.7.1 Functions of Bounded Variation

We now extend our results on Lebesgue–Stieltjes measures to signed measures.

Definition 1.7.20. *A right-continuous function $f : [0, \infty[\,\to \mathbb{R}$ is said to be of bounded variation (or finite variation) if, for $T \in [0, \infty[$ and any increasing sequence $\{t_i\}_{i\in\mathbb{N}} \subset [0, T]$*

$$\sum_i |f(t_{i+1}) - f(t_i)| < \infty.$$

Lemma 1.7.21. *If f is a right-continuous function of bounded variation, then there is a pair of right-continuous nondecreasing functions g and h such that $f = g - h$.*

Proof. Let $\mathcal{D}([0, t])$ denote the set of all increasing sequences in $[0, t]$. We write

$$g(t) = f(0) + \sup_{\{t_i\}\in\mathcal{D}([0,t])} \sum_i (f(t_{i+1}) - f(t_i))^+,$$
$$h(t) = \sup_{\{t_i\}\in\mathcal{D}([0,t])} \sum_i (f(t_{i+1}) - f(t_i))^-.$$

As we can assume without loss of generality that $t_1 = 0$, it is clear that there is a sequence of elements of $\mathcal{D}([0, t])$ which approaches the required suprema for both g and h. Therefore, $f = g - h$. It is straightforward to verify that g and h are right-continuous and nondecreasing. □

Theorem 1.7.22. *There is a one-to-one correspondence between the right-continuous functions of bounded variation (up to addition by a constant) and the finite signed measures on $\mathcal{B}(\mathbb{R})$.*

Proof. For any function of finite variation, we can find mutually singular distribution functions g and h with $f = g - h$. Define μ_g and μ_h to be the Stieltjes measures generated by g and h, then $\mu_f := \mu_g - \mu_h$ is a signed measure. As g and h are uniquely defined up to the addition of a function k (that is, if $f = g - h = g' - h'$ then $g' = g + k$ and $h' = h + k$ for some k), we see that

$$\mu_f = \mu_{g'} - \mu_{h'} = (\mu_g + \mu_k) - (\mu_h + \mu_k) = \mu_g - \mu_h,$$

so this measure is uniquely defined.

Conversely, for any signed measure μ, we have the Jordan–Hahn decomposition $\mu = \mu^+ - \mu^-$. By Theorem 1.2.16, μ^+ has a unique distribution function g, and μ^- a unique distribution function h. Writing $f = g - h$ uniquely defines a function of finite variation. □

Remark 1.7.23. If a function f is not of bounded variation, then we cannot use the above procedure to generate a signed measure corresponding to f. We shall see that, due to this fact, we will not be able to employ the Lebesgue–Stieltjes procedure to define stochastic integrals in general.

1.8 Exercises

Exercise 1.8.1. For an arbitrary index set $\mathcal{A}$, let $\{\Sigma_\alpha\}_{\alpha \in \mathcal{A}}$ be a collection of σ-algebras on a set S. Show that $\Sigma' = \bigcap_{\alpha \in \mathcal{A}} \Sigma_\alpha$ is a σ-algebra on S.

Exercise 1.8.2. Prove that $\sigma(\Sigma_{\mathfrak{J}}) = \mathcal{B}(\mathbb{R})$, in the notation of Example 1.1.4(iii).

Exercise 1.8.3. For a measure space (S, Σ, μ), and a sequence $\{A_n\}_{n \in \mathbb{N}} \subset \Sigma$, show the following properties.

(i) $\mu\big(\bigcup_{n \in \mathbb{N}} A_n\big) \leq \sum_n \mu(A_n)$.
(ii) If $\{A_n\}_{n \in \mathbb{N}}$ is nondecreasing, that is, $A_n \subseteq A_{n+1}$, then $\mu(\bigcup_n A_n) = \lim_n \mu(A_n)$
(iii) If $\{A_n\}_{n \in \mathbb{N}}$ is nonincreasing, that is, $A_{n+1} \subseteq A_n$, and $\mu(S) < \infty$, then $\mu\big(\bigcap_n A_n\big) = \lim_n \mu(A_n)$.

For the final property, give a counterexample to show that this does not necessarily hold when $\mu(S) = \infty$.

Exercise 1.8.4. For $\{f_n\}_{n \in \mathbb{N}}$ a sequence of measurable functions, show that $\sup_n f_n$, $\inf_n f_n$, $\limsup_n f_n$ and $\liminf_n f_n$ are all measurable

Exercise 1.8.5. For f, g measurable functions $(S, \Sigma) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$, show that $\{s : f(s) \geq g(s)\}$ and $\{s : f(s) \leq g(s)\}$ are in Σ.

Exercise 1.8.6. For $f : S \to \overline{\mathbb{R}}$ a measurable function (in the sense of Remark 1.3.2), g a *Borel* measurable function $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}})) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$, show that the composition $g \circ f$ is measurable.

(Note that this is not necessarily true if g is only a Lebesgue measurable function.)

Exercise 1.8.7. For f, g measurable functions $(S, \Sigma) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$, $c \in \mathbb{R}$, show that $f + g$, $f \times g$, $\max\{f, g\}$, $1/f$ and cf are all measurable, where $1/0 := \infty$.

Exercise 1.8.8. Let ϕ be a simple function $S \to \mathbb{R}$, vanishing outside a measurable set B. Let $\{A_n\}_{n\in\mathbb{N}}$ be a sequence of measurable sets, with $A_n \subseteq A_{n+1}$ for all n, and $B \subseteq \bigcup_n A_n$. Show from first principles that

$$\lim_{n\to\infty} \int_{A_n} \phi \, d\mu = \int_B \phi \, d\mu.$$

Exercise 1.8.9. For f, g integrable functions, $f \le g$ a.e., show that

$$\int_S f \, d\mu \le \int_S g \, d\mu.$$

Exercise 1.8.10. Show that a pair of integrable functions $f = g$ a.e. if and only if

$$\int_A f \, d\mu = \int_A g \, d\mu$$

for all $A \in \Sigma$.

Exercise 1.8.11. Suppose $p^{-1} + q^{-1} = 1$, and consider a pair of sequences $\{f_n\}_{n\in\mathbb{N}}$ and $\{g_n\}_{n\in\mathbb{N}}$ such that $f_n \to f$ in L^p and $g_n \to g$ in L^q. Show that $f_n g_n \to fg$ in L^1.

Exercise 1.8.12. Let (S_1, Σ_1, μ) and (S_2, Σ_2, ν) be two measure spaces. Show that every set in $(S_1 \times S_2, \Sigma_1 \otimes \Sigma_2, \mu \times \nu)$ can be written as the union of a countable disjoint union of sets of the form $A \times B$, for $A \in \Sigma_1$, $B \in \Sigma_2$. (Hint: You need only show that this is the smallest σ-algebra containing all sets of this form.)

Exercise 1.8.13. Show that L^∞ is a complete vector space.

Exercise 1.8.14. Show that, if μ and ν are signed measures on a measure space (S, Σ) and $\nu \ll \mu$, then, for any $|\nu|$-integrable f,

$$\int_S f \, d\nu = \int_S f \cdot \frac{d\nu}{d\mu} \, d\mu.$$

Exercise 1.8.15. Show that, if μ and ν are equivalent signed measures on a measure space (S, Σ), then

$$\left(\frac{d\nu}{d\mu}\right)\left(\frac{d\mu}{d\nu}\right) = 1,$$

except possibly on some μ- (or, equivalently, ν-) null set.

Exercise 1.8.16. Let F be a differentiable distribution function, with derivative F'. Show that, if μ is the measure associated with F, then F' is the Radon–Nikodym derivative of μ with respect to Lebesgue measure.

Exercise 1.8.17. Show that a right continuous function of bounded variation has a left limit at every point.

2

Probabilities and Expectation

We now see how general measure theory specializes when we consider applications to probability.

In this context, σ-algebras provide a natural structure with which to model 'information'. As we have seen in the Doob–Dynkin lemma, if we have the σ-algebra generated by a Borel measurable function $f : \mathbb{R} \to \mathbb{R}$, then any $\sigma(f)$-measurable function g can be written in the form $g = h \circ f$, for some Borel function h. We can see that the statement 'g is $\sigma(f)$-measurable' can be interpreted as 'g contains no information not available from knowing the value of f'.

Measure theory also gives a firm mathematical foundation with which to talk about 'probabilities' of events, from an axiomatic standpoint. Using this theory allows statements to be made more carefully and rigorously than using a naïve approach, and many of the philosophical difficulties associated with probability can be avoided.

On the other hand, this approach gives only a mathematical structure within which to work, and the interpretation of what exactly is meant by a 'probability' is not defined – the mathematics remains fundamentally the same whether one takes a frequentist, Bayesian or other philosophical position.

2.1 Probability Spaces

Definition 2.1.1. *Let Ω be a set which contains the outcomes ω of some experiment. We call Ω the* sample space*. Let $\mathcal{F}$ be a σ-algebra on Ω. The elements of $\mathcal{F}$ are called* events*.*

We need the structure of $\mathcal{F}$ being a σ-algebra on Ω so as to consistently define probabilities and expectations. Conceptually, $\mathcal{F}$ contains those events for which we know, at the end of the experiment, whether the event occurred or not. Thus we say "event A has occurred" if $\omega \in A$.

S.N. Cohen, R.J. Elliott, *Stochastic Calculus and Applications*,
Probability and Its Applications, DOI 10.1007/978-1-4939-2867-5_2

From this intuition, it is clear that $\mathcal{F}$ should be a σ-algebra. That is to say, if we know that "event A has occurred", then we should also know that "the opposite of event A has not occurred". Hence, both A and A^c should be in the information set $\mathcal{F}$. Similarly if we know whether or not "each event in a countable collection $\{A_n\}_{n\in\mathbb{N}}$ has occurred", then we know if "at least one of the events has occurred". This translates into the assumption that $\bigcup_{n\in\mathbb{N}} A_n$ should also be in the information set.

Definition 2.1.2. *A* probability measure *is a (finite) measure P on $(\Omega, \mathcal{F})$ satisfying $P(\Omega) = 1$.*

The probability of an event $A \in \mathcal{F}$ is given by $P(A)$. In many cases, we use a simplified notation, for example, we write $P(X > Y)$ for $P\big(\{\omega : X(\omega) > Y(\omega)\}\big)$.

Definition 2.1.3. *A triple $(\Omega, \mathcal{F}, P)$, where Ω is a set, $\mathcal{F}$ a σ-algebra of subsets of Ω and P a probability measure on $\mathcal{F}$, is called a* probability space *or* probability triple.

Remark 2.1.4. As any probability space is a measure space, the concepts of measure theory, such as completeness, absolute continuity, measurable functions, etc., as well as the associated results, all extend to this setting. We shall see that some of these concepts have different names, to highlight their interpretation in the context of random outcomes and probabilities.

Example 2.1.5. Some classic probability spaces.

(i) $([0,1], \mathcal{B}([0,1]), P)$ where P equals Lebesgue measure is a probability space. Here ω is an outcome 'uniformly distributed' on $[0,1]$, and $P(]a,b]) = b - a$ is the probability ω lies in the interval $[a,b]$.

(ii) $(\mathbb{R}, \overline{\mathcal{B}}(\mathbb{R}), P)$, where P is defined by

$$P(A) = \int_A \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

(This is a probability space where $X(\omega) = \omega$ has a normal distribution.)

(iii) Let $\Omega = \{1,2,3,4,5,6\}, \mathcal{F} = 2^{\Omega}$ and P be the counting measure divided by 6. (This is the typical model for the roll of a die.)

(iv) Let $\Omega = [0,\infty[\times[0,\infty[$. Let $\mathcal{F}$ be the sets of the form $A \times B$, for A a Borel set of $[0,\infty[$ and $B \in \{\emptyset, [0,\infty[\}$. Let $P(A \times [0,\infty[) = \int_A \lambda e^{-\lambda x} dx$ for some $\lambda > 0$ and $P(A \times \emptyset) = 0$. Then $(\Omega, \mathcal{F}, \mu)$ is a probability space. In this example, $\mathcal{F}$ gives us no information about the second dimension of Ω – any function which is not constant with respect to the second component of $\omega = (\omega_1, \omega_2) \in [0,\infty[\times[0,\infty[$ will not be $\mathcal{F}$ measurable. On the other hand, the first component ω_1 has an 'exponential distribution', with $P(\omega_1 \le t) = 1 - e^{-\lambda t}$.

Definition 2.1.6. *A measurable function X from a probability space $(\Omega, \mathcal{F}, P)$ to a measurable space $(E, \mathcal{E})$ is called an E-*valued random variable*, or* random element*. If $(E, \mathcal{E}) = (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$, then X is simply called a* random variable*.*

A statement will be said to hold almost surely *(a.s.) if it holds with probability one, that is, if it is true almost everywhere in the measure space $(\Omega, \mathcal{F}, P)$. We write P-a.s. if the measure P needs to be specified.*

If X and Y are random variables with $X = Y$ a.s., then we say that Y is a version *of X.*

This definition makes precise the notion that a random variable X is a *numerical* outcome of an experiment. For every outcome ω, we have a value $X(\omega) \in \mathbb{R}$. As X is $\mathcal{F}$ measurable, at the end of the experiment we will know if $X(\omega) \in B$ for each Borel set B, or equivalently, if $X(\omega) > a$, for each $a \in \mathbb{R}$.

Definition 2.1.7. *For X a random variable, the integral of X with respect to P will be called the* expectation *of X whenever it is defined, and is written*

$$E[X] = \int_\Omega X(\omega)\, dP(\omega).$$

Definition 2.1.8. *A random variable X defined on a probability space $(\Omega, \mathcal{F}, P)$ induces a probability P^X on $(E, \mathcal{E})$, called the* distribution of X *or* law of X *as follows:*

For all $A \in \mathcal{E}$,

$$P^X(A) = P \circ X^{-1}(A) = P\big(\{\omega : X(\omega) \in A\}\big).$$

If $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ then the function F defined by

$$F(a) = P(X \le a)$$

is called the (cumulative) distribution function *of X. It is easy to show this is a distribution function in the sense of Definition 1.2.14, and the law of X is the associated Lebesgue–Stieltjes measure.*

Example 2.1.9. Let X be a random variable on a probability space $(\Omega, \mathcal{F}, P)$ such that $P(X \in A) = \int_{A \cap [0,\infty[} \lambda e^{-\lambda x} dx$ for some $\lambda > 0$. Then X induces a distribution function on $\mathbb{R}$, namely

$$F(a) = P(X \le a) = \begin{cases} 0 & a < 0 \\ 1 - e^{-\lambda a} & a \ge 0 \end{cases}$$

In this case, X is said to have an exponential distribution with rate λ.

The notion of independence is fundamental to most probabilistic modelling. The classic requirement is that two events A, B are independent if $P(A \cap B) = P(A)P(B)$. We here generalize this notion to make use of our richer mathematical setting.

Definition 2.1.10. *Given a probability space $(\Omega, \mathcal{F}, P)$,*

- *a finite collection $\{A_n\}_{n=1}^m$ of events is called* independent *if*

$$P\Big(\bigcap_{k=1}^{j} A_{i_k}\Big) = \prod_{k=1}^{j} P(A_{i_k})$$

  for any $\{A_{i_k}\}_{k=1}^j \subset \{A_n\}_{n=1}^m$;
- *an arbitrary collection of events $\{A_\lambda\}_{\lambda \in I}$ is independent provided every finite subcollection of $\{A_\lambda\}_{\lambda \in I}$ is independent;*
- *a finite collection of σ-algebras $\{\mathcal{F}_n\}_{n=1}^m$ is called independent if for every choice of $A_n \in \mathcal{F}_n$, $n = 1, \ldots, m$, the set $\{A_n\}_{n=1}^m$ is a collection of independent events;*
- *an arbitrary collection of σ-algebras $\{\mathcal{F}_\lambda\}_{\lambda \in I}$ is said to be independent if every finite subcollection of σ-algebras is independent.*

The above definition describes independence of *events*. This naturally extends to independence of random variables, as follows.

Definition 2.1.11. *A collection of E-valued random variables $\{X_i\}$ is called independent if, for every $A, B \in \mathcal{E}$, the events $X_i^{-1}(A), X_j^{-1}(B)$ are independent for all $i \neq j$.*

We shall occasionally use the following classical result, which provides a version of the Fourier transformation for random variables.

Lemma 2.1.12. *The law of a random variable is determined uniquely by its* characteristic function*, defined by*

$$\phi_X(t) = E[e^{itX}].$$

Proof. It is straightforward to show that, if μ is the law of X, then for any $a < b$,

$$\lim_{T \to \infty} \frac{1}{2\pi} \int_{[-T,T]} \frac{e^{-ita} - e^{-itb}}{it} \phi_X(t) dt = \mu\big(]a, b[\big) + \frac{\mu(\{a\}) + \mu(\{b\})}{2}.$$

Taking $a_n \downarrow -\infty$ and $b_n \downarrow b$, we can determine the value of $\mu(]-\infty, b]) = P(X \leq b)$, and hence the law of X. $\square$

2.1.1 Borel–Cantelli Lemma

In this section we prove a useful result about sequences of measurable sets.

Theorem 2.1.13 (Borel–Cantelli Lemma). *Let $\{A_n\}_{n \in \mathbb{N}}$ be a sequence in a σ-algebra Σ, and μ be a measure on Σ.*

(i) If $\sum_n \mu(A_n) < \infty$*, then it follows that* $\mu\left(\bigcap_k \bigcup_{n\geq k} A_n\right) = 0$.
(ii) If μ *is a probability measure,* $\sum_n \mu(A_n) = \infty$ *and the events* $\{A_n\}_{n\in\mathbb{N}}$ *are independent, then*

$$\mu\Big(\big(\bigcap_k \bigcup_{n\geq k} A_n\big)^c\Big) = 0.$$

Proof. To prove (i), as $\sum_n \mu(A_n) < \infty$, we must have $\lim_{k\to\infty}(\sum_{n=k}^{\infty} \mu(A_n)) = 0$. However, it follows that

$$\mu\Big(\bigcap_i \bigcup_{n\geq i} A_n\Big) \leq \inf_k \mu\Big(\bigcup_{n\geq k} A_n\Big) \leq \inf_k \sum_{n\geq k} \mu(A_n) = 0.$$

For (ii), one can easily verify that $\mu\left(\bigcap_{n\geq k} A_n^c\right) = \prod_{n\geq k} \mu(A_n^c)$ for all k. Hence

$$\begin{aligned}
\mu\Big(\big(\bigcap_k \bigcup_{n\geq k} A_n\big)^c\Big) &= \mu\Big(\bigcup_k \bigcap_{n\geq k} A_n^c\Big) = \lim_{k\to\infty} \mu\Big(\bigcap_{n\geq k} A_n^c\Big) \\
&= \lim_{k\to\infty} \prod_{n\geq k} \mu(A_n^c) = \lim_{k\to\infty} \prod_{n\geq k} (1-\mu(A_n)) \\
&\leq \lim_{k\to\infty} \prod_{n\geq k} \exp(-\mu(A_n)) = \lim_{k\to\infty} \exp(-\sum_{n\geq k} \mu(A_n)) = 0.
\end{aligned}$$

□

Remark 2.1.14. The event $\bigcap_i \bigcup_{n\geq i} A_n$ is often interpreted as 'infinitely many of the events $\{A_n\}_{n\in\mathbb{N}}$ occur'. This result indicates that if the probabilities of the individual events can be summed, then the measure of this set (the probability of infinitely many events occurring) is zero; conversely, if the probabilities of the individual events cannot be summed (and the events are independent), then the probability of infinitely many of them occurring is one.

2.2 Conditional Expectation

The notion of conditional expectation is fundamental for probability theory. It is easy to understand in elementary situations, such as probabilities associated with the throwing of a fair die. For example, the probability we roll a four is 1/6; the conditional probability of rolling four, given that an even number is rolled, is $(1/6)/(1/2) = 1/3$, as it is one of the three equally likely possibilities $\{2, 4, 6\}$.

On a general probability space $(\Omega, \mathcal{F}, P)$ the conditional probability of event $B \in \mathcal{F}$ occurring, given that event $A \in \mathcal{F}$ has occurred, can be defined by $P(B|A) = P(A\cap B)/P(A)$, as long as $P(A) \neq 0$. For A fixed, we can verify that $P(\cdot|A)$ is a probability measure on $(\Omega, \mathcal{F})$, and that $P(A|A) = 1$. Using this, if X is a random variable, we can define the conditional expectation of X given A by the integral $\int_\Omega X dP(\cdot|A) = E[X|A]$.

These previous examples are both in the setting of a single event A. When we are told the outcome of a collection of events $\{A_a\}_{a\in\mathcal{A}}$ for $\mathcal{A}$ an index set, the correct mathematical notion becomes more delicate. For example, on a probability space $(\Omega, \mathcal{F}, P)$ consider a random variable Z with values in $\mathbb{N} = \{1, 2, \ldots\}$. Write

$$A_n = \{\omega : Z(\omega) = n\} \in \mathcal{F}.$$

Suppose X is another (real) random variable defined on Ω. A natural definition for the conditional expectation of X, given the outcome of Z, is the random variable defined by

$$Y(\omega) = \frac{\int_{A_n} X\,dP}{P(A_n)} = \int_\Omega X dP(\cdot|A_n), \qquad \text{choosing } n \text{ such that } \omega \in A_n.$$

Equivalently, we could write $Y(\omega) = \sum_n I_{A_n(\omega)} E[X|A_n]$. When $P(A_n) = 0$, Y can be given an arbitrary value (for example, 0). Roughly speaking, X has been averaged over the sets on which Z is constant. Note that Y is here also a random variable, which depends on the outcome of Z (it is measurable with respect to the σ-algebra generated by Z). Note that Y satisfies

$$E[I_{A_n} Y] = E\big[I_{A_n} E[X|A_n]\big] = E[I_{A_n} X]$$

for any n.

Consider now a more general situation. We seek to define the conditional expectation of a $\overline{\mathbb{R}}$-random variable X given a second S-valued random variable Z. This is more complex than the above setting, as we are not conditioning on a fixed event, but rather on another random variable, which may take infinitely many values. The natural meaning of a conditional expectation is then a function $Y : S \to \overline{\mathbb{R}}$ which takes the random outcome $Z(\omega)$ and gives the expectation of X given this outcome. This motivates the following approach.

Theorem 2.2.1. *Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and Z an S-valued random variable defined on $(\Omega, \mathcal{F})$.*

Write Q for the probability measure induced on (S, Σ) by Z (that is, $Q(A) = P(Z^{-1}(A)), A \in \Sigma$) and let X be a P-integrable random variable on $(\Omega, \mathcal{F})$. Then there exists a Q-integrable random variable Y on (S, Σ) such that, for every $A \in \Sigma$,

$$\int_A Y\,dQ = \int_{Z^{-1}(A)} X\,dP. \tag{2.1}$$

Furthermore, if Y' also satisfies (2.1) then $Y' = Y$ P-a.s.

Proof. For $A \in \Sigma$, we can consider the set function given by

$$\nu(A) = \int_{Z^{-1}(A)} X\,dP.$$

As $X \in L^1$, it is straightforward to show that ν is a σ-finite signed measure on (S, Σ). Also, ν is absolutely continuous with respect to Q. Hence, by the Radon–Nikodym Theorem (Thm 1.7.14), there exists a unique function Y such that

$$\nu(A) = \int_A Y\, dQ = \int_{Z^{-1}(A)} X\, dP.$$

□

Remark 2.2.2. The random variable $Y \circ Z : \Omega \to \overline{\mathbb{R}}$ is called the conditional expectation of X given Z. By the Doob–Dynkin Lemma (Theorem 1.3.12) we see that $Y \circ Z$ is $\sigma(Z)$-measurable.

2.3 Conditioning with Respect to a Sub-σ-Algebra

The above definition of conditional expectation is sufficient for many simple applications. It is useful, however, to generalize this result to allow conditioning on an arbitrary σ-algebra $\mathcal{E} \subseteq \mathcal{F}$. It is also convenient to have the conditional expectation defined on the same space Ω as the original random variable. This will allow us to model the information available in more detail. The earlier results then correspond to the special case where $\mathcal{E}$ is generated by Z.

Definition 2.3.1. *Suppose X is a real-valued integrable random variable defined on $(\Omega, \mathcal{F}, P)$. Then the* conditional expectation *of X given $\mathcal{E} \subseteq \mathcal{F}$ is any $\mathcal{E}$-measurable, integrable random variable Y such that, for all $A \in \mathcal{E}$,*

$$E[I_A X] = \int_A X\, dP = \int_A Y\, dP = E[I_A Y].$$

Y is denoted by $E[X|\mathcal{E}]$.

Theorem 2.3.2. *For $\mathcal{E} \subseteq \mathcal{F}$, the conditional expectation $E[X|\mathcal{E}]$ exists for all $X \in L^1$, and is unique up to equality a.s.*

Proof. As in Theorem 2.2.1, let $\nu(A) = \int_A X\, dP$ for $A \in \mathcal{E}$. Then ν is a signed measure absolutely continuous with respect to the restriction of P to $\mathcal{E}$. By the Radon–Nikodym theorem, there exists an a.s. unique $\mathcal{E}$-measurable Y such that

$$\int_A Y\, dP = \nu(A) = \int_A X\, dP.$$

This Y is the conditional expectation $Y = E[X|\mathcal{E}]$, as desired. □

Remark 2.3.3. The conditional expectation Y is defined only almost surely, so Y should be called a version of the conditional expectation. Roughly speaking, Y is the average of X over the coarser sets of $\mathcal{E}$. If $\mathcal{E}$ represents the events that are 'known' at the present time, then $E[X|\mathcal{E}]$ is the average value of X given our present knowledge.

If $\mathcal{E}$ is the σ-algebra generated by a family $\{Z_a\}_{a\in\mathcal{A}}$, of random variables, that is, $\mathcal{E} = \sigma\big(\{Z_a\}_{a\in\mathcal{A}}\big)$, then Y is called the conditional expectation of X given the family $\{Z_a\}_{a\in\mathcal{A}}$.

If $X = I_A$, where I_A is the indicator function of A, then Y is called the conditional probability of A given $\mathcal{E}$. We emphasize that Y is not a number, but rather a random variable. (See Section 2.6 for further discussion.)

Remark 2.3.4. If $\mathcal{E}$ is the trivial σ-algebra (that is, $\mathcal{E} = \{\emptyset, \Omega\}$), then $E[X|\mathcal{E}] = E[X]$ (as only constant functions are measurable).

2.4 Properties of Conditional Expectations

In what follows, let $(\Omega, \mathcal{F}, P)$ be a probability space on which the random variables are defined, and let $\mathcal{E}$ be any sub-σ-algebra of $\mathcal{F}$.

Lemma 2.4.1 (Linearity). *Suppose X and Y are integrable random variables, and α, β, $\gamma \in \mathbb{R}$ are constants. Then*

$$E[\alpha X + \beta Y + \gamma|\mathcal{E}] = \alpha E[X|\mathcal{E}] + \beta E[Y|\mathcal{E}] + \gamma \quad a.s.$$

(a.s. here means that any *version of the left-hand side equals a version of the right-hand side.)*

Proof. For any $A \in \mathcal{E}$,

$$\begin{aligned}\int_A E[\alpha X + \beta Y + \gamma|\mathcal{E}]\, dP &= \int_A (\alpha X + \beta Y + \gamma)\, dP \\ &= \alpha \int_A X\, dP + \beta \int_A Y\, dP + \gamma \int_A dP \\ &= \alpha \int_A E[X|\mathcal{E}]\, dP + \beta \int_A E[Y|\mathcal{E}]\, dP + \gamma \int_A dP \\ &= \int_A (\alpha E[X|\mathcal{E}] + \beta E[Y|\mathcal{E}] + \gamma)\, dP.\end{aligned}$$

The integrand in the final integral is $\mathcal{E}$-measurable and we have shown it has the same integral over all sets $A \in \mathcal{E}$ as the ($\mathcal{E}$-measurable) integrand in the first integral. Therefore they are equal almost surely (Exercise 1.8.10). □

Lemma 2.4.2. *Suppose X and Y are integrable random variables and $X \le Y$ a.s. Then*

$$E[X|\mathcal{E}] \le E[Y|\mathcal{E}] \quad a.s.$$

Proof. The proof is obtained immediately by integrating over arbitrary sets $A \in \mathcal{E}$. □

Lemma 2.4.3 (Dominated/Monotone Convergence). *Suppose* $\{X_n\}_{n\in\mathbb{N}}$, *is a sequence of integrable random variables which converge almost surely to an integrable random variable* X. *If either there exists a nonnegative integrable random variable* Y *with* $|X_n| \leq Y$, *or* X_n *is monotone increasing, or* X_n *is monotone decreasing, then*

$$E[X|\mathcal{E}] = \lim_n E[X_n|\mathcal{E}] \quad a.s.$$

Proof. By the appropriate convergence theorem (Theorem 1.3.34, Theorem 1.3.29 or Corollary 1.3.32), for any $A \in \mathcal{E}$,

$$\begin{aligned}\int_A \lim_n E[X_n|\mathcal{E}]\, dP &= \lim_n \int_A E[X_n|\mathcal{E}]\, dP \\ &= \lim_n \int_A (X_n)\, dP = \int_A (X)\, dP = \int_A E[X|\mathcal{E}]\, dP.\end{aligned}$$

□

Remark 2.4.4. By linearity and monotone convergence, we can uniquely extend our definition of conditional expectation to all random variables X such that at least one of X^+ and X^- is integrable. Most of the stated results follow directly.

Remark 2.4.5. If $\{X_n\}_{n\in\mathbb{N}}$ converges only in probability, then, by Lemma 1.3.39, the result of Lemma 2.4.3 remains true, but the limit on the right-hand side needs to be taken in probability also.

Lemma 2.4.6 (Fatou's Inequality). *Suppose* $\{X_n\}_{n\in\mathbb{N}}$ *is a sequence of nonnegative integrable random variables. Then*

$$E[\liminf_n X_n|\mathcal{E}] \leq \liminf_n E[X_n|\mathcal{E}] \quad a.s.$$

Proof. As in the previous lemma, simply apply Fatou's inequality (Theorem 1.3.33) to $I_A X_n$ for all $A \in \mathcal{E}$. □

Lemma 2.4.7. *Suppose* X *is an integrable random variable and* $\mathcal{E}$ *is a sub-*σ*-algebra of* $\mathcal{F}$. *Then* $X = E[X|\mathcal{E}]$ *a.s. if and only if* X *is* $\mathcal{E}$*-measurable.*

Proof. By definition, $E[X|\mathcal{E}]$ is $\mathcal{E}$-measurable. Conversely, if X is $\mathcal{E}$-measurable then X satisfies Definition 2.3.1, so by uniqueness, $X = E[X|\mathcal{E}]$ a.s. □

Lemma 2.4.8 (Tower Property). *Suppose* $\mathcal{D}$ *and* $\mathcal{E}$ *are sub-*σ*-algebras of* $\mathcal{F}$ *such that* $\mathcal{D} \subseteq \mathcal{E} \subseteq \mathcal{F}$. *Then, for any integrable random variable* X,

$$E\big[E[X|\mathcal{E}]\big|\mathcal{D}\big] = E[X|\mathcal{D}] \quad a.s.$$

Proof. Again, the two sides are $\mathcal{D}$-measurable and have the same integral over any $A \in \mathcal{D}$. □

Corollary 2.4.9. *Suppose $\mathcal{D}$ is the trivial σ-algebra $\{\emptyset, \Omega\}$, so $E[X|\mathcal{D}]=E[X]$. Then $E\big[E[X|\mathcal{E}]\big|\mathcal{D}\big] = E[E[X|\mathcal{E}]] = E[X]$.*

Lemma 2.4.10 (Taking out what is known). *Suppose X is an integrable random variable and Y an $\mathcal{E}$-measurable random variable, such that the product XY is integrable. Then $E[XY|\mathcal{E}] = YE[X|\mathcal{E}]$ a.s.*

Proof. Suppose first that Y is a simple function, that is, Y takes only countably many values $\{a_i\}_{i\in\mathbb{N}}$. Write $A_i = Y^{-1}(a_i) \in \mathcal{E}$. Then for any $A \in \mathcal{E}$

$$\begin{aligned}\int_A E[XY|\mathcal{E}]\,dP &= \int_A XY\,dP = \sum_i \int_{A\cap A_i} a_i X\,dP = \sum_i \int_{A\cap A_i} a_i E[X|\mathcal{E}]\,dP \\ &= \int_A YE[X|\mathcal{E}]\,dP.\end{aligned}$$

Because an nonnegative random variable is the limit of a monotone increasing sequence of simple functions, the result follows from Lemma 2.4.3 and linearity. □

Lemma 2.4.11 (Jensen's Inequality). *Suppose ϕ is a convex map of $\mathbb{R}$ into $\mathbb{R}$ and suppose X is an integrable random variable such that $\phi \circ X$ is integrable. Then*

$$\phi\big(E[X|\mathcal{E}]\big) \leq E[\phi \circ X|\mathcal{E}] \quad \textit{a.s.}$$

Proof. By convexity, as $\phi(X)$ is finite, at least on the essential range of X, we know ϕ is continuous and is the upper envelope of a countable family of affine functions

$$\lambda_n(x) = \alpha_n x + \beta_n, \quad x \in \mathbb{R}, \quad n \in \mathbb{N},$$

that is, $\phi(x) = \sup_n\{\lambda_n(x)\}$. The random variables $\lambda_n \circ X$ are integrable and

$$\lambda_n \circ E[X|\mathcal{E}] = E[\lambda_n \circ X|\mathcal{E}] \leq E[\phi \circ X|\mathcal{E}] \quad \text{a.s.}$$

Taking the supremum with respect to n, the result follows. □

Remark 2.4.12. Taking $\mathcal{E} = \{\emptyset, \Omega\}$, we obtain $\phi\big(E[X]\big) \leq E[\phi(X)]$. For a general finite measure space (S, Σ, μ), by rescaling μ we see that

$$\phi\Big(\frac{1}{\mu(S)}\int_S X\,d\mu\Big) \leq \frac{1}{\mu(S)}\int_S \phi(X)d\mu.$$

Remark 2.4.13. Applying Jensen's inequality to the convex function

$$\phi(x) = |x|^p, \quad 1 \leq p < \infty,$$

we have

$$\|E[X|\mathcal{E}]\|_p \leq \|X\|_p$$

for $X \in L^p$. Also

$$\operatorname{ess\,sup} |E[X|\mathcal{E}]| \leq \operatorname{ess\,sup} |X|,$$

so

$$\|E[X|\mathcal{E}]\|_\infty \leq \|X\|_\infty.$$

The map $X \to E[X|\mathcal{E}]$ is, therefore, a continuous map, with operator norm ≤ 1, from L^p to itself for $1 \leq p \leq \infty$.

Lemma 2.4.14. *For X in L^2, the conditional expectation $Y = E[X|\mathcal{E}]$ is the unique (up to equality a.s.) $\mathcal{E}$-measurable random variable which minimizes*

$$E[(X - Y)^2].$$

Proof. Note that

$$E[(X - Y)^2] = E[X^2 - 2YX + Y^2].$$

As Y is $\mathcal{E}$-measurable, by the Tower property, minimizing this is equivalent to minimizing

$$Y^2 - 2YE[X|\mathcal{E}].$$

It is straightforward to show that this is achieved by setting $Y = E[X|\mathcal{E}]$. □

Remark 2.4.15. This result shows that $E[X|\mathcal{E}]$ can be thought of as the projection of X onto $L^2(\Omega, \mathcal{E}, P|_\mathcal{E})$, which forms a subspace of $L^2(\Omega, \mathcal{F}, P)$. In the extreme case where $\mathcal{E}$ is trivial, this shows that the expectation can be thought of as the least-squares approximation to X.

It is possible to use this fact to construct the conditional expectation without direct use of the Radon–Nikodym theorem, and to use this construction to prove the Radon–Nikodym theorem in turn. Doing so again rests on the properties of Hilbert spaces, in particular their completeness (which is embedded in the Riesz Representation theorem for Hilbert spaces, which we used to prove the Radon–Nikodym theorem).

2.5 Uniform Integrability

The important concept of uniform integrability is often not covered in measure theory courses. This is primarily a form of integrability for *collections* of random variables, (of particular interest are the properties of sequences of random variables). Uniform integrability ensures that the integrals of the random variables are *uniformly* well behaved. It is exceedingly useful when dealing with stochastic processes, particularly as it implies certain types of convergence, which do not follow from simple integrability.

Our presentation of this concept will be focussed on uniform integrability for random variables; however, there is a completely analogous theory in the context of general measurable functions.

Again $(\Omega, \mathcal{F}, P)$ is a probability space, and L^1 is the space of (equivalence classes of) real valued random variables X such that

$$\|X\|_1 = E\big[|X|\big] < \infty.$$

Definition 2.5.1. *Suppose $K \subset L^1(\Omega, \mathcal{F}, P)$. Then K is said to be a* uniformly integrable *subset of $L^1(\Omega, \mathcal{F}, P)$ if*

$$\int_{\{|X| \geq c\}} |X(\omega)| \, dP(\omega)$$

converges to 0 uniformly in $X \in K$ as $c \to +\infty$.

Remark 2.5.2. The key idea here is that the convergence must be uniform for $X \in K$, that is, for every $\epsilon > 0$, there exists a $c > 0$ such that

$$\sup_{X \in K} \left\{ \int_{\{|X| \geq c\}} |X(\omega)| \, dP(\omega) \right\} < \epsilon.$$

Equivalently, if X is a random variable and $c > 0$, define

$$X^c(\omega) := \begin{cases} X(\omega), & \text{if } |X(\omega)| \leq c, \\ 0, & \text{if } |X(\omega)| > c, \end{cases}$$

and

$$X_c(\omega) := X(\omega) - X^c(\omega) = I_{\{|X| \geq c\}}(\omega) X(\omega).$$

Then $K \subset L^1(\Omega, \mathcal{F}, P)$ is uniformly integrable if, and only if, for any $\epsilon > 0$, there is a $c > 0$ such that $\|X_c\|_1 < \epsilon$ for all $X \in K$.

Example 2.5.3. Let X be an integrable random variable. Then the set $K = \{Y : |Y| \leq |X|\}$ is uniformly integrable.

Theorem 2.5.4. *Suppose K is a subset of $L^1(\Omega, \mathcal{F}, P)$. Then K is uniformly integrable if and only if both*

(i) there is a number $k < \infty$ such that for all $X \in K$, $E\big[|X|\big] < k$, and
(ii) for any $\epsilon > 0$ there is a $\delta > 0$ such that, for all $A \in \mathcal{F}$ with $P(A) \leq \delta$, we have $\int_A |X(\omega)| \, dP(\omega) < \epsilon$ for all $X \in K$.

Proof. Necessity. Define X_c as in Remark 2.5.2. Note that for any integrable X, any set $A \in \mathcal{F}$ and any $c > 0$

$$\int_A |X(\omega)|\, dP(\omega) \le cP(A) + E\big[|X_c|\big].$$

Fix $\epsilon > 0$. If K is uniformly integrable, we can find a $c > 0$ such that $E\big[|X_c|\big] < \epsilon/2$ for all $X \in K$. Then

$$E\big[|X|\big] \le c + \epsilon/2$$

for all $X \in K$, establishing (i). For the same c, if $P(A) \le \delta = \epsilon/(2c)$ we have

$$\int_A |X(\omega)|\, dP(\omega) < \epsilon,$$

proving (ii).

Sufficiency. Fix $\epsilon > 0$ and suppose conditions (i) and (ii) are satisfied. There is then a $\delta > 0$ such that $\int_A |X(\omega)|\, dP(\omega) < \epsilon$ for all $A \in \mathcal{F}$ with $P(A) \le \delta$. Take

$$c = \delta^{-1} \sup_{X \in K} E\big[|X|\big] < \infty.$$

For each $X \in K$, let $A_X = \{|X| \ge c\}$, so that, by Markov's inequality (Exercise 2.7.3),

$$P(A_X) = P\big(\omega : |X(\omega)| \ge c\big) \le c^{-1} E\big[|X|\big] \le \delta.$$

Then

$$\int_{\{|X| \ge c\}} |X(\omega)|\, dP(\omega) = \int_{A_X} |X(\omega)|\, dP(\omega) < \epsilon,$$

for all $X \in K$, so K is uniformly integrable. □

Corollary 2.5.5 (de la Vallée Poussin Criterion). *Let K be a subset of $L^1(\Omega, \mathcal{F}, P)$. Suppose there is a positive function ϕ defined on $[0, \infty[$ such that $\lim_{t\to\infty} t^{-1}\phi(t) = +\infty$ and $\sup_{X \in K} E\big[\phi(|X|)\big] < \infty$. Then K is uniformly integrable.*

Proof. Write $\lambda = \sup_{X\in K} E[\phi \circ |X|]$ and fix $\epsilon > 0$. Put $a = \epsilon^{-1}\lambda$ and choose c large enough that $t^{-1}\phi(t) \ge a$ if $t \ge c$. Then, on the set $\{|X| \ge c\}$, we have

$$|X| \le a^{-1}(\phi \circ |X|),$$

so

$$\int_{\{|X|\ge c\}} |X(\omega)|\, dP(\omega) \le a^{-1} \int_{\{|X|\ge c\}} (\phi \circ |X|)\, dP \le a^{-1} E\big[\phi \circ |X|\big] \le \epsilon.$$

Therefore, K is uniformly integrable. □

Remark 2.5.6. A common application of the above result is when $\phi(x) = x^p$, for $p > 1$. Then, if K is a subset of L^p with $\sup_{X \in K} E[X^p] < \infty$, we know K is uniformly integrable.

We have seen various notions of convergence for sequences of random variables. Applying the notion of convergence in measure (cf. Definition 1.3.37), we obtain the concept of convergence in probability.

Definition 2.5.7. *A sequence of random variables $\{X_n\}_{n\in\mathbb{N}}$ is said to converge* in probability *to X if, for all $\epsilon > 0$, there is an N such that, for all $n \geq N$,*

$$P\big(\{\omega : |X_n(\omega) - X(\omega)| \geq \epsilon\}\big) < \epsilon.$$

In general, almost sure convergence of a sequence $\{X_n\}_{n\in\mathbb{N}}$ of random variables neither implies nor is implied by convergence in L^1. We have already seen (Lemma 1.3.38) that any sequence converging almost surely converges in probability, and any sequence converging in probability has a subsequence which converges almost surely.

The following theorem states that, given convergence in probability, uniform integrability and convergence in $L^1(\Omega, \mathcal{F}, P)$ are equivalent. One can see, from Example 2.5.3, that the dominated convergence theorem is a special case of this result.

Theorem 2.5.8 (Vitali Convergence Theorem). *Suppose $\{X_n\}_{n\in\mathbb{N}}$ is a sequence of integrable random variables which converge in probability to a random variable X. Then the following are equivalent:*

(i) X_n converges to X in the norm of L^1,
(ii) the collection $K = \{X_n\}_{n\in\mathbb{N}}$ is uniformly integrable.

In either case, the limit X is also integrable.

Proof. (i $\Rightarrow$ ii) Suppose $X_n \to X$ in the norm of L^1, so that X itself is in L^1. For any n, $\|X_n\|_1 \leq \|X_n - X\|_1 + \|X\|_1$, and we see that the expectations $E\big[|X_n|\big] = \|X_n\|_1$ are uniformly bounded.

For any $\epsilon > 0$, let N be such that

$$\|X_n - X\|_1 < \epsilon/3$$

for all $n \geq N$. For any $n \geq N$ and any set $A \in \mathcal{F}$, this implies

$$\int_A |X_n|\,dP < \int_A |X|\,dP + \|X_n - X\|_1 < \int_A |X|\,dP + \epsilon/3.$$

For any $n < N$ and any set $A \in \mathcal{F}$,

$$\begin{aligned}\int_A |X_n|\,dP &\leq \int_A |X|\,dP + \int_A |X_n - X_N|\,dP + \|X_N - X\|_1 \\ &< \int_A |X|\,dP + \int_A |X_n - X_N|\,dP + \epsilon/3.\end{aligned}$$

As X is integrable, we can find a $\delta_\infty > 0$ such that $\int_A |X|\,dP < \epsilon/3$ whenever $P(A) \leq \delta_\infty$. Similarly, for each $n \leq N$ we can find a $\delta_n > 0$ such that $\int_A |X_n - X_N|\,dP < \epsilon/3$ whenever $P(A) \leq \delta_n$. Let $\delta = \delta_\infty \wedge \min_{n \leq N} \delta_n$. Then, whenever $P(A) < \delta$, we have $\int_A |X_n|\,dP < \epsilon$. By Theorem 2.5.4, this shows that $\{X_n\}_{n\in\mathbb{N}}$ is uniformly integrable.

(ii $\Rightarrow$ i) Conversely, suppose the set $\{X_n\}_{n\in\mathbb{N}}$ is uniformly integrable. Then the set of expectations $E[|X_n|]$ is bounded and so, by Fatou's inequality (Theorem 1.3.33) applied to an almost surely converging subsequence,

$$E[|X|] = E[\lim_n |X_n|] \leq \liminf_n E[|X_n|] < \infty.$$

Now, using the notation of Remark 2.5.2,

$$\|X_n - X\|_1 \leq \|(X_n)^c - X^c\|_1 + \|(X_n)_c\|_1 + \|X_c\|_1.$$

Fix $\epsilon > 0$. Because the collection $\{X_n\}_{n\in\mathbb{N}}$ is uniformly integrable, there exists a number $c > 0$ such that $\|X_c\|_1 < \epsilon/3$ and $\|(X_n)_c\|_1 < \epsilon/3$ for all n. We know that X_n^c converges to X^c in probability and $|X_n^c - X^c| \leq 2c$ so by Lebesgue's dominated convergence theorem (Lemma 1.3.39), $\lim_n \|X_n^c - X^c\|_1 = 0$.

There is, therefore, an integer N such that $\|X_n^c - X^c\|_1 \leq \epsilon/3$ if $n > N$. Consequently, if $n > N$, we have $\|X_n - X\| < \epsilon$, and $X_n \to X$ in L^1. Because

$$\big|\, \|X_n\|_1 - \|X\|_1 \,\big| \leq \|X_n - X\|_1,$$

$E\big[|X_n|\big]$ converges to $E\big[|X|\big]$. □

Corollary 2.5.9. *Let $\{X_n\}_{n\in\mathbb{N}}$ be as in Theorem 2.5.8. If $X_n \geq 0$ a.s. for each $n \in \mathbb{N}$, it is necessary and sufficient for convergence in L^1 (and hence uniform integrability) that* $\lim_n E[X_n] = E[X] < \infty$.

Proof. We show that $\lim_n E[X_n] = E[X] < \infty$ implies convergence in L^1. The converse is easy. Suppose that, for each n, $X_n \geq 0$ and that $\lim_n E[X_n] = E[X] < \infty$. Now

$$X_n + X = (X \vee X_n) + (X \wedge X_n) \quad \text{and} \quad |X_n - X| = (X \vee X_n) - (X \wedge X_n).$$

By dominated convergence,

$$\lim_n E[X \wedge X_n] = E[X].$$

Also, by hypothesis,

$$\lim_n E[X + X_n] = 2E[X].$$

Consequently, $\lim_n E[X \vee X_n] = E[X]$, and so

$$\lim_n \|X_n - X\|_1 = \lim_n E\big[|X_n - X|\big] = E[X] - E[X] = 0.$$

□

Theorem 2.5.10. *Let $X \in L^1(\Omega, \mathcal{F}, P)$ and $\mathfrak{G}$ be a (possibly uncountable) family of sub-σ-algebras of $\mathcal{F}$. Then the family of random variables $\{E[X|\mathcal{G}]\}_{\mathcal{G}\in\mathfrak{G}}$ is uniformly integrable.*

Proof. We prove this using Theorem 2.5.4. From Jensen's inequality, we know that for any $A \in \mathcal{F}$,

$$E\big[I_A|E[X|\mathcal{G}]|\big] \leq E[I_A|X|] \quad \text{for all } \mathcal{G} \in \mathfrak{G}.$$

Setting $A = \Omega$, we obtain a uniform bound on $E\big[|E[X|\mathcal{G}]|\big]$. For each $\delta > 0$, let $A_\delta(\mathcal{G})$ be the largest set of the form $\{|E[X|\mathcal{G}]| > k\}$ such that $P(A_\delta(\mathcal{G})) \leq \delta$, that is,

$$A_\delta(\mathcal{G}) = \bigcup_{\{k: P(|E[X|\mathcal{G}]|>k)\leq\delta\}} \{\omega : |E[X|\mathcal{G}]| > k\}.$$

Note that $A_\delta(\mathcal{G}) \in \mathcal{G}$ and, by construction, for $A \in \mathcal{F}$ with $P(A) \leq \delta$, we have $E\big[I_A|E[X|\mathcal{G}]|\big] \leq E\big[I_{A_\delta(\mathcal{G})}|E[X|\mathcal{G}]|\big]$. For any $\epsilon > 0$, we can find a $\delta > 0$ such that $E[I_{A_\delta(\mathcal{F})}|X|] < \epsilon$, and hence, for any $A \in \mathcal{F}$ with $P(A) \leq \delta$,

$$\begin{aligned} E\big[I_A|E[X|\mathcal{G}]|\big] &\leq E\big[I_{A_\delta(\mathcal{G})}|E[X|\mathcal{G}]|\big] \leq E[I_{A_\delta(\mathcal{G})}|X|] \\ &\leq E[I_{A_\delta(\mathcal{F})}|X|] < \epsilon \end{aligned}$$

for all $\mathcal{G} \in \mathfrak{G}$. By Theorem 2.5.4, we see that the family $\{E[X|\mathcal{G}]\}_{\mathcal{G}\in\mathfrak{G}}$ is uniformly integrable. □

The following theorem, which we present without proof, shows that uniformly integrable sets are the weakly relatively compact sets in L^1, and fills a gap left in Theorem 1.7.19. A proof can be found in Royden and Fitzpatrick [160, p.412].

Theorem 2.5.11 (Dunford–Pettis Theorem). *For a probability space (or more generally, a finite measure space) and a sequence $\{X_n\}_{n\in\mathbb{N}}$ bounded in L^1, the following are equivalent.*

(i) $\{X_n\}_{n\in\mathbb{N}}$ *is uniformly integrable,*
(ii) Every subsequence of $\{X_n\}_{n\in\mathbb{N}}$ *has a further subsequence* $\{X_{n_k}\}_{n\in\mathbb{N}}$ *that converges weakly in* L^1*, that is, there exists* $X \in L^1$ *such that* $E[X_{n_k}Y] \to E[XY]$ *for all bounded random variables* Y*.*

In other words, a set is uniformly integrable if and only if it is weakly relatively compact in L^1*.*

In addition to the notions already considered, we also have the concept of convergence in distribution for a sequence of random variables, which can be defined as follows.

Definition 2.5.12. *A sequence of random variables* $\{X_n\}_{n\in\mathbb{N}}$ *is said to converge* in distribution *or* in law *if the functions* $F_{X_n}(x) := P(X_n \leq x)$ *converge pointwise, for* dx*-almost all* x*.*

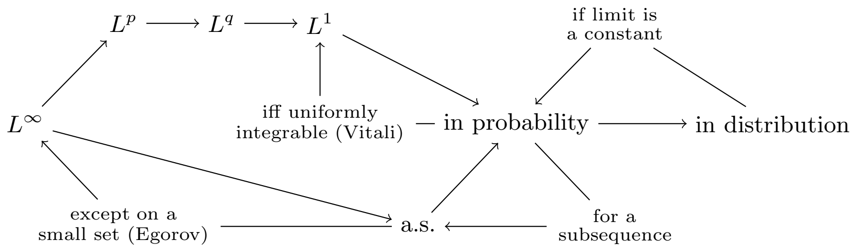

Fig. 2.1. Relations between notions of convergence, where $p \geq q$. Names in parentheses refer to the relevant theorem.

For a probability measure (or more generally, for any finite measure) we have the following relations between the different notions of convergence considered (Fig. 2.1). For simplicity, we present these in the following diagram. Some of these notions we have proven, others we leave as an exercise. By a small set, we mean one with $P(A) < \epsilon$, for any fixed $\epsilon > 0$.

2.6 Regular Conditional Probability

Before finishing this chapter, we consider the problem of defining conditional probability distributions. This is significantly more delicate than defining a conditional expectation, and depends in a fine way on how the σ-algebra on Ω has been obtained. The general theory which underlies these concerns is treated in some detail in Bogachev [21]; however we shall restrict our attention to a relatively simple case.

Definition 2.6.1. *Let Ω be a space with a σ-algebra $\mathcal{F}$ and a σ-algebra $\mathcal{G}$. Let μ be a signed measure on $(\Omega, \mathcal{F} \vee \mathcal{G})$. We say that a function*

$$\mu|_{\mathcal{G}} : \mathcal{F} \times \Omega \to \mathbb{R}$$

is a regular conditional measure on $\mathcal{F}$ with respect to $\mathcal{G}$ if

(i) for every ω, the function $\mu|_{\mathcal{G}}(\cdot, \omega)$ is a measure on $\mathcal{F}$,
(ii) for every $A \in \mathcal{F}$, the function $\mu|_{\mathcal{G}}(A, \cdot)$ is $\mathcal{G}$-measurable and $|\mu|$-integrable,
(iii) for all $A \in \mathcal{F}$, $B \in \mathcal{G}$,

$$\mu(A \cap B) = \int_B \mu_{\mathcal{G}}(A, \omega)|\mu|(d\omega).$$

Remark 2.6.2. If $\mu = P$ is a probability measure, then we have already considered how to use the Radon–Nikodym theorem to define $P(A|\mathcal{G}) := E[I_A|\mathcal{G}]$. The problem is to guarantee that this defines a *measure* (in particular a probability measure), that is, it is *σ-additive and defined simultaneously for almost* *all* ω. We can easily see that, for any disjoint sequence $\{A_n\}_{n\in\mathbb{N}}$, by monotone convergence we have

$$\sum_n P(A_n|\mathcal{G}) = \sum_n E[I_{A_n}|\mathcal{G}] = E\Big[\sum_n I_{A_n}\Big|\mathcal{G}\Big] = E[I_{\cup_n A_n}|\mathcal{G}] = P(\cup_n A_n|\mathcal{G}),$$

which might suggest that this is trivially true. However, $E[I_A|\mathcal{G}]$ is only defined P-almost everywhere, so different sequences $\{A_n\}_{n\in\mathbb{N}}$ may have different null sets on which the limit above fails to hold. As we want $\mu|_\mathcal{G}$ to be a measure for every ω (or at least for μ-almost all ω, by changing the value of $\mu|_\mathcal{G}$ on a null set), we need to rely on some finer analysis.

Definition 2.6.3. *Let μ be a nonnegative set function on a class $\mathcal{F}$ of subsets of a set Ω. Let $\mathcal{K}$ be another class of subsets of Ω.*

(i) We say that $\mathcal{K}$ is an approximating class for μ if, for any $\epsilon > 0$ and any $A \in \mathcal{F}$, there exist $C_\epsilon \in \mathcal{K}$ and $A_\epsilon \in \mathcal{F}$ such that $A_\epsilon \subseteq C_\epsilon \subseteq A$ and $|\mu(A) - \mu(A_\epsilon)| < \epsilon$.
(ii) We say that $\mathcal{K}$ is a compact class if for any sequence K_n of its elements with $\bigcap_{n\in\mathbb{N}} K_n = \emptyset$, there exists N such that $\bigcap_{n=1}^N K_n = \emptyset$.

Remark 2.6.4. It is easy to show that an arbitrary family of compact sets (in a general topological space) is a compact class, see Bogachev [21, p.13]. One can also prove that if $\mathcal{F}$ is a σ-algebra, $\mathcal{K}$ is a compact approximating class, $\mathcal{K} \subset \mathcal{F}$ and μ is additive (and hence regular, in the sense of Lemma A.2.3), then μ is countably additive.

The following theorem is not the most general which is possible, however is sufficient for many applications. A full proof of the theorem is best understood within a general framework of measure theory for metric spaces, and so we omit the details.

Theorem 2.6.5. *Let Ω be a Souslin space*[1] *with its Borel σ-algebra $\mathcal{B}(\Omega)$. Let μ be a (countably additive) finite measure on $\mathcal{B}(\Omega)$. Then there exists a compact approximating class for μ.*

Proof. See, for example, Bogachev [21, Chapter 7]. □

Remark 2.6.6. In Lemma A.2.3, we show directly that this result holds for $\Omega = \mathbb{R}$.

[1] Recall that a Polish space is a separable completely metrizable topological space, that is, a space with a countable dense subset, where the topology can be generated by some (unspecified) metric on the space, which is complete with respect to this metric. A Souslin space is a space which can be obtained as the image (in a Hausdorff space) of a Polish space under some continuous mapping. Most familiar spaces, for example $\mathbb{R}^n$, or any separable Banach space, fall into these categories.

The result of this theorem is particularly important in our context, as it allows us to apply the following existence result.

Theorem 2.6.7. *Consider a (countably additive signed) finite measure μ on a measurable space $(\Omega, \mathcal{F})$.*

(i) Suppose that $\mathcal{F}$ is countably generated (that is, there exists a sequence of sets $\{A_n\}_{n\in\mathbb{N}}$ such that $\mathcal{F} = \sigma(\{A_n\}_{n\in\mathbb{N}})$) and that μ has a compact approximating class in $\mathcal{F}$. Then for any sub-σ-algebra $\mathcal{G}$ of $\mathcal{F}$, there exists a regular conditional measure $\mu|_{\mathcal{G}}$ on $\mathcal{F}$.

(ii) More generally, let $\tilde{\mathcal{F}}$ be a sub-σ-algebra of $\mathcal{F}$ generated by a countable algebra of sets $\mathcal{U}$. Suppose that there is a compact class $\mathcal{K}$ such that for every $A \in \mathcal{U}$ and $\epsilon > 0$, there exist $K_\epsilon \in \mathcal{K}$ and $A_\epsilon \in \mathcal{F}$ with $A_\epsilon \subseteq K_\epsilon \subseteq A$ and $|\mu|(A \setminus A_\epsilon) < \epsilon$. Then, for every sub-$\sigma$-algebra $\mathcal{G} \subseteq \mathcal{F}$, there exists a regular conditional measure $\mu|_{\mathcal{G}}$ on $\mathcal{F}$ with respect to $\mathcal{G}$ (which can be taken to be a probability measure if μ is nonnegative).
In addition, for every $\tilde{\mathcal{F}}$-measurable μ-integrable function f, one has

$$\int_\Omega f\, d\mu = \int_\Omega \int_\Omega f(\omega')\mu|_{\mathcal{G}}(d\omega', \omega)|\mu|(d\omega).$$

Proof. See Appendix A.3. □

Remark 2.6.8. As mentioned before, Souslin spaces are not the most general class of spaces in which these results hold. Blackwell [18] considers spaces $(\Omega, \mathcal{G})$ which he calls Lusin spaces, by which he means that $\mathcal{G}$ is countably generated and $f(\Omega)$ is an analytic set[2] for every $\mathcal{G}$-measurable real-valued function f. Dellacherie and Meyer [54] call such spaces *Blackwell spaces* (to distinguish from classical Lusin spaces); however, this differs from the alternative definition of a Blackwell space as given in, for example, Bogachev [21] and references therein.

Remark 2.6.9. As discussed by Blackwell and Dubins [19], some intuitively reasonable properties for the regular conditional measures frequently fail. For example, it is not typically the case that $\mu|_{\mathcal{G}}(\omega, A) = 1$ for all $\omega \in A$.

2.7 Exercises

Exercise 2.7.1. Suppose $\Omega = [0, 1]$, $\mathcal{F}$ is the Borel σ-algebra on $[0, 1]$, and P is Lebesgue measure. If $f(x) = x^2/2$, $g(x) = 2(x - 1/2)^2$, $\mathcal{E} = \sigma(f)$ and $\mathcal{D} = \sigma(g)$, find $E[g|\mathcal{E}]$ and $E[f|\mathcal{D}]$.

---

[2] In this context, an analytic set is a continuous image of a Polish space. Equivalently, $f(\Omega)$ is a Souslin space which is a subset of the real line.

Exercise 2.7.2. Show that $L^2(\Omega, \mathcal{F}, P)$ is a Hilbert space with inner product $\langle X, Y\rangle = E[XY]$ and if $\mathcal{E}$ is a sub-σ-algebra of $\mathcal{F}$, then $L^2(\Omega, \mathcal{E}, P)$ is a subspace of $L^2(\Omega, \mathcal{F}, P)$. Show that if X is a random variable in $L^2(\Omega, \mathcal{F}, P)$, then $E[X|\mathcal{E}]$ is the orthogonal projection of X onto the subspace $L^2(\Omega, \mathcal{E}, P)$.

Exercise 2.7.3. For $X \in L^1(\Omega, \mathcal{F}, P)$, prove Markov's inequality:

$$P\big(|X| \geq k\big) \leq k^{-1}E\big[|X|\big] \quad \text{for all } k \in \mathbb{R}.$$

Exercise 2.7.4. For $X \in L^2(\Omega, \mathcal{F}, P)$, prove Chebyshev's inequality:

$$P\big(|X - E[X]| \geq k\big) \leq k^{-2}\mathrm{Var}(X), \quad \text{for all } k \in \mathbb{R}.$$

where $\mathrm{Var}(X) = E\big[(X - E[X])^2\big]$.

Exercise 2.7.5. Give counterexamples which show that, in general, almost sure convergence of a sequence $\{X_n\}_{n\in\mathbb{N}}$ of random variables neither implies nor is implied by convergence of $\{X_n\}_{n\in\mathbb{N}}$ in L^1.

Exercise 2.7.6. Show that L^1 convergence or almost sure convergence implies convergence in probability. Give counterexamples which show that the converse is not true in either case.

Exercise 2.7.7. Suppose X, Y are two random variables in $L^2(\Omega, \mathcal{F}, P)$, with $E[X|Y] = Y$ and $E[Y|X] = X$. Show that $X = Y$ P-a.s.
Extension: Show this for X, Y in $L^1(\Omega, \mathcal{F}, P)$.

Exercise 2.7.8. Suppose $\{X_n\}_{n\in\mathbb{N}}$ is a sequence of random variables converging in $L^1(\Omega, \mathcal{F}, P)$ to X. Show that for any σ-algebra $\mathcal{G} \subseteq \mathcal{F}$, $E[X_n|\mathcal{G}]$ also converges in L^1 to $E[X|\mathcal{G}]$.

Exercise 2.7.9. Suppose K is a uniformly integrable family of random variables and J is another family of random variables such that, for every $X \in J$, there is $Y \in K$ such that $|X| \leq Y$. Show that J is uniformly integrable.

Exercise 2.7.10. Show that, if two random variables X, Y are independent, then $E[X|Y] = E[X]$ and hence $E[XY] = E[X]E[Y]$. Give a counterexample to the converse statement.

Exercise 2.7.11. For P, Q equivalent probability measures on $(\Omega, \mathcal{F})$, a common quantity considered in information theory is the relative entropy $H(P, Q) = E_P[-\log(\frac{dP}{dQ})]$. Show that $H(P, Q) \geq 0$ with equality if and only if $P = Q$.

Exercise 2.7.12. Show that random variables X, Y are independent if and only if for every Borel measurable g, $E[g(X)|Y] = E[g(X)]$ a.s.

Exercise 2.7.13. For X and Y random variables, let F_X, F_Y be the distribution functions of their laws. Show that X and Y are independent if and only if $P(X \leq x, Y \leq y) = F_X(x)F_Y(y)$ for all $x, y \in \mathbb{R}$.

Exercise 2.7.14. Let $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}(\Omega)$, P be Lebesgue measure and Q be defined by the Radon–Nikodym derivative $dQ/dP = 2\omega$. Let $X_n(\omega) := (n\omega)^{-1}$. Show that $X_n \to 0$ in $L^1(Q)$ and almost surely, but not in $L^1(P)$. What does this imply about the dependence of uniform integrability on the choice of measure?

Exercise 2.7.15. Suppose rainfall on a given day has a 20% chance of being zero and an 80% chance of being exponentially distributed with parameter λ. Describe the law of the amount of rainfall X (for example, by writing down its distribution function). Now suppose $\mathcal{G} = \sigma(\{X > 0\})$ (note that $\mathcal{G}$ is a σ-algebra, not an event). Describe the random variable $E[X^2|\mathcal{G}]$.

Part II

Stochastic Processes

3

Filtrations, Stopping Times and Stochastic Processes

In many situations, we have more than a single random variable to consider. In particular, we may have new observations at different points in time, each of which is random. Our goal in this section is to build a mathematical understanding of these 'stochastic processes', that is, of collections of random variables, the values of which become revealed through time.

To understand this, we need to carefully model the flow of information – we wish to model the fact that we usually know the values of random outcomes in the past, but not of those in the future. We do this using the concepts of σ-algebras developed in the previous chapters.

3.1 Filtrations and Stopping Times

Suppose $(\Omega, \mathcal{F})$ is a measurable space. We wish to model the development in time of information about some random phenomenon. This is done by considering an increasing family of sub-σ-algebras of $\mathcal{F}$.

Definition 3.1.1. *Let* $\mathbb{T}$ *denote the* time index set*, that is, the collection of times at which we observe random outcomes. For our purposes, we shall assume* $\mathbb{T}$ *is either* $\overline{\mathbb{R}}^+ = [0, \infty]$ *or* $\mathbb{R}^+ = [0, \infty[$ *(*continuous time*) or* $\mathbb{T} = \overline{\mathbb{Z}}^+ = \{0, 1, 2, \dots, \infty\}$ *or* $\mathbb{Z}^+ = \{0, 1, 2, \dots\}$ *(*discrete time*).*

Definition 3.1.2. *A* filtration $\{\mathcal{F}_t\}_{t \in \mathbb{T}}$ *of* $(\Omega, \mathcal{F})$ *is a family of sub-σ-algebras of* $\mathcal{F}$ *such that if* $s \leq t$ *then* $\mathcal{F}_s \subseteq \mathcal{F}_t$*.*

Remark 3.1.3. The family of σ-algebras $\{\mathcal{F}_t\}_{t \in \mathbb{T}}$ can be considered as describing the history of some phenomenon. For this reason, $\mathcal{F}_t$ is sometimes called the σ-algebra of events up to time t.

S.N. Cohen, R.J. Elliott, *Stochastic Calculus and Applications*,
Probability and Its Applications, DOI 10.1007/978-1-4939-2867-5_3

A probability space with a filtration will, unsurprisingly, be called a filtered probability space.

Definition 3.1.4. *Given a filtration $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$, we define*

$$\mathcal{F}_{t+} = \bigcap_{s>t} \mathcal{F}_s,$$

the σ-algebra of events immediately after t*, and, for* $t > 0$*,*

$$\mathcal{F}_{t-} = \bigvee_{s<t} \mathcal{F}_s,$$

the σ-algebra of events strictly prior to t*.*

The filtration $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$ *is said to be* right-continuous *if* $\mathcal{F}_t = \mathcal{F}_{t+}$ *for all* t*. (Note that the filtration* $\{\mathcal{F}_{t+}\}_{t\in\mathbb{T}}$ *is always right-continuous.)*

Remark 3.1.5. In discrete time, (when $\mathbb{T} = \{0, 1, 2, \dots, \infty\}$), $\mathcal{F}_{n+} = \mathcal{F}_{n+1}$ (so a right-continuous filtration is constant), and $\mathcal{F}_{n-} = \mathcal{F}_{n-1}$.

In continuous time, a filtration is right continuous if there is no information available to you immediately after time t (that is, at times $t+\epsilon$ for all ϵ) which is not already available at time t.

Definition 3.1.6. *Suppose* $(\Omega, \mathcal{F})$ *is a measurable space with a filtration* $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$*. A random variable* $T : \Omega \to \mathbb{T} \cup \{\infty\}$ *is called a random time. A random time is said to be a* stopping time *with respect to the filtration if*

$$\{T \le t\} = \{\omega : T(\omega) \le t\} \in \mathcal{F}_t \quad \textit{for every} \quad t \in \mathbb{T}.$$

Intuitively, a stopping time is a random time T such that the event "T has occurred by time t" is known at time t. That is, it depends only on the history up to time t, and not on any information about what happens after time t.

For example, suppose you are driving down a road without a map. If T is the time you need to turn, then the instruction "take the third turn on the right" gives a stopping time – when you reach the turn, you will know that you have reached it. On the other hand, the instruction "take the second-to-last turn on the left" does not give a stopping time – to know whether to turn, you need to know how many roads remain on the street, which you will not know until you drive past them!

Example 3.1.7. Some intuitive examples of stopping times:

(i) When repeatedly tossing a coin, with the filtration given by observing the sequence of outcomes, the time when the third head is observed is a stopping time. The time when 'the last tail before the fifth head' is observed is not a stopping time (as you would need to know the value of the next coin toss).

(ii) The time during a day when a stock price reaches its maximum is not (in general) a stopping time in the filtration of information available from observing market prices. The first time the price reaches \$10 is a stopping time.

(iii) A constant random variable $T = t \in \mathbb{T}$ is a stopping time.

(iv) If T is a stopping time and $s \in \mathbb{T}$ then $T + s$ is a stopping time.

Lemma 3.1.8. *For $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$ a filtration, if T is a stopping time with respect to $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$, then $\{T < t\} \in \mathcal{F}_t$ for all $t \in \mathbb{T}$. The converse is true if $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$ is right-continuous.*

Proof. If T is a stopping time, then $\{T \le t - \frac{1}{n}\} \in \mathcal{F}_{t-\frac{1}{n}} \subseteq \mathcal{F}_t$ for any $n \ge 1$. Since

$$\{T < t\} = \bigcup_{n=1}^{\infty} \Big\{T \le t - \frac{1}{n}\Big\},$$

the result follows.

Conversely, given right-continuity, if $\{T < t\} \in \mathcal{F}_t$ for all $t \ge 0$, then $\{T \le t\} \in \mathcal{F}_{t+\epsilon}$ for any $\epsilon > 0$. Thus $\{T \le t\} \in \mathcal{F}_{t+} = \mathcal{F}_t$, so T is a stopping time. □

Lemma 3.1.9. *Suppose S and T are stopping times. Then $S \wedge T = \min\{S, T\}$ and $S \vee T = \max\{S, T\}$ are stopping times. If $\{T_n\}_{n\in\mathbb{N}}$, is a sequence of stopping times, then $\bigvee_n T_n = \sup_n\{T_n\}$ is a stopping time. If $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$ is right-continuous then $\bigwedge_n T_n = \inf_n\{T_n\}$ is also a stopping time.*

Proof. To prove the first part, we simply note that

$$\{S \wedge T \le t\} = \{S \le t\} \cup \{T \le t\} \in \mathcal{F}_t,$$
$$\{S \vee T \le t\} = \{S \le t\} \cap \{T \le t\} \in \mathcal{F}_t.$$

For the second assertion, note that

$$\Big\{\bigvee_n T_n \le t\Big\} = \bigcap_{n=1}^{\infty} \{T_n \le t\} \in \mathcal{F}_t,$$

and

$$\Big\{\bigwedge_n T_n < t\Big\} = \bigcup_{n=1}^{\infty} \{T_n < t\} \in \mathcal{F}_t$$

by Lemma 3.1.8. Thus $\{\bigwedge_n T_n \le t\} \in \mathcal{F}_t$, again by Lemma 3.1.8, since $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$ is right-continuous. □

Just as $\mathcal{F}_t$ represents the information available at time t, we wish to define $\mathcal{F}_T$, the information available at a stopping time T. These will be the events A where the occurrence of A will be known at time t, provided $T \le t$, that is, when we have reached the stopping time T at or before t.

For example, when repeatedly tossing a coin, let T be the time we observe the first head. Clearly the number of tails observed prior to stopping should be known at time T, however, this does not mean that there is any fixed time t when we can be sure of knowing its value. Nevertheless, for every t, if we have stopped by time t, then, at that time we know the number of tails observed.

Definition 3.1.10. *Suppose T is a stopping time with respect to the filtration $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$. Then the σ-algebra $\mathcal{F}_T$ of events occurring up to time T is the σ-algebra consisting of those events $A \in \mathcal{F}$ such that*

$$A \cap \{T \leq t\} \in \mathcal{F}_t \quad \text{for every} \quad t \in \mathbb{T}.$$

Remark 3.1.11. Note that T is $\mathcal{F}_T$-measurable and if $T = t$ then $\mathcal{F}_T = \mathcal{F}_t$ (Exercise 3.4.4). Also, by Exercise 3.4.5, for T a.s. finite, we can equally define $\mathcal{F}_T$ as the σ-algebra of events $A \in \mathcal{F}_\infty$ (rather than $A \in \mathcal{F}$) such that $A \cap \{T \leq t\} \in \mathcal{F}_t$.

Lemma 3.1.12. *The collection of sets $\mathcal{F}_T$, as defined in Definition 3.1.10, is a σ-algebra.*

Proof. Clearly $\emptyset \in \mathcal{F}_T$. As $A^c \cap \{T \leq t\} = \{T \leq t\} \setminus (A \cap \{T \leq t\})$ and T is a stopping time, we know $A^c \in \mathcal{F}_T$ for all $A \in \mathcal{F}_T$. Finally if $A_i \in \mathcal{F}_T$ for all $i \in \mathbb{N}$, then $(\cup_{i\in\mathbb{N}} A_i) \cap \{T \leq t\} = \cup_{i\in\mathbb{N}}(A_i \cap \{T \leq t\})$, so $\cup_{i\in\mathbb{N}} A_i \in \mathcal{F}_T$. □

Theorem 3.1.13. *Suppose S and T are stopping times.*

(i) If $S \leq T$ then $\mathcal{F}_S \subseteq \mathcal{F}_T$.
(ii) If $A \in \mathcal{F}_S$ then $A \cap \{S \leq T\} \in \mathcal{F}_T$.

Proof. (i) Suppose $B \in \mathcal{F}_S$ and $t \in \mathbb{T}$. Then

$$B \cap \{T \leq t\} = B \cap \{S \leq t\} \cap \{T \leq t\} \in \mathcal{F}_t.$$

(ii) Suppose $A \in \mathcal{F}_S$. Then

$$\begin{aligned} &A \cap \{S \leq T\} \cap \{T \leq t\} \\ &\qquad = (A \cap \{S \leq t\}) \cap \{T \leq t\} \cap \{S \wedge t \leq T \wedge t\}. \end{aligned}$$

Each of these three sets is in $\mathcal{F}_t$: the first because $A \in \mathcal{F}_S$, the second because T is a stopping time, and the third because $S \wedge t$ and $T \wedge t$ are $\mathcal{F}_t$-measurable random variables.

□

Lemma 3.1.14. *If S and T are stopping times, then*

$$\mathcal{F}_{S\wedge T} = \mathcal{F}_S \cap \mathcal{F}_T.$$

Proof. Since $S \wedge T \leq S$ and $S \wedge T \leq T$, by Theorem 3.1.13(i)

$$\mathcal{F}_{S\wedge T} \subseteq \mathcal{F}_S \cap \mathcal{F}_T.$$

Now suppose $A \in \mathcal{F}_S \cap \mathcal{F}_T$. Then

$$\begin{aligned} A \cap \{S \wedge T \leq t\} &= A \cap (\{S \leq t\} \cup \{T \leq t\}) \\ &= (A \cap \{S \leq t\}) \cup (A \cap \{T \leq t\}) \in \mathcal{F}_t, \end{aligned}$$

so $A \in \mathcal{F}_{S\wedge T}$. The result follows. □

Theorem 3.1.15. *Suppose S and T are stopping times. Then the events $\{S < T\}$, $\{S = T\}$ and $\{S > T\}$ belong to both $\mathcal{F}_S$ and $\mathcal{F}_T$.*

Proof. From part (ii) of Theorem 3.1.13, we have $\{S \leq T\} \in \mathcal{F}_T$. By Lemma 3.1.9 and Remark 3.1.11, $S \wedge T$ is a stopping time that is $\mathcal{F}_{S\wedge T}$-measurable. □

Lemma 3.1.16. *For any integrable random variable X, any stopping times S and T,*

$$I_{\{S\leq T\}} E[X|\mathcal{F}_S] = I_{\{S\leq T\}} E[X|\mathcal{F}_{S\wedge T}].$$

Proof. As $\mathcal{F}_{S\wedge T} \subseteq \mathcal{F}_S$, and $I_{\{S<T\}}$ is $\mathcal{F}_{S\wedge T}$-measurable,

$$I_{\{S\leq T\}} E[X|\mathcal{F}_{S\wedge T}] = E\big[I_{\{S\leq T\}} E[X|\mathcal{F}_S]\big|\mathcal{F}_{S\wedge T}\big].$$

Therefore, it is enough to show that $I_{\{S\leq T\}} E[X|\mathcal{F}_S]$ is $\mathcal{F}_{S\wedge T}$-measurable. This follows from Theorem 3.1.13(ii), and the $\mathcal{F}_S$-measurability of $E[X|\mathcal{F}_S]$. □

Lemma 3.1.17. *Suppose $\mathbb{T} = \{0, 1, \ldots, \infty\}$ and $\{\mathcal{F}_n\}_{n\in\mathbb{T}}$ is a discrete-time filtration of the measurable space $(\Omega, \mathcal{F})$. If $\{X_n\}_{n\in\mathbb{T}}$ is a sequence of random variables such that each X_n is $\mathcal{F}_n$-measurable, then the random variable $X_T = X_{T(\omega)}(\omega)$ is $\mathcal{F}_T$-measurable.*

Proof. We must show that $\{X_T \in A\} \in \mathcal{F}_T$ for every $A \in \mathcal{B}(\mathbb{R})$. That is, we must show that for each $n \in \mathbb{T}$, $\{X_T \in A\} \cap \{T \leq n\} \in \mathcal{F}_n$.

Observe that

$$\{X_T \in A\} \cap \{T \leq n\} = \bigcup_{k=1}^{n} \{X_T \in A\} \cap \{T = k\}$$

and the result follows from

$$\{X_T \in A\} \cap \{T = k\} \in \mathcal{F}_k \subset \mathcal{F}_n.$$

□

Remark 3.1.18. This clearly extends to where X is an E-valued random variable. We might hope that this would also extend to the situation $\mathbb{T} = [0, \infty[$; however, this is not the case unless we assume some measurability on X with respect to time. (See Theorem 3.2.29 for a positive result in this direction, and Remark 3.2.23 for a counterexample.)

3.2 Stochastic Processes

In Lemma 3.1.17 we have considered a sequence of random variables parameterized by time. Such an object is called a *stochastic process* and gives a mathematical model for a process whose evolution in time is random. We now aim to describe and give properties of such processes.

Definition 3.2.1. *Suppose the time index set $\mathbb{T}$ is as in Definition 3.1.1, and $(\Omega, \mathcal{F})$ is a measurable space. If $(E, \mathcal{E})$ is another measurable space, a* stochastic process *defined on $(\Omega, \mathcal{F})$ with values in $(E, \mathcal{E})$ is a family $\{X_t\}_{t\in\mathbb{T}}$ of E-valued random variables indexed by $t \in \mathbb{T}$. The pair $(\Omega, \mathcal{F})$ is called the* base space *and $(E, \mathcal{E})$ is the* state space *of X.*

For $t \in \mathbb{T}$, X_t is the state *of X at time t. For a fixed $\omega \in \Omega$, the set $\{X_t(\omega); t \in \mathbb{T}\}$ is called the* sample path *or* trajectory *associated with ω.*

Note that we will write X_t for the value of X at time t (as usual, omitting the ω for simplicity), and either $\{X_t\}_{t\in\mathbb{T}}$ or X for the process considered as a whole. We can equivalently think of X as a map $X : \mathbb{T} \times \Omega \to E$, where $X(t, \omega) = X_t(\omega)$.

Remark 3.2.2. Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and X is a stochastic process. If "Π" is some property of a sample path (for example, "Π" might be right-continuity, or having left-hand limits for every $t \in \mathbb{T}$, or being of bounded variation), then we say the process has property "Π" if every sample path $t \mapsto X_t(\omega)$ has property "Π". A weaker statement is to say that the process almost surely has property "Π" by which we mean that, for P-almost all ω, the sample path $t \mapsto X_t(\omega)$ has property "Π". That is, the event $\{\omega : t \mapsto X_t(\omega)\,\text{has property}\,\Pi\}$ has probability one.

Note the subtlety of this concept: it is possible for a continuous-time process to be almost surely continuous at each given time t with probability one, but for its paths to be discontinuous with probability one. See Example 3.2.6 below.

Definition 3.2.3. *When the state space $(E, \mathcal{E})$ is a topological space we shall be particularly concerned with processes which are both continuous on the right and have limits on the left; such processes will be said to be* càdlàg *(or RCLL or 'corlol'). Similarly, the term* càglàd *(or 'collor') would be used to describe a process which is continuous on the left and has limits on the right, and* càg *for one which only is assumed continuous on the left.*

These terms are acronyms derived from the French, for example, càdlàg from 'continu à droite, limite à gauche'.

Remark 3.2.4. Typically, our attention shall be focussed on the case $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, that is, when X is a real valued process. In this case, it is worth observing that if X is càdlàg, then both the process *and the left limits* take values in $\mathbb{R}$. For example, the process $X_t = I_{\{t<1\}}(t-1)^{-1}$ is not càdlàg.

3.2.1 Equivalence of Processes

It is natural to ask when two stochastic processes are, in fact, modelling the same phenomenon. The first definition is the following.

Definition 3.2.5. *Suppose X and Y, are two processes defined on the same probability space $(\Omega, \mathcal{F}, P)$ with values in $(E, \mathcal{E})$. Then Y is said to be a* modification *of X if, for every $t \in \mathbb{T}$, $X_t = Y_t$ a.s.*

Example 3.2.6. Suppose that $\Omega = [0,1]$, $\mathcal{F}$ is the Borel σ-algebra on Ω and P is Lebesgue measure. With $\mathbb{T} = [0,\infty[$, define the process X by $X_t(\omega) := 0$ for all ω and all t. Define Y by $Y_t(\omega) := 0$ if $t - \lfloor t \rfloor \neq \omega$, and $Y_t(\omega) = 1$ if $t - \lfloor t \rfloor = \omega$. (Here $\lfloor t \rfloor$ denotes the greatest integer less than or equal to t.)

The process Y is a modification of X as $P(Y_t = X_t) = P(\omega \neq t - \lfloor t \rfloor) = 1$, but while all sample paths of X are continuous, all sample paths of Y are discontinuous.

Remark 3.2.7. Intuitively, X and Y are modifications of one another if, for any given time t, the probability that $X_t = Y_t$ is one. However, as the above example demonstrates, this does not guarantee that the probability that $X_t = Y_t$ for *every t simultaneously* is one. This motivates the following, stronger, definition.

Definition 3.2.8. *Again, X and Y are two processes defined on the probability space $(\Omega, \mathcal{F}, P)$ with values in $(E, \mathcal{E})$. We say X and Y are* indistinguishable *if, for almost every $\omega \in \Omega$,*

$$X_t(\omega) = Y_t(\omega) \quad \textit{for all} \quad t \in \mathbb{T}.$$

Remark 3.2.9. The key distinction is that, in Definition 3.2.5 the set of measure zero on which X_t and Y_t may differ depends on $t \in \mathbb{T}$. In Definition 3.2.8, there is just one set of measure zero outside which $X_t(\omega) = Y_t(\omega)$ for all t.

In discrete time, these two definitions are equivalent (Exercise 3.4.6), but, as the above example exhibits, they are different in continuous time.

In practice, it is often easier to show that two processes are modifications of each other than it is to directly show they are indistinguishable. If our processes are sufficiently continuous, the following lemma makes the final connection between these concepts.

Lemma 3.2.10. *Suppose that X is a modification of Y, where $\mathbb{T} = [0,\infty[$ or $[0,\infty]$, and that both processes are almost surely right-(or left-) continuous. Then X and Y are indistinguishable.*

Proof. We prove the lemma for the right-continuous case. Let D_X and D_Y be the sets of measure zero on which X and Y are not right-continuous respectively. For each rational number r, the set $\{X_r \neq Y_r\}$ has measure zero. Consequently, the set $D = \bigcup_{r \in \mathbb{Q}} \{\omega : X_r(\omega) \neq Y_r(\omega)\}$ has measure zero (note this is a *countable* union of measurable sets).

For any $t \in \mathbb{T}$, let $\{t_n\}_{n\in\mathbb{N}}$ be a decreasing sequence with $t_n \in \mathbb{Q}$, such that $\lim_n t_n = t$. Since $X_{t_n}(\omega) = Y_{t_n}(\omega)$ for $\omega \notin D \cup D_X \cup D_Y$, by right-continuity

$$X_t(\omega) = \lim_n X_{t_n}(\omega) = \lim_n Y_{t_n}(\omega) = Y_t(\omega).$$

Hence $X_t(\omega) = Y_t(\omega)$ for *all* t, except possibly on the null set $D \cup D_X \cup D_Y$, which does not depend on t. Hence X and Y are indistinguishable. □

Definition 3.2.11. *Suppose A is a subset of $\mathbb{T} \times \Omega$ and that $I_A(t, \omega) = I_A$ is the indicator function of A. Then A is said to be* evanescent *if I_A is indistinguishable from the zero process.*

Remark 3.2.12. This definition allows us to say that, 'if X and Y are indistinguishable, then $X = Y$, except possibly on some evanescent set'.

In fact $A \subset \mathbb{T} \times \Omega$ is evanescent if and only if the projection of A on Ω is a set of measure zero, i.e. $P\big(\{\omega : (\omega, t) \in A \text{ for some } t\}\big) = 0$. Intuitively, this means that A is evanescent if and only if the collection of paths which pass through A at some point occurs with probability zero. Note that, if our probability space is complete, an evanescent set is always measurable.

Remark 3.2.13. We now have four different ways in which we can say two processes X and Y satisfy $X = Y$.

- First, we could have $X_t(\omega) = Y_t(\omega)$ for *all* t and ω.
- A slightly weaker statement is to say that X and Y are indistinguishable, that is, $X_t(\omega) = Y_t(\omega)$ for all t, except possibly for some collection of ω with probability zero.
- Weaker still is to say that X and Y are a modification of each other, that is, for each t, $X_t = Y_t$ P-a.s.
- Finally, we could say that $X_t = Y_t$ $dP \times dt$-a.s., where $dP \times dt$ denotes the product measure on $\Omega \times \mathbb{T}$. In this final case, we can only say that $X_t = Y_t$ P-a.s. for *almost* all t. (Note, this assumes that the X and Y are measurable in the product space $\Omega \times \mathbb{T}$, which will be discussed below, Definition 3.2.22.)

3.2.2 Measurability in Time and Space

We have described the information available at a time t by reference to the σ-algebra $\mathcal{F}_t$. We have also described stochastic processes, that is, collections of random variables indexed by t. We now wish to combine these concepts in such a way that, for a stochastic process X, we can say that the value X_t is known at time t.

Definition 3.2.14. *Suppose $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$ is a filtration of the measurable space $(\Omega, \mathcal{F})$, and that X is a process defined on $(\Omega, \mathcal{F})$ with values in $(E, \mathcal{E})$. Then X is said to be* adapted *to $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$ if X_t is $\mathcal{F}_t$-measurable for each $t \in \mathbb{T}$.*

Remark 3.2.15. If $(\Omega, \mathcal{F}, P)$ is a probability space and X is adapted to the filtration $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$, then a modification Y of X is also adapted, if each $\mathcal{F}_t$ contains all null sets of $\mathcal{F}$.

Remark 3.2.16. Two simple relationships between adapted processes and stopping times are as follows:

(i) For X an adapted continuous process, consider $T = \inf\{t : X_t \le c\}$, for some $c \in \mathbb{R}$. Using the continuity of X, for any $t \in \mathbb{T}$ we can write

$$\{T \le t\} = \Big(\bigcap_{n \in \mathbb{N}} \bigcup_{s \in \mathbb{Q} \cap [0,t]} \{X_s \le c + n^{-1}\} \Big) \in \mathcal{F}_t,$$

so T is a stopping time.

(ii) For X an adapted càdlàg process, if the filtration is right-continuous, then consider $T = \inf\{t : X_t < c\}$. In this case, using the right continuity of X we see that $X_s < c$ for some $s \in [0,t]$ if and only if $X_s < c$ for some $s \in \mathbb{Q} \cap [0,t]$. Therefore, we have

$$\begin{aligned}\{T \le t\} &= \Big(\bigcup_{s \in \mathbb{Q} \cap [0,t]} \{X_s < c\} \Big) \cup \big\{ X_{t_n} < c \text{ for some } t_n \downarrow t \big\} \\ &= \Big(\bigcup_{s \in \mathbb{Q} \cap [0,t]} \{X_s < c\} \Big) \cup \Big(\bigcap_{n \in \mathbb{N}} \bigcup_{s \in]t, t+n^{-1}] \cap \mathbb{Q}} \{X_s \le c + n^{-1}\} \Big) \\ &\in \mathcal{F}_{t+} = \mathcal{F}_t\end{aligned}$$

so T is again a stopping time.

Definition 3.2.17. *Suppose $(\Omega, \mathcal{F}^0, P)$ is a probability space equipped with a filtration $\{\mathcal{F}_t^0\}_{t \in \mathbb{T}}$. By $\mathcal{F}$, we denote the completion of $\mathcal{F}^0$, and, for each $t \in \mathbb{T}$, $\mathcal{F}_t$ denotes the σ-algebra generated by $\mathcal{F}_t^0$ and the P-null sets of $\mathcal{F}$. Then $\{\mathcal{F}_t\}_{t \in \mathbb{T}}$ is a filtration on $(\Omega, \mathcal{F}, P)$ and is called the* completion *of the filtration $\{\mathcal{F}_t^0\}_{t \in \mathbb{T}}$. A filtration is said to be* complete *if $\mathcal{F}$ is complete and each $\mathcal{F}_t$ contains all P-null sets of $\mathcal{F}$.*

Definition 3.2.18. *Suppose $\{X_t\}_{t \in \mathbb{T}}$ is a stochastic process defined on the probability space $(\Omega, \mathcal{F}, P)$. Then $\{X_t\}_{t \in \mathbb{T}}$ is certainly adapted to the filtration $\{\mathcal{F}_t^X\}$ where $\mathcal{F}_t^X = \sigma\{X_s : s \le t\}$ (the σ-algebra on Ω generated by all the random variables X_s, $s \le t$). We call this the filtration* generated *by the process X, or the* natural *filtration of X.*

Remark 3.2.19. This filtration may not be, in general, complete or right-continuous, but it can be completed by adding all P-null sets of $\mathcal{F}$ as in the above definition, and can be made right-continuous by replacing $\mathcal{F}_t^X$ with $\mathcal{F}_{t+}^X$. (See Exercise 3.4.15).

Example 3.2.20. A few examples of adapted processes:

(i) If T is a stopping time, then $I_{\{T \le t\}}$, the stochastic process which is 0 before T and 1 after, is adapted. If T is not a stopping time, then $I_{\{T \le t\}}$ is not adapted.

(ii) When tossing a coin, let $\mathcal{F}_t$ be the σ-algebra generated by the outcome of the first t throws. The process which counts the number of heads already observed is adapted to $\{\mathcal{F}_t\}_{t\in\mathbb{Z}^+}$. The process which indicates the number of heads in the next three throws is not adapted.

(iii) Suppose $\{X_t\}_{t\in\mathbb{T}}$ is an adapted process. Then $\tilde{X}_t := X_{t-\epsilon}$ defines an adapted process for all $\epsilon > 0$, where $X_{-\epsilon} := X_0$ for all $\epsilon > 0$.

Intuitively, a process $\{X_t\}_{t\in\mathbb{T}}$ is adapted if, for all $t \in \mathbb{T}$, the value of X_t is 'known' at time t, that is, X_t is a $\mathcal{F}_t$-measurable random variable. While this notion is fundamental, it only determines measurability of $X_t(\omega)$ as a function of ω, not as a function of t. Particularly in continuous time, this is not quite sufficient, which motivates the following definitions.

Remark 3.2.21. As before, for $t \in \mathbb{T} = [0,\infty[$, we shall write $\mathcal{B} = \mathcal{B}([0,\infty[)$ and $\mathcal{B}([0,t])$ for the Borel σ-algebras on $[0,\infty[$ and $[0,t]$.

Definition 3.2.22. *Suppose $\mathbb{T} = [0,\infty[$ or $\mathbb{T} = [0,\infty]$ and $\{X_t\}_{t\in\mathbb{T}}$ is a stochastic process defined on the measurable space $(\Omega,\mathcal{F})$, with values in $(E,\mathcal{E})$. Then X is said to be a* measurable *process if the map $(t,\omega) \mapsto X_t(\omega)$ is measurable when $\mathbb{T}\times\Omega$ is given the product σ-algebra $\mathcal{B}(\mathbb{T})\otimes\mathcal{F}$.*

If $\{\mathcal{F}_t\}_{t\in\mathbb{T}=[0,\infty[}$ is a filtration of $(\Omega,\mathcal{F})$, then to say X is adapted says something about the measurability (in ω) of $X_t(\omega)$ for each t. To say X is a measurable process is a very weak statement about joint measurability in t and ω. However, measurability of a process does not relate to any filtration, so a measurable process need not be adapted.

Remark 3.2.23. Note that we do not complete the σ-algebra $\mathcal{B}(\mathbb{T})\otimes\mathcal{F}$ in Definition 3.2.22. This is vital, as otherwise various simple events cease to be measurable. For example, let $\Omega = [0,1]$ and $\mathcal{F} = \overline{\mathcal{B}}(\Omega)$ the Lebesgue measurable sets. Let V be a non-Lebesgue-measurable subset of $[0,1]$. Then the process $X_t(\omega) = I_{\{\omega=t\in V\}}$ is zero except on the diagonal $\{t=\omega\}$, which is of Lebesgue measure zero in $\Omega\times[0,\infty[$. Therefore, X is measurable in the Lebesgue-completed product space, but the event $\{\omega : X_\omega = 1\} = \{\omega \in V\}$ is not measurable, so X_ω is not a random variable. As we would like to be able to consider the value of X at a randomly chosen time, this is problematic.

Remark 3.2.24. Di Nunno and Rozanov [58] give necessary and sufficient conditions under which a general process admits a measurable modification.

We now give a definition which relates measurability in t and ω with the filtration. This will also allow us to generalize Lemma 3.1.17 to continuous time, under some conditions.

Definition 3.2.25. *Suppose $\mathbb{T} = [0,\infty[$ or $[0,\infty]$, $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$ is a filtration on $(\Omega,\mathcal{F})$ and that X is a stochastic process defined on $(\Omega,\mathcal{F})$. Then X is said to be* progressively measurable *or* progressive *if, for every $t\in\mathbb{T}$, the map $(s,\omega)\mapsto X_s(\omega)$ of $[0,t]\times\Omega$ into $(E,\mathcal{E})$ is measurable, when $[0,t]\times\Omega$ is given the product σ-algebra $\mathcal{B}([0,t])\otimes\mathcal{F}_t$.*

A progressive process is adapted. In discrete time, an adapted process is progressive. However, in continuous time, an adapted process need not be either measurable or progressive, as the following simple example shows.

Example 3.2.26. Let $f : [0, \infty[\to \mathbb{R}$ be any non-Borel measurable function, (for example, the indicator function of a non-measurable set). For $(\Omega, \mathcal{F}, P)$ a probability space with filtration $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$, and time index $\mathbb{T} = [0, \infty[$, let $\{X_t\}_{t\in\mathbb{T}}$ be the 'stochastic' process defined by $X_t(\omega) := f(t)$. Clearly, as X_t is independent of ω, it is $\mathcal{F}_t$-measurable for all t, and hence is adapted. Conversely, it is also clear that $X : \mathbb{T} \times \Omega \to \mathbb{R}$ is not a measurable function for the product σ-algebra $\mathcal{B}([0, \infty[) \otimes \mathcal{F}$. Hence X is not measurable (and, by consequence, it is not progressive).

The following theorem gives a positive result in this direction.

Theorem 3.2.27. *Suppose $\mathbb{T} = [0, \infty[$ or $[0, \infty]$, $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$ is a filtration on $(\Omega, \mathcal{F})$ and $\{X_t\}_{t\in\mathbb{T}}$ is an adapted right-continuous process with values in a metric space E (which has the Borel σ-algebra $\mathcal{E}$). Then $\{X_t\}_{t\in\mathbb{T}}$ is progressively measurable. The same result is true if $\{X_t\}_{t\in\mathbb{T}}$ is adapted and left-continuous.*

Proof. First fix $t \in [0, \infty[$ and consider a partition of $[0, t[$ into 2^n equal intervals. For $s \in [(k-1)2^{-n}t, k2^{-n}t[$, $1 \le k \le 2^n$, write

$$X_s^n(\omega) = X_{(k-1)2^{-n}t}(\omega),$$

and $X_t^n(\omega) = X_t(\omega)$.

Consider X^n as a map of $[0, t] \times \Omega$ into E. As $X_{(k-1)2^{-n}t}$ is measurable for each t, k, n, the preimages of sets $A \subseteq E$ under X^n are of the form

$$(X^n)^{-1}(A) = \bigcup_{k=1}^{2^n} \Big\{ \{s \in [(k-1)2^{-n}t, k2^{-n}t[\} \times (X_{(k-1)2^{-n}t})^{-1}(A) \Big\}$$

which are all measurable when $[0, t] \times \Omega$ is given the σ-algebra $\mathcal{B}([0, t]) \otimes \mathcal{F}_t$. Hence $\{X^n\}_{t\in\mathbb{N}}$ is a sequence of progressive processes.

Letting $n \to \infty$ we see that the map $(t, \omega) \mapsto X_t(\omega)$ is the pointwise limit $X^n \to X$. By Lemma 1.3.28, we see that X is $\mathcal{B}([0, t]) \otimes \mathcal{F}_t$-measurable, that is, $\{X_t\}_{t\in\mathbb{T}}$ is progressive. $\square$

Definition 3.2.28. *Suppose X is a progressive process on the space $(\Omega, \mathcal{F})$ equipped with the filtration $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$. If S is a stopping time with respect to $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$ then $X^S = \{X_{S\wedge t}\}_{t\in\mathbb{T}}$ is called the process "stopped" at time S.*

Theorem 3.2.29. *Suppose $\{X_t\}_{t\in\mathbb{T}}$, is a progressive process on the space $(\Omega, \mathcal{F})$ equipped with the filtration $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$. If S is a stopping time with respect to $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$ then the random variable $X_S = X_{S(\omega)}(\omega)$ is $\mathcal{F}_S$-measurable, and the process stopped at S, defined by $X_t^S := X_{t\wedge S}$, is progressive.*

Proof. To establish the first result we see that, if $(E, \mathcal{E})$ is the state space of $\{X_t\}_{t\in\mathbb{T}}$, then, for every $B \in \mathcal{E}$, the set $\{X_S \in B\} \cap \{S \leq t\}$ is in $\mathcal{F}_t$. However, $\{X_S \in B\} \cap \{S \leq t\} = \{X_{t\wedge S} \in B\} \cap \{S \leq t\}$, so it is enough to prove the second part of the theorem.

Now $t \wedge S$ is a stopping time less than or equal to t, so $t \wedge S$ is $\mathcal{F}_t$-measurable. Therefore, the map $(s, \omega) \to \big(s \wedge S(\omega), \omega\big)$ is measurable as a map from $([0, t] \times \Omega, \mathcal{B}([0, t]) \otimes \mathcal{F}_t)$ to itself. It follows that the map $(s, \omega) \to X_{s\wedge S(\omega)}(\omega)$ is measurable, by considering preimages of sets through the composition $(s, \omega) \to (s \wedge S(\omega), \omega) \to X_{s\wedge S(\omega)}(\omega)$. Consequently, $\{X_t^S\}_{t\in\mathbb{T}}$ is progressive. □

Remark 3.2.30. The result of Theorem 3.2.29 does not hold for general processes. As we will use the technique of stopping a process extensively, we will usually require any process we consider to be (at least) progressive. In fact, we will even define slightly more restrictive notions of measurability (see Chapter 7) which are important in the study of stochastic integration.

3.3 Localization of Processes

In many cases, a stochastic process may not have a desired property over the entire interval $[0, \infty[$. In the deterministic setting, it is often sufficient to assume that the desired property holds on the interval $[0, T]$ for all finite times T.

In a stochastic setting, it is useful to extend this notion by considering properties holding on a sequence of intervals of the form $[0, T_n]$ where the $\{T_n\}_{n\in\mathbb{N}}$ are stopping times with $T_n \to \infty$.

Definition 3.3.1. *If $\mathcal{C}$ is some family of processes (for example, $\mathcal{A}$ or $\mathcal{A}^+$), then $\mathcal{C}_{\mathrm{loc}}$, the localized class of $\mathcal{C}$, will denote the family of processes which are 'locally' in $\mathcal{C}$. That is, $Y \in \mathcal{C}_{\mathrm{loc}}$ if there is an increasing sequence of stopping times $\{T_n\}_{n\in\mathbb{N}}$ such that*

(i) $\lim_n T_n = \infty$ *a.s. and*
(ii) each stopped process $\{Y_t^{T_n}\}_{t\in\mathbb{T}} = \{Y_{t\wedge T_n}\}_{t\in\mathbb{T}}$ is in $\mathcal{C}$.

The sequence $\{T_n\}_{n\in\mathbb{N}}$ is called a localizing sequence *for Y in $\mathcal{C}$.*

Example 3.3.2. Consider the probability space $(\Omega, \mathcal{F}, P) = \big(]0, 1[, \mathcal{B}(]0, 1[), \mu\big)$, where μ is Lebesgue measure. We give this the trivial filtration $\{\mathcal{F}_t\}_{t\in[0,\infty]}$, where $\mathcal{F} = \mathcal{F}_t = \mathcal{F}_0$, so any random time is a stopping time, and any stochastic process is adapted.

Consider the stochastic process $\{X_t\}_{t\in[0,\infty[}$ given by $X_t(\omega) = t/\omega$. Then the random variable $X_t(\cdot)$ is not bounded for any t, and the path $X_{(\cdot)}(\omega)$ is not bounded for any ω. However, taking the stopping times $T_n = nI_{\{\omega\geq n^{-1}\}}$ we see $X_t^{T_n} \leq n^2$ a.s. for all t, so X is *locally* bounded.

A useful characterization is given by the following lemma, the proof of which is left as an exercise (Exercise 3.4.16).

Lemma 3.3.3. *Let $\mathcal{C}$ be a set of processes such that, if X is a process with $X^T, X^S \in \mathcal{C}$, for S, T stopping times, then $X^{S\vee T} \in \mathcal{C}$. A process Y satisfies $Y \in \mathcal{C}_{\mathrm{loc}}$ if and only if, for any $t > 0$ and any $\epsilon > 0$, there exists a stopping time T such that $Y^T \in \mathcal{C}$ and $P(T > t) > 1 - \epsilon$.*

We shall return to this concept frequently in the coming chapters. Typically, this will be because we prefer to work with processes which have some nice property everywhere (for example, boundedness), but we can only prove that they have this property locally. By using a localization technique, instead of working with a locally bounded process Y, we can instead work with a bounded stopped processes Y^{T_n}, and then infer results for Y.

Remark 3.3.4. It is worth noting that we do not require that the stopping times $\{T_n\}_{n\in\mathbb{N}}$ in the localization converge *uniformly* to ∞. That is, we cannot guarantee that for any $\varepsilon > 0$ there exists N such that $T_n > \varepsilon$ a.s. for all $n \geq N$.

3.4 Exercises

Exercise 3.4.1. Suppose the filtration $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$ is right-continuous and $\{T_n\}_{n\in\mathbb{N}}$ is a decreasing sequence of stopping times whose limit is the stopping time T. Show that $\mathcal{F}_T = \bigcap_n \mathcal{F}_{T_n}$.

Exercise 3.4.2. If T is a stopping time for $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$, prove that $\mathcal{F}_T$ is a σ-algebra.

Exercise 3.4.3. Show that the evanescent sets form a σ-algebra.

Exercise 3.4.4. Show that T is $\mathcal{F}_T$-measurable and, if $T = t$ for some deterministic t, then $\mathcal{F}_T = \mathcal{F}_t$.

Exercise 3.4.5. For T an a.s. finite stopping time, show that $\mathcal{F}_T \subseteq \mathcal{F}_\infty$, where if $\infty \notin \mathbb{T}$ we define $\mathcal{F}_\infty = \bigvee_{t\in\mathbb{T}} \mathcal{F}_t$.

Exercise 3.4.6. Consider a filtered probability space in discrete time, $\mathbb{T} = \mathbb{Z}^+$ or $\mathbb{T} = \overline{\mathbb{Z}}^+$. Let X and Y be two stochastic processes. Show that if X is a modification of Y, then X and Y are indistinguishable.

Exercise 3.4.7. Let $\{X_t\}_{t\in\mathbb{T}} = \{X(t,\omega)\}_{t\in\mathbb{T}}$ be a measurable process, in the sense of Definition 3.2.22. Show that the paths $X(\cdot;\omega)$ are $\mathcal{B}(\mathbb{T})$-measurable for P-almost all ω.

Exercise 3.4.8. Let X be an a.s. càdlàg process. Show that, for any $\epsilon > 0$ and any $T > 0$, there are almost surely finitely many $t \leq T$ such that

$$|\Delta X_t| := |X_t - X_{t-}| \geq \epsilon.$$

(Hint: Compare with Lemma 1.3.42.)

Exercise 3.4.9. Let X be an adapted càdlàg process, for a general filtration $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$. For $c \in \mathbb{R}$, in each case show that T is a stopping time, or give a counterexample.

(i) $T = \sup\{t : X_t \geq c\}$.
(ii) $T = \sup\{t : X_s \leq c \text{ for all } s \leq t\}$
(iii) $T = \inf\{t : \int_{[0,t[} X_t dt = c\}$
(iv) $T = \sup\{t : X_t < c/t\}$
(v) $T = \sup T_n$ where $T_n = \inf\{t : X_t = n\}$.

Which of these require X to be càdlàg? What if $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$ is right continuous?

Exercise 3.4.10. Let X be an adapted process on a filtered probability space $(\Omega, \mathcal{F}, P)$, with a filtration $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$, where $\mathcal{F}_t = \sigma(\{X_u\}_{u\leq t})$, $\mathcal{F} = \mathcal{F}_\infty$ and time index $\mathbb{T} = [0, \infty]$. Let $\mathcal{G}_s = \sigma(\{X_u\}_{u\geq 1/s})$. Show that

(i) $\{\mathcal{G}_s\}_{s\in\mathbb{T}}$ is also a filtration on $(\Omega, \mathcal{F}, P)$.
(ii) $\{X_{1/s}\}_{s\geq 0}$ is an adapted process with respect to the filtration $\{\mathcal{G}_s\}_{s\in\mathbb{T}}$.
(iii) $\sigma(\mathcal{G}_t, \mathcal{F}_t) = \mathcal{F}$ for all $t \geq 1$.
(iv) If $\mathcal{F}_0$ is trivial, that is, $\mathcal{F}_0 = \{\emptyset, \Omega\}$, then the event $X_\infty = X_0$ is in $\mathcal{G}_s$ for all s.
(v) Conversely, if X_∞ is not $\mathcal{F}_t$-measurable for any $t < \infty$, then show that the event $X_\infty = X_0$ is not in $\mathcal{F}_t$ for any t.

Exercise 3.4.11. Let T be an exponentially distributed random variable, so $P(T > t) = e^{-\lambda t}$ for some $\lambda > 0$ and each $t \in [0, \infty[$. Let $\mathcal{F}_t = \sigma(I_{\{T\leq s\}} : s \leq t)$.

(i) Show that T is an $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$ stopping time.
(ii) Find an expression for $E[T|\mathcal{F}_t]$.
(iii) Are either T^2 or $T^{1/2}$ also $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$ stopping times?

Exercise 3.4.12. Give a general condition on a map $\phi : \mathbb{R} \to \mathbb{R}$ such that $\phi(T)$ is a stopping time for any stopping time T.

Exercise 3.4.13. Let $\{X_t\}_{t\in\mathbb{T}}$ be an a.s. càdlàg process, and let $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$ be the completed filtration generated by X. Show that $\mathcal{F}_t$ has the countable representation

$$\mathcal{F}_t = \sigma(X_t) \vee \Big(\bigvee_{\{s\in\mathbb{Q}:s<t\}} \sigma(X_s)\Big) \vee \{\text{null sets}\}.$$

Exercise 3.4.14. Let $\{X_t\}_{t\in[0,\infty[}$ be a càdlàg process. Show that $Y_t = \int_{[0,t]} X_s ds$ defines an adapted process such that, if $\{\mathcal{F}_t\}_{t\in[0,\infty[}$ is the filtration generated by X, and $\{\mathcal{G}_t\}_{t\geq 0}$ is the filtration generated by Y, then $\mathcal{G}_{t+} = \mathcal{F}_{t+}$ for all t. Give a counterexample if X is measurable but not càdlàg.

Exercise 3.4.15. Let $\Omega = C([0,T])$ be the space of continuous functions $\omega : [0,T] \to \mathbb{R}$. Let $\mathcal{F}_t = \sigma(X_s : s \leq t)$ be the filtration generated by $X_t = \omega_t$. Show that $\mathcal{F}_t = \bigvee_{\{s \in \mathbb{Q}: s<t\}} \mathcal{F}_s$, and so the filtration is left-continuous at deterministic times (that is, $\mathcal{F}_t = \mathcal{F}_{t-}$ for all $t > 0$). By considering the event $\{X \text{ is differentiable at time } t\}$, show that this filtration is not right-continuous, even though X is a continuous process.

Exercise 3.4.16. Let $\mathcal{C}$ be a set of processes such that if X is a process with $X^T, X^S \in \mathcal{C}$, for S, T stopping times, then $X^{S \vee T} \in \mathcal{C}$. Show that a process Y satisfies $Y \in \mathcal{C}_{\text{loc}}$, in the sense of Definition 3.3.1, if and only if, for any $t > 0$ and any $\epsilon > 0$, there exists a stopping time T such that $Y^T \in \mathcal{C}$ and $P(T > t) > 1 - \epsilon$.

4

Martingales in Discrete Time

In this chapter and the next, we consider one of the most important classes of stochastic processes, the class of *martingales*. Their significance was first emphasized in the now classical book of Doob [62].

Results for discrete time martingales are established in this chapter and extended to continuous time martingales in the next.

4.1 Definitions and Basic Properties

Throughout this chapter, unless otherwise stated, we take as given a probability space $(\Omega, \mathcal{F}, P)$ equipped with a filtration $\{\mathcal{F}_n\}_{n\in\mathbb{T}}$, where $\mathbb{T} = \mathbb{Z}^+ = \{0, 1, 2, \ldots\}$ or $\mathbb{T} = \overline{\mathbb{Z}}^+ = \{0, 1, 2, \ldots, \infty\}$.

Definition 4.1.1. *A real-valued stochastic process $\{X_n\}_{n\in\mathbb{T}}$ is called a* super-martingale *with respect to the filtration $\{\mathcal{F}_n\}_{n\in\mathbb{T}}$ if*

(i) each X_n is $\mathcal{F}_n$-measurable, i.e. $\{X_n\}_{n\in\mathbb{T}}$ is adapted to $\{\mathcal{F}_n\}_{n\in\mathbb{T}}$,
(ii) $E\big[|X_n|\big] < \infty$, for all $n \in \mathbb{T}$, and
(iii) $X_n \geq E[X_m|\mathcal{F}_n]$ almost surely, for all $m \geq n$.

If "$\geq$" in property (iii) is replaced by "$\leq$", then X is called a submartingale. *If the sequence X is both a supermartingale and a submartingale, then it is called a* martingale.

Remark 4.1.2. The term "martingale" has an interesting history. Originally a term for hose (i.e. trousers) which fasten at the back, a martingale became known as the part of a horse's harness which prevents the horse from rearing its head. Through horse racing the word became a gambling term (see Exercise 4.7.7), and the mathematical definition above can be thought of as a

S.N. Cohen, R.J. Elliott, *Stochastic Calculus and Applications*,
Probability and Its Applications, DOI 10.1007/978-1-4939-2867-5_4

model of a gambler's winnings in a fair game of chance (and the term in this sense is due to Ville [179]). Likewise supermartingales[1] and submartingales correspond to games which are respectively unfavourable and favourable for a gambler. A detailed etymological study is given by Mansuy [128].

Remark 4.1.3. Note that X is a submartingale if and only if $-X$ is a supermartingale. Also, for any martingale X, $E[X_{n+1}|\mathcal{F}_n] = X_n$ for all n, and this statement is equivalent to (iii) in the definition whenever $\mathbb{T} = \mathbb{Z}^+$. This is commonly known as the *martingale property* (and similarly we have the supermartingale and submartingale properties).

Martingales form one of the most interesting and useful classes of stochastic processes in all of probability theory. As we shall see, many results which are usually proven for sequences of independent random variables (such as the law of large numbers) have analogues for martingales.

Example 4.1.4. Some examples of martingales.

(i) Let $\{Y_n\}_{n\in\mathbb{Z}^+}$ be a sequence of independent integrable random variables defined on $(\Omega, \mathcal{F}, P)$, with $E[Y_n] = 0$ for all n. Let $\mathcal{F}_n$ be the σ-algebra generated by $\{Y_0, Y_1, \ldots, Y_n\}$. The process X defined by the partial sums $X_n = Y_0 + Y_1 + \ldots + Y_n = \sum_{i=0}^n Y_i$ is a $\{\mathcal{F}_n\}_{n\in\mathbb{Z}^+}$ martingale.
(ii) Suppose $Y \in L^1(\Omega, \mathcal{F}, P)$ and $X_n = E[Y|\mathcal{F}_n]$. Then $\{X_n\}_{n\in\mathbb{T}}$ is a martingale.
(iii) Let X be a martingale and H a bounded adapted process, that is, H_n is $\mathcal{F}_n$-measurable. Then the process Y defined by

$$Y_n = \sum_{i=1}^n H_{i-1}(X_i - X_{i-1})$$

is a martingale.

Lemma 4.1.5. *Suppose $\{X_n\}_{n\in\mathbb{T}}$ is an $\{\mathcal{F}_n\}_{n\in\mathbb{T}}$-martingale (resp. submartingale) and ϕ is a convex (resp. convex, nondecreasing) function defined on $\mathbb{R}$ such that the random variables $\phi \circ X_n$ are integrable for every n. Then $\{\phi \circ X_n\}_{n\in\mathbb{T}}$ is an $\{\mathcal{F}_n\}_{n\in\mathbb{T}}$ submartingale.*

[1]The term 'supermartingale' is not directly related to horse racing (it is historically due to connections with 'superharmonic functions' from classical analysis) but comes about as it says the current value is 'above' the expectation in the future. Nevertheless, relating the term to a horse's harness can help one to remember that a supermartingale is a process which is being 'pulled down' through time more strongly than a martingale, while a submartingale is being 'pulled down' less strongly. Another easy way to remember which is a sub- or supermartingale is that the tail of the 'p/b' points in the direction that the expected value is changing.

Proof. If X is a martingale, for $n \in \mathbb{T}$, write $Y_n = \phi \circ X_n$. Then, by Jensen's inequality (Lemma 2.4.11), for any $m \geq n$,

$$E[Y_m|\mathcal{F}_n] = E[\phi \circ X_m|\mathcal{F}_n] \geq \phi \circ E[X_m|\mathcal{F}_n] = \phi \circ X_n = Y_n.$$

If X is a submartingale, then $X_n \leq E[X_m|\mathcal{F}_n]$ for all $m \geq n$ and if ϕ is convex and nondecreasing, we have

$$Y_n = \phi \circ X_n \leq \phi \circ E[X_m|\mathcal{F}_n] \leq E[\phi \circ X_m|\mathcal{F}_n] = E[Y_m|\mathcal{F}_n].$$

□

Remark 4.1.6. Commonly encountered examples of functions satisfying the above conditions are:

$$\begin{aligned}
&\phi(x) = |x|^p && \text{for } p \geq 1, && \text{(convex)},\\
&\phi(x) = x \vee 0 = x^+, && && \text{(convex nondecreasing)},\\
&\phi(x) = (x-\alpha)^+ && \text{for } \alpha \in \mathbb{R}, && \text{(convex nondecreasing)}.
\end{aligned}$$

4.2 Optional Stopping

For many problems, we wish to replace the fixed times in our (super)martingale property with stopping times. The ability to do this allows us to establish remarkable convergence results for martingales, which underpin much of the theory of stochastic processes.

We shall begin by establishing the result for stopping times which are almost-surely bounded, with a view to later extending this to possibly infinite stopping times (Theorem 4.6.7).

Theorem 4.2.1 (Optional Stopping–Bounded Stopping Times). *Suppose that $\{X_n\}_{n\in\mathbb{T}}$ is an $\{\mathcal{F}_n\}_{n\in\mathbb{T}}$-supermartingale. If S and T are bounded $\{\mathcal{F}_n\}_{n\in\mathbb{T}}$ stopping times and $S \leq T$ a.s. then $E[X_T|\mathcal{F}_S] \leq X_S$ a.s.*

Proof. Let $M < \infty$ be an integer such that $S \vee T \leq M$ a.s. We must show that for every $A \in \mathcal{F}_S$

$$E[I_A X_S] \geq E[I_A X_T].$$

Suppose first that $S \leq T \leq S+1$ and write

$$B_n = A \cap \{S = n\} \cap \{T > S\} = A \cap \{S = n\} \cap \{T > n\}$$

and

$$\tilde{B} = A \cap \{S = T\}.$$

Now $A \cap \{S = n\} \in \mathcal{F}_n$, as S is a stopping time and $A \in \mathcal{F}_S$, and $\{T > n\}$ is the complement of $\{T \leq n\} \in \mathcal{F}_n$. Consequently, each $B_n \in \mathcal{F}_n$. It is also clear that $\tilde{B} \cap B_n = B_n \cap B_m = \emptyset$ for any $n \neq m$ and $A = \tilde{B} \cup (\bigcup_{n=0}^{M} B_n)$. By construction, $X_S - X_T = X_n - X_{n+1}$ on B_n and $X_S - X_T = 0$ on $\tilde{B}$. Therefore,

$$E[I_A(X_S - X_T)] = E\Big[I_{\tilde{B}}(X_S - X_T) + \sum_{n=0}^{M} I_{B_n}(X_S - X_T)\Big]$$
$$= E\Big[\sum_{n=0}^{M} I_{B_n} E[X_n - X_{n+1}|\mathcal{F}_n]\Big] \geq 0,$$

because $E[X_{n+1}|\mathcal{F}_n] \leq X_n$ a.s. The result is, therefore, proven when $T - S \leq 1$.

In the general case, write

$$R_n = T \wedge (S + n), \quad n = 0, 1, 2, \ldots, M,$$

so that, from Example 3.1.7 and Lemma 3.1.9, the R_n are $\{\mathcal{F}_n\}_{n \in \mathbb{T}}$-stopping times and, because $S \leq R_n$ for each n, $\mathcal{F}_S \subset \mathcal{F}_{R_n}$. Consequently, $A \in \mathcal{F}_{R_n}$ for each n and $R_{n+1} - R_n \leq 1$. Now $R_0 = S$ and $R_M = T$, so from the case discussed above,

$$E[I_A X_S] = E[I_A X_{R_0}] \geq E[I_A X_{R_1}] \geq \cdots \geq E[I_A X_{R_M}] = E[I_A X_T].$$

□

Corollary 4.2.2. *Suppose S is a bounded stopping time.*

(i) If X is a supermartingale, so is X^S.
(ii) If X is a submartingale, so is X^S.
(iii) If X is a martingale, so is X^S.
(iv) If X is uniformly integrable, so is X^S.

4.3 Upcrossing and Downcrossing Inequalities

A fundamental property of martingales is their convergence in time. We now establish Doob's upcrossing and downcrossing inequalities, which will allow us to establish these convergence properties. Intuitively, for a submartingale X, these inequalities bound the expected variation of the function $n \mapsto X_n(\omega)$, by considering the number of times the process can cross an arbitrary interval in an upward or downward direction.

Definition 4.3.1. *For a discrete-time stochastic process $\{X_n\}_{n \in \mathbb{T}}$ and a given interval $[\alpha, \beta]$, we say that X upcrosses $[\alpha, \beta]$ over a period $\{n_0, \ldots, n_k\}$ if $X_{n_0} < \alpha$ and $\beta < X_{n_k}$.*

We denote by $M(\omega, X; [\alpha, \beta])$ *the number of distinct upcrossings of* $[\alpha, \beta]$, *that is, the largest integer* $k \in \overline{\mathbb{Z}}$ *such that we can find random times*

$$0 \leq s_1 < t_1 < s_2 < t_2 < \ldots < s_k < t_k$$

with $X_{s_i} < \alpha \leq \beta < X_{t_i}$ *for all* i *(Fig. 4.1).*

Similarly, we define $D(\omega, X; [\alpha, \beta])$, *the number of distinct downcrossings of* $[\alpha, \beta]$ *as the largest integer* $k \in \overline{\mathbb{Z}}$ *such that we can find* $0 \leq s_1 < t_1 \ldots < s_k < t_k$ *with* $X_{s_i} > \beta \geq \alpha > X_{t_i}$ *for all* i. *The inequalities are naturally*

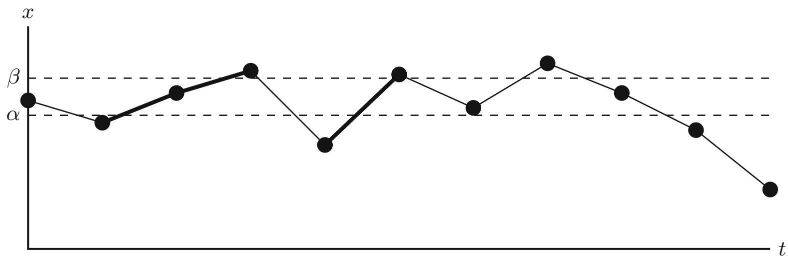

Fig. 4.1. A depiction of the number of upcrossings of $[\alpha, \beta]$, in this case, $M(\omega, X; [\alpha, \beta]) = 2$ (inspired by Williams [183]).

weakened to $\geq$ *or* $\leq$ *in the corresponding definitions of upcrossings and downcrossings of open or half-open intervals.*

We now give a bound on the mean number of upcrossings and downcrossings of a stopped supermartingale.

Theorem 4.3.2. *Suppose* $\{X_n\}_{n \in \mathbb{T}}$ *(where* $\mathbb{T} = \mathbb{Z}^+$ *or* $\overline{\mathbb{Z}}^+$*) is a submartingale and that* S *is a bounded stopping time with respect to* $\{\mathcal{F}_n\}_{n \in \mathbb{N}}$. *Let* α *and* β *be real numbers with* $\alpha < \beta$. *Define*

$$M := M(\omega, X^S; [\alpha, \beta]), \qquad D := D(\omega, X^S; [\alpha, \beta]).$$

Then, almost surely,

$$E[M|\mathcal{F}_0] \leq (\beta - \alpha)^{-1} \big(E\big[(X_S - \alpha)^+ \big| \mathcal{F}_0\big] - (X_0 - \alpha)^+ \big),$$
$$E[D|\mathcal{F}_0] \leq (\beta - \alpha)^{-1} E[(X_S - \beta)^+ | \mathcal{F}_0].$$

Proof. From Lemma 4.1.5 and Corollary 4.2.2, $Y_n := (X_n^S - \alpha)^+$ defines a submartingale which is constant after time S. We also know

$$M = M(\omega, Y^S; [0, \beta - \alpha]), \qquad D := D(\omega, Y^S; [0, \beta - \alpha]).$$

Let M^1 and D^1 be the number of upcrossings and downcrossings, respectively, of the open interval $]0, \beta - \alpha[$ by Y_n. Note that $M \leq M^1$ and $D \leq D^1$.

Taking $\min\{\emptyset\} = \infty$, define a sequence of stopping times $\{T_n\}_{n\in\mathbb{N}}$ as follows:

$$\begin{aligned}
T_0(\omega) &= 0 \quad \text{for all } \omega \in \Omega,\\
T_1(\omega) &= S \wedge \min\{n : n > T_0(\omega) \text{ and } Y_n(\omega) = 0\},\\
T_2(\omega) &= S \wedge \min\{n : n > T_1(\omega) \text{ and } Y_n(\omega) \geq \beta - \alpha\},
\end{aligned}$$

and so on, so that

$$\begin{aligned}
T_{2k+1}(\omega) &= S \wedge \min\{n : n > T_{2k}(\omega) \text{ and } Y_n(\omega) = 0\},\\
T_{2k+2}(\omega) &= S \wedge \min\{n : n > T_{2k+1}(\omega) \text{ and } Y_n(\omega) \geq \beta - \alpha\}.
\end{aligned}$$

Eventually we reach $T_{2p}(\omega) = S(\omega)$ a.s., for some fixed $p < \infty$, because S is a bounded stopping time, and we are considering a discrete time process. Then

$$\begin{aligned}
Y_S(\omega) - Y_0(\omega) &= [Y_{T_1}(\omega) - Y_{T_0}(\omega)] + [Y_{T_2}(\omega) - Y_{T_1}(\omega)]\\
&\quad + \cdots + [Y_{T_{2p}}(\omega) - Y_{T_{2p-1}}(\omega)]\\
&= \sum_{i=0}^{2p} [Y_{T_i} - Y_{T_{i-1}}].
\end{aligned}$$

Consider the terms $Y_{T_{2k}}(\omega) - Y_{T_{2k-1}}(\omega)$ in this sum. As there are M^1 upcrossings of $]0, \beta - \alpha[$, M^1 terms of this sum correspond to a jump of Y_n from 0 to a value at least $\beta - \alpha$. Furthermore, as Y is a submartingale, by Theorem 4.2.1 and Lemma 2.4.8, the expectation of each term in the sum is nonnegative. Therefore

$$E[Y_S|\mathcal{F}_0] - Y_0 = E\Big[\sum_{i=0}^{2p} [Y_{T_i} - Y_{T_{i-1}}]\Big|\mathcal{F}_0\Big] \geq E[(\beta - \alpha)M^1|\mathcal{F}_0].$$

It then follows that

$$E[M|\mathcal{F}_0] \leq E[M^1|\mathcal{F}_0] \leq (\beta - \alpha)^{-1} \left(E[Y_S|\mathcal{F}_0] - Y_0\right).$$

To prove the second inequality, define a sequence of stopping times $\{S_n\}_{n\in\mathbb{N}}$ similar to our sequence $\{T_n\}_{n\in\mathbb{N}}$ by

$$\begin{aligned}
S_0(\omega) &= 0 \qquad \text{for all } \omega \in \Omega,\\
S_{2k+1}(\omega) &= S \wedge \min\{n : n > S_{2k}(\omega) \text{ and } Y_n(\omega) \geq \beta - \alpha\},\\
S_{2k+2}(\omega) &= S \wedge \min\{n : n > S_{2k+1}(\omega) \text{ and } Y_n(\omega) = 0\}.
\end{aligned}$$

Continue in this manner, so that eventually $S_{2p}(\omega) = S(\omega)$. By Theorem 4.2.1,

$$E[(Y_{S_2} - Y_{S_1}) + \cdots + (Y_{S_{2p}} - Y_{S_{2p-1}})|\mathcal{F}_0] \geq 0.$$

However, each nonzero term in this sum, except possibly the final one, corresponds to a descent to 0 from a value greater than or equal to $(\beta - \alpha)$ and the final term has a value at most equal to

$$(Y_S - (\beta - \alpha))^+ = (X_S - \beta)^+.$$

Consequently, as there are precisely D^1 downcrossings of $]0, \beta - \alpha[$ by Y,

$$E[(X_S - \beta)^+ | \mathcal{F}_0] - (\beta - \alpha) E[D^1 | \mathcal{F}_0] \geq 0$$

and we conclude

$$E[D | \mathcal{F}_0] \leq (\beta - \alpha)^{-1} E[(X_S - \beta)^+ | \mathcal{F}_0].$$

□

Corollary 4.3.3. *Suppose X is a supermartingale and S a bounded stopping time. Applying the above inequalities to the submartingale $-X$ over the interval $[-\beta, -\alpha]$ and taking expectations, we have:*

$$\begin{aligned} E[D(\omega, X^S; [\alpha, \beta])] &= E[M(\omega, -X^S; [-\beta, -\alpha])] \\ &\leq (\beta - \alpha)^{-1} E[(-X_S + \beta)^+ - (-X_0 + \beta)^+] \\ &\leq (\beta - \alpha)^{-1} E[X_0 \wedge \beta - X_S \wedge \beta], \end{aligned}$$

and

$$\begin{aligned} E[M(\omega, X^S; [\alpha, \beta])] &= E[D(\omega, -X^S; [-\beta, -\alpha])] \\ &\leq (\beta - \alpha)^{-1} E[(-X_S + \alpha)^+] \\ &= (\beta - \alpha)^{-1} E[(X_S - \alpha)^-] \\ &\leq (\beta - \alpha)^{-1} (E[X_S^-] + |\alpha|) \\ &\leq (\beta - \alpha)^{-1} (E\big[|X_S|\big] + |\alpha|). \end{aligned}$$

4.4 Convergence Results

We now use the inequalities from the previous section to prove convergence results for supermartingales (and hence, for submartingales). Essentially, we aim to take a supermartingale $\{X_t\}_{t \in \mathbb{Z}^+}$ and show that it converges almost surely and in L^1 (as $t \to \infty$) to a random variable X_∞. Furthermore, the new process $\{X_t\}_{t \in \overline{\mathbb{Z}}^+ = \mathbb{Z}^+ \cup \{\infty\}}$ is still a supermartingale.

For the sake of notational simplicity, if our setting is $\mathbb{T} = \mathbb{Z}^+$, then we define $\mathcal{F}_\infty = \bigvee_n \mathcal{F}_n$, and so have a filtration defined for $\mathbb{T} = \overline{\mathbb{Z}}^+$. We write $\mathcal{F}_{\infty-} = \bigvee_{n < \infty} \mathcal{F}_n$, and so know $\mathcal{F}_\infty \supseteq \mathcal{F}_{\infty-}$.

Theorem 4.4.1. *Suppose $\{X_n\}_{n \in \mathbb{Z}^+}$ is a supermartingale such that*

$$\sup_n E[X_n^-] < \infty.$$

Then the sequence $\{X_n\}_{n \in \mathbb{Z}^+}$ converges almost surely to an integrable random variable $X_\infty \in L^1(\mathcal{F}_{\infty-})$.

Proof. For any finite integer k we work with X^k, the supermartingale stopped at k. For any interval $[\alpha,\beta]$, consider $M(\omega,X^k;[\alpha,\beta])$, the number of upcrossings of $[\alpha,\beta]$ by X^k. By Corollary 4.3.3,

$$E[M(\omega,X^k;[\alpha,\beta])] \le (\beta-\alpha)^{-1}(E[X_k^-]+|\alpha|).$$

Now the sequence $\{M(\omega,X^k;[\alpha,\beta])\}_{k=0}^{\infty}$ is monotonic increasing to $M(\omega,X;[\alpha,\beta])$, so by monotone convergence (Theorem 2.4.3),

$$E[M(\omega,X;[\alpha,\beta])] = \lim_k E[M(\omega,X^k;[\alpha,\beta])].$$

Therefore,

$$E[M(\omega,X;[\alpha,\beta])] \le (\beta-\alpha)^{-1}(\sup_k E[X_k^-]+|\alpha|) < \infty,$$

and so $M(\omega,X;[\alpha,\beta]) < \infty$ for almost every $\omega \in \Omega$. Consequently, the event

$$\begin{aligned} H_{\alpha,\beta} &= \{\omega : M(\omega,X;[\alpha,\beta]) = \infty\} \\ &= \Big\{\omega : \limsup_n X_n(\omega) \ge \beta \text{ and } \liminf_n X_n(\omega) \le \alpha\Big\}. \end{aligned}$$

has probability zero.

Now suppose H is the union of all events $H_{\alpha,\beta}$, where α, β are rational numbers with $\alpha<\beta$. Then H has probability zero. Therefore, for almost all ω, that is, for all $\omega \notin H$,

$$\liminf_n X_n(\omega) = \limsup_n X_n(\omega)$$

and so $\lim_{n\to\infty} X_n(\omega)$ exists almost surely (and by defining $X_\infty(\omega) = \liminf_n X_n(\omega)$, we have a measurable limit X_∞ defined for all ω).

We now have a version of the limit X_∞ and merely need to show that X_∞ is integrable. We write

$$\|X_n\|_1 = E[X_n^+ + X_n^-] = E[X_n] + 2E[X_n^-] \le E[X_0] + 2E[X_n^-].$$

So, by Fatou's lemma

$$E\big[|X_\infty|\big] = E\big[\lim_n |X_n|\big] \le \liminf_n E\big[|X_n|\big] \le \sup_n \|X_n\|_1 < \infty.$$

□

Remark 4.4.2. Since

$$E[X_n^-] \le \|X_n\|_1 \le E[X_0] + 2E[X_n^-] \le 3\|X_n\|_1,$$

the hypothesis of the theorem is equivalent to saying $\sup_n \|X_n\|_1 < \infty$.

Corollary 4.4.3. *Suppose X is a nonnegative supermartingale, then certainly $\sup_n E[X_n^-] = 0 < \infty$, so the sequence $\{X_n\}_{n\in\mathbb{Z}^+}$ converges almost surely to an integrable random variable X_∞.*

Corollary 4.4.4. *Suppose X is a uniformly integrable $\{\mathcal{F}_n\}_{n\in\mathbb{T}}$ supermartingale, that is, the family $\{X_n\}_{n\in\mathbb{Z}^+}$ of random variables is uniformly integrable. Then, defining $X_\infty = \lim_n X_n$, the process $\{X_n\}_{n\in\overline{\mathbb{Z}}^+}$ is a supermartingale and $\lim_n \|X_n - X_\infty\|_1 = 0$.*

Proof. Because $\{X_n\}_{n\in\mathbb{Z}^+}$ is uniformly integrable, we know $\sup_n \|X_n\|_1 < \infty$ and the sequence $\{X_n\}_{n\in\mathbb{Z}^+}$ converges almost surely to an integrable random variable X_∞. Since $\{X_n\}_{n\in\mathbb{Z}^+}$ is uniformly integrable, the convergence also takes place in L^1, by Theorem 2.5.8.

We need to show $\{X_n\}_{n\in\overline{\mathbb{Z}}^+}$ is a supermartingale. As $\mathcal{F}_\infty \supseteq \bigvee_n \mathcal{F}_n$, we know X_∞ is $\mathcal{F}_\infty$-measurable. Suppose $m < n$ and $A \in \mathcal{F}_m$. Then

$$E[I_A X_m] \geq E\big[I_A E[X_n|\mathcal{F}_m]\big] = E[I_A X_n],$$

so, by Fatou's lemma, letting $n \to \infty$, we see that $X_m \geq E[X_\infty|\mathcal{F}_m]$ a.s. $\square$

Corollary 4.4.5. *Suppose $\{X_n\}_{n\in\mathbb{Z}^+}$ is a uniformly integrable martingale. Then $\lim_n X_n = X_\infty$ a.s. and in L^1, and $X_n = E[X_\infty|\mathcal{F}_n]$ a.s. for each $n \in \mathbb{Z}^+$.*

Proof. Apply Corollary 4.4.4 to the supermartingales $\{X_n\}_{n\in\mathbb{Z}^+}$ and $\{-X_n\}_{n\in\mathbb{Z}^+}$. $\square$

A converse to this corollary is the following result.

Theorem 4.4.6. *Suppose $\{\mathcal{F}_n\}_{n\in\mathbb{T}}$, is a filtration of $(\Omega, \mathcal{F})$. If $Y \in L^1(\Omega, \mathcal{F}, P)$ and $X_n = E[Y|\mathcal{F}_n]$ a.s., then $\{X_n\}_{n\in\mathbb{Z}^+}$ is a uniformly integrable martingale. Furthermore, if Y is $\mathcal{F}_{\infty-}$-measurable, then $\lim_{n\to\infty} X_n = Y$ a.s.*

Proof. By Jensen's inequality

$$|X_n| = \big|E[Y|\mathcal{F}_n]\big| \leq E\big[|Y|\,\big|\mathcal{F}_n\big] < \infty.$$

For $m \geq 0$,

$$E[X_{n+m}|\mathcal{F}_n] = E\big[E[Y|\mathcal{F}_{n+m}]\big|\mathcal{F}_n\big] = E[Y|\mathcal{F}_n] = X_n \text{ a.s.},$$

and as $E\big[|X_n|\big] \leq E\big[|Y|\big] < \infty$ we see $\{X_n\}_{n\in\mathbb{Z}^+}$ is a martingale.

For $\lambda \geq 0$, let

$$I(n,\lambda) := \int_{\{|X_n|>\lambda\}} |X_n|\, dP \leq \int_{\{|X_n|>\lambda\}} |Y|\, dP.$$

By Markov's inequality (Exercise 2.7.3),

$$P(|X_n| > \lambda) \leq \frac{E\big[|X_n|\big]}{\lambda} \leq \frac{E\big[|Y|\big]}{\lambda},$$

which tends to zero uniformly in n, as $\lambda \to \infty$. As the measure $\nu(A) := \int_A |Y| dP$ is absolutely continuous with respect to P and is finite, by Lemma 1.6.2 we see that $I(n, \lambda)$ tends to zero uniformly in n as $\lambda \to \infty$. Therefore $\{X_n\}_{n \in \mathbb{Z}^+}$ is a uniformly integrable family.

Now suppose Y is $\mathcal{F}_{\infty-}$-measurable. From Corollary 4.4.5 we know that $\lim_n X_n = X_\infty$ exists and that $X_n = E[X_\infty|\mathcal{F}_n] = E[Y|\mathcal{F}_n]$. We wish to show $X_\infty = Y$. Write $\mathcal{G}$ for the family of events $A \in \mathcal{F}_{\infty-}$ such that

$$\int_A Y\, dP = \int_A X_\infty\, dP.$$

Now for each $n \in \mathbb{Z}^+$, $\mathcal{F}_n \subset \mathcal{G}$ and $\mathcal{G}$ is closed under countable unions and intersections. Therefore, by the monotone class theorem (Theorem 1.1.14), $\mathcal{G} = \mathcal{F}_{\infty-}$. As X_∞ and Y are $\mathcal{F}_{\infty-}$ measurable, this implies $X_\infty = Y$ a.s. $\square$

Remark 4.4.7. Corollary 4.4.5 and Theorem 4.4.6 establish necessary and sufficient conditions for a martingale $\{X_n\}_{n \in \mathbb{Z}^+}$ to converge to a limit X_∞ in L^1. Specifically, there exists an $\mathcal{F}_\infty$-integrable random variable X_∞ such that $X_n = E[X_\infty|\mathcal{F}_n]$ a.s. if and only if $\{X_n\}_{n \in \mathbb{Z}^+}$ is uniformly integrable.

4.5 Maximal Inequalities

We now seek to derive bounds on the maximum value attained by a supermartingale. Together with the up and downcrossing inequalities, this determines much of the behaviour of these processes.

Lemma 4.5.1. *Suppose $\{X_n\}_{n \in \mathbb{Z}^+}$ is a supermartingale. For every $\alpha \geq 0$,*

$$\alpha P\big(\sup_n X_n \geq \alpha\big) \leq E[X_0] + \sup_n E[X_n^-] \leq 2 \sup_n \|X_n\|_1.$$

Proof. Put $T(\omega) = \min\{n : X_n \geq \alpha\}$ and define a sequence of stopping times $\{T_k = T \wedge k\}_{k \in \mathbb{Z}^+}$. By Theorem 4.2.1, for each k, $E[X_{T_k}] \leq E[X_0]$. Either

$$X_{T_k}(\omega) \geq \alpha \quad \text{or} \quad X_{T_k}(\omega) = X_k(\omega),$$

therefore,

$$\alpha P\big(\sup_{n \leq k} X_n \geq \alpha\big) + \int_{\{\sup_{n \leq k} X_n < \alpha\}} X_k\, dP \leq E[X_{T_k}] \leq E[X_0].$$

As $\alpha \geq 0$, we know $\{X_k < 0\} \subseteq \{X_k < \alpha\} \subseteq \{\sup_{n \leq k} X_n < \alpha\}$. Hence

$$-E[X_k^-] = \int_{\{X_k \leq 0\}} X_k\, dP \leq \int_{\{\sup_{n \leq k} X_n < \alpha\}} X_k\, dP,$$

and so

$$\alpha P\big(\sup_{n\le k} X_n \ge \alpha\big) \le E[X_0] + E[X_k^-].$$

Letting $k \to \infty$ we have

$$\alpha P\big(\sup_{n} X_n \ge \alpha\big) \le E[X_0] + \sup_n E[X_n^-] \le 2\sup_n \|X_n\|_1.$$

□

Lemma 4.5.2. *Suppose $\{X_n\}_{n\in\mathbb{Z}^+}$, is a supermartingale. For every $\alpha \ge 0$,*

$$\alpha P\big(\inf_n X_n \le -\alpha\big) \le \sup_n E[X_n^-].$$

Proof. Put $S(\omega) = \min\{n : X_n(\omega) \le -\alpha\}$ and define a sequence of stopping times $\{S_k := S \wedge k\}_{k\in\mathbb{Z}^+}$. By Theorem 4.2.1, similarly to in the previous lemma we know $E[X_{S_k}] \ge E[X_k]$ for every $k \in \mathbb{Z}^+$. Therefore,

$$E[X_k] \le -\alpha P\big(\inf_{n\le k} X_n \le -\alpha\big) + \int_{\{\inf_{n\le k} X_n > -\alpha\}} X_k\, dP,$$

so

$$\begin{aligned}\alpha P\big(\inf_{n\le k} X_n \le -\alpha\big) &\le E[-X_k] + \int_{\{\inf_{n\le k} X_n > -\alpha\}} X_k\, dP\\ &= \int_{\{\inf_{n\le k} X_n \le -\alpha\}} (-X_k)\, dP \\ &\le E[X_k^-].\end{aligned} \tag{4.1}$$

Letting $k \to \infty$, the result follows. □

The following result is a corollary to Lemmata 4.5.1 and 4.5.2.

Corollary 4.5.3. *Suppose $\{X_n\}_{n\in\mathbb{Z}^+}$ is a supermartingale. For every $\alpha \ge 0$,*

$$\alpha P\big(\sup_n |X_n| \ge \alpha\big) \le 3\sup_n \|X_n\|_1.$$

Corollary 4.5.4 (Doob's Maximal Inequality). *Suppose $\{X_n\}_{n\in\mathbb{Z}^+}$ is a martingale. For every $\alpha \ge 0$,*

$$\alpha P\big(\sup_n |X_n| \ge \alpha\big) \le \sup_n \|X_n\|_1.$$

Proof. From Lemma 4.1.5, if $Y_n = -|X_n|$, then Y is a (negative) supermartingale and

$$\|Y_n\|_1 = \|X_n\|_1 = E[Y_n^-].$$

Also

$$\big\{\inf_n Y_n \le -\alpha\big\} = \big\{\sup_n |X_n| \ge \alpha\big\},$$

so the result follows from Lemma 4.5.2. □

This result gives us control over the probability the maximum exceeds a given value. Another useful result would give control over the value of the maximum in L^p norm. To prove this, we first prove the following lemma. Similar estimates are given in Lemmata 8.2.18 and 11.5.1.

Lemma 4.5.5. *Suppose X and Y are two nonnegative random variables defined on the probability space $(\Omega, \mathcal{F}, P)$ such that $X \in L^p$ for some $p \in]1, \infty[$, and for every $\alpha > 0$, $\alpha P(Y \geq \alpha) \leq \int_{\{Y \geq \alpha\}} X \, dP$. Then $\|Y\|_p \leq q\|X\|_p$, where $p^{-1} + q^{-1} = 1$.*

Proof. Let $\tilde{F}(\lambda) = 1 - F_Y(\lambda) = P(Y > \lambda)$ where F_Y is the distribution function of Y. As λ^p is continuous, integration by parts (Theorem 1.3.43) yields,

$$
\begin{aligned}
E[Y^p] &= -\int_{[0,\infty]} \lambda^p \, dF(\lambda) \\
&= \int_{[0,\infty]} \tilde{F}(\lambda) \, d(\lambda^p) - \lim_{h \to \infty} [\lambda^p \tilde{F}(\lambda)]_0^h \\
&\leq \int_{[0,\infty]} \tilde{F}(\lambda) \, d(\lambda^p) \\
&\leq \int_{[0,\infty]} \lambda^{-1} \Big(\int_{\{Y \geq \lambda\}} X \, dP \Big) \, d(\lambda^p) && \text{by hypothesis} \\
&= E\Big[X \int_{[0,Y]} \lambda^{-1} \, d(\lambda^p) \Big] && \text{by Fubini's theorem} \\
&= \Big(\frac{p}{p-1} \Big) E[XY^{p-1}] \\
&\leq q\|X\|_p \|Y^{p-1}\|_q && \text{by Hölder's inequality.}
\end{aligned}
$$

We have, therefore, proved that

$$E[Y^p] \leq q\|X\|_p (E[Y^{pq-q}])^{\frac{1}{q}} = q\|X\|_p (E[Y^p])^{\frac{1}{q}}.$$

If $\|Y\|_p$ is finite, as $1 - q^{-1} = p^{-1}$ the inequality follows immediately. Otherwise, the random variable $Y_n := Y \wedge n$ satisfies the hypotheses and is in $L^p(\Omega, \mathcal{F}, P)$ for every n. Therefore

$$\|Y_n\|_p \leq q\|X\|_p,$$

and the result follows by letting $n \to \infty$ and monotone convergence. □

Theorem 4.5.6 (Doob's L^p Inequality). *Suppose X is a martingale or nonnegative submartingale. Then, for $p \in]1, \infty]$, we have*

$$\sup_n |X_n| \in L^p \text{ if and only if } \sup_n \|X_n\|_p < \infty.$$

Furthermore, for $p \in]1, \infty[$ and $p^{-1} + q^{-1} = 1$ we have

$$\big\| \sup_n |X_n| \big\|_p \leq q \sup_n \|X_n\|_p.$$

Proof. When $p = \infty$ the first part of the theorem is immediate. Clearly, for $1 < p \leq \infty$ if $\sup_n |X_n| \in L^p$, then $\sup_n \|X_n\|_p \leq \| \sup_n X_n\|_p < \infty$.

To show the converse, we know by assumption that

$$\sup_n E[(-X_n)^-] = \sup_n E[X_n^+] \leq \sup_n E\big[|X|^p\big] = \sup_n \|X_n\|_p^p < \infty$$

and from Theorem 4.4.1 applied to the supermartingale $-X$ we know $\lim_{n\to\infty} X_n(\omega) = X_\infty(\omega)$ exists and is integrable. By Fatou's lemma

$$\begin{aligned} E\big[|X_\infty|^p\big] &= E\big[|\lim_n X_n|^p\big] = E\big[\lim_n |X_n|^p\big] \\ &\leq \liminf_n E\big[|X_n|^p\big] \leq \sup_n E\big[|X_n|^p\big] < \infty, \end{aligned}$$

so $X_\infty \in L^p$ and $\|X_\infty\|_p \leq \sup_n \|X_n\|_p$.

From (4.1), as $-|X|$ is a supermartingale, for any $\alpha > 0$ we have

$$\begin{aligned} \alpha P\big(\sup_{n\leq k} |X_n| \geq \alpha\big) &= \alpha P\big(\inf_{n\leq k} (-|X_n|) \leq -\alpha\big) \\ &\leq \int_{\{\sup_{n\leq k} |X_n| \geq \alpha\}} X_k \, dP \\ &\leq \int_{\{\sup_n |X_n| \geq \alpha\}} X_k^+ \, dP. \end{aligned}$$

Letting $k \to \infty$, as $X_k^+ \leq \sup_n |X_n|$, which is integrable, by dominated convergence we have that for any $\alpha > 0$

$$\alpha P(\sup_n |X_n| \geq \alpha) \leq \lim_k \int_{\{\sup_n |X_n| \geq \alpha\}} X_k^+ \, dP = \int_{\{\sup_n |X_n| \geq \alpha\}} X_\infty^+ \, dP.$$

Consequently, we can apply Lemma 4.5.5 with $Y = \sup_n |X_n|$ and $X = X_\infty^+$ to deduce that

$$\| \sup_n |X_n| \|_p \leq q \|X_\infty^+\|_p \leq q \|X_\infty\|_p.$$

□

4.6 Decomposition of Supermartingales

Definition 4.6.1. *Suppose $X = \{X_n\}_{n\in\mathbb{Z}^+}$ is a nonnegative supermartingale. Then X is said to be a* potential *if* $\lim_n E[X_n] = 0$.

Example 4.6.2. An example of a potential is provided by the wealth of a man condemned to play an unfair game until he loses all his money (for example, where his expected wealth after a round is some fixed fraction of his wealth before the round).

Remark 4.6.3. From Corollary 4.4.3 we know that, for any potential X, the limit $\lim_n X_n(\omega) = X_\infty(\omega)$ exists almost surely, and by Fatou's inequality $E[X_\infty] = 0$. Consequently $X_\infty = 0$ a.s., and the convergence also takes place in L^1.

Remark 4.6.4. For X a nonnegative supermartingale with $X_t \to 0$ a.s., by Fatou's inequality we know $\lim_n E[X_n] = 0$. Therefore, a nonnegative supermartingale X is a potential if and only if

(i) $X_t(\omega) \geq 0$ a.s.
(ii) $\lim_{t\to\infty} X_t(\omega) = 0$ a.s.

The following Riesz decomposition for supermartingales can now be established.

Theorem 4.6.5 (Riesz Decomposition). *Suppose $\{X_n\}_{n\in\mathbb{Z}^+}$ is a supermartingale. Then the following are equivalent:*

(i) $\lim_n E[X_n] > -\infty$,
(ii) there is a submartingale Y' such that $Y'_n \leq X_n$ a.s. for all $n \in \mathbb{Z}^+$,
(iii) there is a martingale Y and a potential Z such that for each $n \in \mathbb{Z}^+$, we have $X_n = Y_n + Z_n$.

These two processes Y and Z are then unique and, if Y' is any submartingale such that $Y'_n \leq X_n$ a.s. for all $n \in \mathbb{Z}^+$, then $Y'_n \leq Y_n$ a.s. for all $n \in \mathbb{Z}^+$.

Proof. We shall show that (iii)⇒(ii)⇒(i)⇒(iii). If (iii) is satisfied, then $Y_n \leq X_n$ so (ii) is true. If (ii) is satisfied, then

$$E[X_n] \geq E[Y'_n] \geq E[Y'_0] \geq -\infty,$$

so $E[X_n]$ is bounded by $E[Y'_0]$, hence (i) is true.

Suppose (i) holds. For $p \in \mathbb{Z}^+$ write

$$X_{n,p} = E[X_{n+p}|\mathcal{F}_n] \leq X_n \quad \text{a.s.},$$

so

$$\begin{aligned} X_{n,p+1} &= E\big[E[X_{n+p+1}|\mathcal{F}_{n+p}]\big|\mathcal{F}_n\big] \\ &\leq E[X_{n+p}|\mathcal{F}_n] = X_{n,p} \quad \text{a.s.} \end{aligned}$$

Therefore, $X_{n,p}$ is almost surely decreasing in p.

Define Y by $Y_n = \lim_p X_{n,p}$, so

$$X_n \geq \lim_p X_{n,p} = Y_n \quad \text{a.s.},$$

and, by conditional monotone convergence (Lemma 2.4.3), for $m \geq 0$

$$\begin{aligned} E[Y_{n+m}|\mathcal{F}_n] &= E[\lim_p X_{n+m,p}|\mathcal{F}_n] \\ &= \lim_p E[X_{n+m+p}|\mathcal{F}_n] \\ &= Y_n \quad \text{a.s.} \end{aligned}$$

We can also check that Y_n is integrable, so $\{Y_n\}_{n\in\mathbb{Z}^+}$ is a martingale. Write

$$Z_n = X_n - Y_n.$$

Clearly $Z_n(\omega) \geq 0$ a.s., so $\{Z_n\}_{n\in\mathbb{Z}^+}$ is a nonnegative supermartingale.

From the definition of Y,

$$\lim_p E[Z_{n+p}|\mathcal{F}_n] = \lim_p E[X_{n+p}|\mathcal{F}_n] - Y_n = 0 \quad \text{a.s.}$$

for each $n \in \mathbb{N}$. As Z is a supermartingale, $E[Z_p|\mathcal{F}_0]$ is a.s. nonincreasing in p. By the corollary to the monotone convergence theorem (Corollary 1.3.32), $\lim_p E[Z_p] = E[\lim_p E[Z_p|\mathcal{F}_0]] = 0$, so Z is a potential. Hence (iii) holds.

Finally, suppose $X = Y' + Z'$ is a second decomposition with Y' a martingale and Z' a potential. Then for any $n \in \mathbb{Z}^+$

$$\begin{aligned} E[X_{n+p}|\mathcal{F}_n] &= E[Y'_{n+p}|\mathcal{F}_n] + E[Z'_{n+p}|\mathcal{F}_n] \\ &= Y'_n + E[Z'_{n+p}|\mathcal{F}_n]. \end{aligned}$$

Letting $p \to \infty$, $\lim_p E[X_{n+p}|\mathcal{F}_n] = Y_n$ a.s. and $\lim_p E[Z^*_{n+p}|\mathcal{F}_n] = 0$ a.s. The result follows. □

We can now extend the optional stopping theorem (Theorem 4.2.1) to possibly infinite stopping times. To do so, we shall use the following lemma.

Lemma 4.6.6. *For any integrable random variable X, any stopping time S and any $m \in \mathbb{Z}^+$,*

$$E[I_{\{m\geq S\}}X|\mathcal{F}_{\{S\wedge m\}}] = I_{\{m\geq S\}}E[X|\mathcal{F}_S] = I_{\{m\geq S\}}E[X|\mathcal{F}_m].$$

Proof. Note that $I_{m\geq S}E[X|\mathcal{F}_S]$ is $\mathcal{F}_{S\wedge m}$ measurable. As $\mathcal{F}_{S\wedge m} \subseteq \mathcal{F}_S$, by Lemmata 2.4.7 and 2.4.8 we have

$$I_{\{m\geq S\}}E[X|\mathcal{F}_S] = E[I_{\{m\geq S\}}E[X|\mathcal{F}_S]|\mathcal{F}_{S\wedge m}] = E[I_{\{m\geq S\}}X|\mathcal{F}_{S\wedge m}].$$

The second equality follows by a similar argument. □

Theorem 4.6.7 (Optional Stopping–Unbounded Stopping Times). *Suppose $\{X_n\}_{n\in\mathbb{Z}^+}$ is a uniformly integrable or nonnegative $\{\mathcal{F}_n\}_{n\in\mathbb{Z}^+}$ supermartingale and $X_\infty = \lim_n X_n$. If S and T are two stopping times such that $S \leq T$ a.s., then X_S and X_T are integrable random variables and $E[X_T|\mathcal{F}_S] \leq X_S$ a.s. If $\{X_n\}_{n\in\mathbb{Z}^+}$ is a martingale, then $X_S = E[X_T|\mathcal{F}_S] = E[X_\infty|\mathcal{F}_S]$.*

Proof. Put $Y_n = E[X_\infty|\mathcal{F}_n] \leq X_n$ a.s. and $Z_n = X_n - Y_n$, so that Y is a uniformly integrable martingale and Z is a potential (Remark 4.6.4). The result will be proved for Y and Z.

Consider first the uniformly integrable martingale Y. From the optional stopping theorem for bounded stopping times, we have that, for any $m \in \mathbb{N}$,

$$Y_{S\wedge m} = E[Y_{T\wedge m}|\mathcal{F}_{S\wedge m}].$$

We can then write, as $S \leq T$,

$$Y_{T\wedge m} = I_{\{m<S\}} Y_m + I_{\{m\geq S\}} Y_{T\wedge m}$$

and so, by Lemma 4.6.6,

$$Y_{S\wedge m} = E[Y_{T\wedge m}|\mathcal{F}_{S\wedge m}] = I_{\{m<S\}} Y_m + I_{\{m\geq S\}} E[Y_{T\wedge m}|\mathcal{F}_s] = E[Y_{T\wedge m}|\mathcal{F}_S].$$

We now note that $Y_{S\wedge m} = Y^S_m$ and, as Y^S is also a uniformly integrable martingale, by convergence of uniformly integrable martingales a.s. and in L^1 (cf. Exercise 2.7.8) we have

$$Y_S = Y^S_\infty = \lim_m Y_{S\wedge m} = E[\lim_m Y_{T\wedge m}|\mathcal{F}_S] = E[Y_T|\mathcal{F}_S].$$

Similarly, for the potential Z we have

$$Z_{S\wedge m} \geq E[Z_{T\wedge m}|\mathcal{F}_{S\wedge m}] = E[Z_{T\wedge m}|\mathcal{F}_S].$$

Taking $m \to \infty$, as Z is a nonnegative supermartingale it converges a.s., so we have $Z_S = \lim_m Z^S_m$ and by Fatou's inequality

$$Z_S = \lim_m Z_{S\wedge m} \geq \lim_m E[Z_{T\wedge m}|\mathcal{F}_S] \geq E[\lim_m Z_{T\wedge m}|\mathcal{F}_S] = E[Z_T|\mathcal{F}_S].$$

Hence

$$X_S = Y_S + Z_S \geq E[Y_T + Z_T|\mathcal{F}_S] = E[X_T|\mathcal{F}_S].$$

If X is a martingale, then $Z_t = 0$ a.s. for all t, and the equality is clear. □

Remark 4.6.8. Fundamentally, the key to this theorem is that X must be a supermartingale for $n \in \overline{\mathbb{Z}}^+$, and not just $n \in \mathbb{Z}^+$. In particular, X_∞ must exist and have well-defined conditional expectations, which is guaranteed when X is uniformly integrable or nonnegative.

Corollary 4.6.9. *Suppose that $\{\mathcal{F}_n\}_{n\in\overline{\mathbb{Z}}^+}$ is a filtration, that $\mathcal{F}_{\infty-} = \mathcal{F}_\infty$ and that S and T are two stopping times. (We do not suppose $S \leq T$.) Then the projection operators $E[\cdot|\mathcal{F}_S]$ and $E[\cdot|\mathcal{F}_T]$ commute, and their composition is $E[\cdot|\mathcal{F}_{S\wedge T}]$.*

Proof. Suppose that Y is any integrable $\mathcal{F}_\infty$-measurable random variable. Consider the martingale $\{X_n = E[Y|\mathcal{F}_n]\}_{n\in\mathbb{Z}^+}$, so that $X_\infty = Y$ a.s. Then by Theorem 4.6.7 and Lemma 4.6.6, as $T \wedge n \leq T$

$$X_n^T = X_{T\wedge n} = E[X_T|\mathcal{F}_{T\wedge n}] = I_{\{T<n\}}X_n + E[I_{\{T\geq n\}}X_T|\mathcal{F}_n] = E[X_T|\mathcal{F}_n].$$

Again applying Theorem 4.6.7,

$$X_{T\wedge S} = X_S^T = E[X_T|\mathcal{F}_S] = E\big[E[X_\infty|\mathcal{F}_T]\big|\mathcal{F}_S\big].$$

However, the left-hand side is symmetric in S and T and has the value $E[X_\infty|\mathcal{F}_{S\wedge T}]$. Because $X_\infty = Y$ is arbitrary, the result is proven. □

Remark 4.6.10. Once the optional stopping theorem is proved for continuous-time filtrations (see Theorem 5.3.1), it is immediate that the result of Corollary 4.6.9 is valid for stopping times with respect to a continuous-time filtration. The above proof applies almost word for word.

Remark 4.6.11. If S and T are bounded, then the assumption $\mathcal{F}_{\infty-} = \mathcal{F}_\infty = \mathcal{F}$ is unnecessary. (This follows by defining $\mathcal{F}'_t = \mathcal{F}_{t\wedge k}$, where k is an upper bound on $T \vee S$, and applying the result with filtration $\{\mathcal{F}'_t\}_{t\in\mathbb{T}}$.)

4.7 Exercises

Exercise 4.7.1. Prove the assertions made in Example 4.1.4.

Exercise 4.7.2. Suppose $X_1, X_2, \ldots$ are iid $N(0,1)$ random variables and $S_n := X_1 + \cdots + X_n$. If $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$, which of the following processes are martingales with respect to $\{\mathcal{F}_n\}_{n\in\mathbb{Z}^+}$?

(i) $Y_n = S_n$
(ii) $Y_n = S_n^2 - n$
(iii) $Y_n = \left(\sum_{m\leq n} X_m^2\right) - n$
(iv) $Y_n = I_{\{n>0\}}S_{n-1}$,
(v) $Y_n = S_n/n$
(vi) $Y_n = e^{(S_n - n/2)}$

Exercise 4.7.3. Let $X_0, X_1, X_2, \ldots$ be independent random variables in $L^2(\Omega, \mathcal{F}, P)$ with the same expectation μ and $\sup_i \|X_i\|_2 < \infty$. Let $\{\mathcal{F}_n\}_{n\in\mathbb{Z}^+}$ be a filtration such that the process $\{X_n\}_{n\in\mathbb{Z}^+}$ is adapted.

(i) Show that the process

$$\hat{X}_n = \sum_{i\leq n} \frac{X_i - \mu}{i}$$

is a martingale.

(ii) Show that $E[\hat{X}_n^2]$ is uniformly bounded, and hence $\hat{X}_n$ is uniformly integrable.
(iii) Hence show that $\hat{X}_n$ converges almost surely to $\hat{X}_\infty$.
(iv) Using Kronecker's lemma (below), show the *strong law of large numbers*, that is, $n^{-1}\sum_{i\le n}(X_n - \mu) \to 0$ a.s.

Kronecker's lemma: Let $\{a_n\}_{n\in\mathbb{Z}^+}$ and $\{b_n\}_{n\in\mathbb{Z}^+}$ be sequences of real numbers, with $0 < b_n < b_{n+1}$ for all n and $b_n \to \infty$. If $\sum_{i\le n} a_i/b_i$ converges to a finite limit, then $b_n^{-1}\sum_{i\le n} a_n \to 0$.

Exercise 4.7.4. Suppose that $\{X_n\}_{n\in\mathbb{Z}^+}$ is a stochastic process such that $E\big[|X_n|\big] < \infty$ for each $n \in \mathbb{Z}^+$. Show that we can write $X = Y + Z$, where Y is a submartingale and Z is a supermartingale.

Exercise 4.7.5. Let S be the space of points with integer coordinates in d-dimensional space, $d \ge 1$. A function h on S is said to be *harmonic* if h is equal to the average of h on its $2d$ nearest neighbours.

Define a random walk X on S as follows: Let $X_0 = s_0$ be any arbitrary initial point. For each $n \ge 1$, on the set $X_n = s_n$, we choose X_{n+1} with uniform probability $1/(2d)$ from the $2d$ neighbours of s_n, independently of $\{X_k\}_{k<n}$. If h is harmonic on S, show that the sequence $\{h(X_n)\}_{n\in\mathbb{Z}^+}$ is a martingale.

Similarly a function h on S is said to be *superharmonic* (*subharmonic*) if h is greater than or equal (less than or equal to) to the average of h on its $2d$ nearest neighbours. If $\{X_n\}_{n\in\mathbb{Z}^+}$ is a walk on S and h is superharmonic (subharmonic), show that $\{h(X_n)\}_{n\in\mathbb{Z}^+}$ is a supermartingale (submartingale).

Exercise 4.7.6. (Doob decomposition) Let $\{X_n\}_{n\in\mathbb{Z}^+}$ be an adapted process with respect to $\{\mathcal{F}_n\}_{n\in\mathbb{Z}^+}$, and suppose $E\big[|X_n|\big] < \infty$ for all n. Show that there is a unique process $\{C_n\}_{n\in\mathbb{Z}^+}$ where C_n is $\mathcal{F}_{n-1}$-measurable for all $n \ge 1$, and a unique martingale $\{Y_n\}_{n\in\mathbb{Z}^+}$, such that $X_n = C_n + Y_n$ a.s. and $E[Y_0] = 0$. Show that C is nondecreasing if and only if X is a submartingale.

Exercise 4.7.7. Consider the classic coin-tossing game, where an initial stake X_n is doubled when a head is thrown (with probability 1/2), and lost when a tail is thrown. Let $Y_n(\omega) \in \mathbb{R}$ be the running profit after n turns of a player who can take on an arbitrarily large debt, and starts with $Y_0 = 1$.

(i) Show that, for any integrable sequence of stakes, $\{Y_n\}_{n\in\mathbb{Z}^+}$ is a martingale.
(ii) Hence explain, using the optional stopping theorem, why there is no sequence of stakes which guarantee the player a positive mean profit over any finite horizon.
(iii) Now consider the *doubling strategy* (also known as a 'martingale' strategy, in a different sense of the term), where a player begins with a stake of \$1, and proceeds to double their bet after each loss. Show that the profit following the first win is \$1 (almost surely). Explain why the optional stopping theorem does not hold in this case.

Exercise 4.7.8. Give an example of a martingale which converges a.s. but not in L^1.

Exercise 4.7.9. Let Z and Z' be potentials

(i) Show that $Z - Z'$ is a martingale if and only if $Z = Z'$.
(ii) Give conditions under which $Z - Z'$ is a potential.

Exercise 4.7.10. Let Y be an adapted process with conditional distributions $Y_n|_{\mathcal{F}_{n-1}} \sim N(Y_{n-1}, n^{-\alpha})$ and $Y_0 = 1$, for some $\alpha > 1$. Show that $\lim_n Y_n$ exists almost surely and find its distribution. (The results of Lemma 5.5.3 may be useful.)

5

Martingales in Continuous Time

In this chapter we extend our discussion of martingales to allow a continuous-time processes. Throughout we take $\mathbb{T} = [0, \infty[$ or $[0, \infty]$.

5.1 Definitions and Basic Properties

The definition of a continuous-time martingale is analogous to that of a discrete-time martingale.

Definition 5.1.1. *A real-valued stochastic process* $\{X_t\}_{t \in \mathbb{T}}$ *is said to be a* supermartingale *with respect to the filtration* $\{\mathcal{F}_t\}$ *if*

(i) each X_t *is* $\mathcal{F}_t$*-measurable, i.e.* $\{X_t\}$ *is adapted to* $\{\mathcal{F}_t\}$*,*
(ii) $E\big[|X_t|\big] < \infty$*, for all* $t \in \mathbb{T}$*, and*
(iii) $X_t \geq E[X_s|\mathcal{F}_t]$ *almost surely for* $s \geq t$*.*

If "$\geq$*" in property (iii) is replaced by "*$\leq$*", then* X *is said to be a* submartingale*. If the process* X *is both a submartingale and a supermartingale, then it is called a* martingale*.*

Suppose X is a real-valued stochastic process, α and β are real numbers with $\alpha < \beta$, and I is a subset of $[0, \infty[$. Similarly to the notation of the previous chapter, $M(\omega, X, I, [\alpha, \beta])$ will denote the (possibly infinite) number of distinct upcrossings of $[\alpha, \beta]$ by the function $X_t(\omega)$ as t runs through I in an increasing manner. $D(\omega, X, I, [\alpha, \beta])$ will denote the number of downcrossings. We first extend to continuous parameter martingales the inequalities of the last chapter. We will denote the rational numbers by $\mathbb{Q}$.

S.N. Cohen, R.J. Elliott, *Stochastic Calculus and Applications*,
Probability and Its Applications, DOI 10.1007/978-1-4939-2867-5_5

Theorem 5.1.2 (Doob's Inequalities). *Let X be a right-continuous supermartingale and $J = [u, v]$ an interval contained in $\mathbb{T}$. Then, for any $\lambda > 0$ and any real interval $[\alpha, \beta]$, the following inequalities hold:*

(i) $\lambda P(\sup_{t \in J} X_t \geq \lambda) \leq E[X_u] + E[X_v^-]$,
(ii) $\lambda P(\inf_{t \in J} X_t \leq -\lambda) \leq E\big[|X_v|\big]$,
(iii) $E[M(\omega, X, J; [\alpha, \beta])] \leq (\beta - \alpha)^{-1} E[(X_v - \alpha)^-]$,
(iv) $E[D(\omega, X, J; [\alpha, \beta])] \leq (\beta - \alpha)^{-1} E[(X_u \wedge \beta) - (X_v \wedge \beta)]$.

Proof. Suppose F is a finite subset of $\mathbb{Q}_J := (\mathbb{Q} \cap J) \cup \{u, v\}$. Then these inequalities, with F replacing J, follow from Lemma 4.5.1, Lemma 4.5.2 and Corollary 4.3.3. By considering an increasing family J_n of subsets of $\mathbb{Q}_J$ such that $\bigcup_{n=1}^{\infty} J_n = \mathbb{Q}_J$, we see that the inequalities hold when J is replaced by $\mathbb{Q}_J$. By right-continuity, the left-hand sides of the inequalities are unchanged when $\mathbb{Q}_J$ is replaced by J, and so the result follows. □

Theorem 5.1.3 (Doob's L^p Inequality). *Let X be a right-continuous martingale (or nonnegative submartingale) and $J = [u, v] \subset \mathbb{T}$. If $p \in]1, \infty[$, $p^{-1} + q^{-1} = 1$ and $X_v \in L^p(\Omega, \mathcal{F}, P)$, then*

$$\Big\| \sup_{t \in J} |X_t| \Big\|_p \leq q \sup_{t \in J} \|X_t\|_p.$$

Proof. In a similar way to the last proof, consider an increasing sequence of finite subsets of $\mathbb{Q}_J$. The result on each subset follows from Theorem 4.5.6. By right-continuity, the result holds on J. □

5.1.1 Continuity of Paths

We next investigate the behaviour of the sample paths of a supermartingale.

Theorem 5.1.4. *Suppose X is a right-continuous supermartingale. Then X a.s. has left-hand limits at every time, that is, X is a.s. càdlàg, and almost every sample path is bounded on every compact interval.*

Proof. The boundedness follows immediately from inequalities (i) and (ii) of Theorem 5.1.2. For $J = [0, n]$ and $\alpha, \beta \in \mathbb{Q}$ with $\alpha < \beta$, from inequality (iii) of Theorem 5.1.2 we have that $E[M(\omega, X, [0, n]; [\alpha, \beta])] < \infty$ a.s. Therefore, the event

$$H_{\alpha,\beta}^{(n)} = \{\omega \in \Omega : M(\omega, X, [0, n]; [\alpha, \beta]) = \infty\}$$

has probability zero. Write

$$H^{(n)} = \bigcup_{\substack{\alpha, \beta \in \mathbb{Q} \\ \alpha < \beta}} H_{\alpha,\beta}^{(n)},$$

then $H^{(n)}$ also has probability zero. But $H^{(n)}$ corresponds to those sample paths $X_t(\omega)$ for which there is a $t \in [0, n]$ with

$$\liminf_{s \uparrow t} X_s(\omega) < \limsup_{s \uparrow t} X_s(\omega).$$

Therefore, if $\omega \notin H^{(n)}$, for any t,

$$\liminf_{s \uparrow t} X_s(\omega) = \limsup_{s \uparrow t} X_s(\omega) = X_{t-}(\omega).$$

□

A related result is the following.

Lemma 5.1.5. *Suppose X is a supermartingale with respect to the filtration $\{\mathcal{F}_t\}_{t \in \mathbb{T}}$. Then the restriction to $\mathbb{Q}$ of the map $s \mapsto X_s(\omega)$ has a left and right limit at every point $t \in \mathbb{T}$ for almost every $\omega \in \Omega$.*

Proof. Adapt the notation of Theorem 5.1.4, and write

$$H^{(n)}_{\alpha,\beta} = \{\omega \in \Omega : M(\omega, X, \mathbb{Q} \cap [0, n]; [\alpha, \beta]) = \infty\}$$

so that, from the proof of Theorem 5.1.2, $H^{(n)} = \bigcup_{\alpha,\beta \in \mathbb{Q}} H^{(n)}_{\alpha,\beta}$ has probability zero. Therefore, for $\omega \notin H^{(n)}$ and $t \in [0, n[$

$$\limsup_{s \uparrow t, s \in \mathbb{Q}} X_s(\omega) = \liminf_{s \uparrow t, s \in \mathbb{Q}} X_s(\omega) = X_{t-}(\omega)$$

and

$$\limsup_{s \downarrow t, s \in \mathbb{Q}} X_s(\omega) = \liminf_{s \downarrow t, s \in \mathbb{Q}} X_s(\omega) = X_{t+}(\omega).$$

□

To prove further properties of these limits, a uniform integrability result is needed.

Lemma 5.1.6. *Let X be a supermartingale with respect to the filtration $\{\mathcal{F}_t\}_{t \in \mathbb{T}}$. Let S_n be a nonincreasing sequence of stopping times (such that for each n, $X_{S_{n+1}} \geq E[X_{S_n} | \mathcal{F}_{S_{n+1}}]$ a.s.). Then $\{X_{S_n}\}_{n \in \mathbb{N}}$ is uniformly integrable.*

Remark 5.1.7. If $\{S_n\}_{n \in \mathbb{N}}$ is a deterministic family, the condition $X_{S_{n+1}} \geq E[X_{S_n} | \mathcal{F}_{S_{n+1}}]$ is simply the definition of a supermartingale. Using this simple case, we shall prove the optional stopping theorem (Theorem 5.3.1). However, once we have that result, we shall see that the condition $X_{S_{n+1}} \geq E[X_{S_n} | \mathcal{F}_{S_{n+1}}]$ is satisfied for *any* nonincreasing sequence of stopping times, so can be omitted.

Proof. As $E[X_{S_n}]$ is an increasing function of n, the limit $\alpha = \lim_n E[X_{S_n}]$ exists. As $E[X_{S_n}] \leq E[X_0]$ for all n, we know $\alpha < \infty$. For any $\epsilon > 0$, there exists an integer k such that $\alpha - E[X_{S_k}] < \epsilon/2$ and so, for all $n \geq k$,

$$0 \leq E[X_{S_n}] - E[X_{S_k}] \leq \epsilon/2.$$

Consider any $\lambda > 0$ and suppose $n \geq k$. Then

$$\begin{aligned} I(n,\lambda) &:= \int_{\{|X_{S_n}|>\lambda\}} |X_{S_n}|\,dP \\ &= \int_{\{X_{S_n}<-\lambda\}} (-X_{S_n})\,dP + \int_{\{X_{S_n}>\lambda\}} X_{S_n}\,dP \\ &= -\int_{\{X_{S_n}<-\lambda\}} X_{S_n}\,dP + E[X_{S_n}] - \int_{\{X_{S_n}\leq\lambda\}} X_{S_n}\,dP \\ &\leq -\int_{\{X_{S_n}<-\lambda\}} X_{S_n}\,dP + E[X_{S_k}] - \int_{\{X_{S_n}\leq\lambda\}} X_{S_n}\,dP + \epsilon/2. \end{aligned}$$

As $X_{S_n} \geq E[X_{S_k}|\mathcal{F}_{S_n}]$, we have that

$$\begin{aligned} &\int_{\{X_{S_n}\leq\lambda\}} X_{S_n}\,dP + \int_{\{X_{S_n}<-\lambda\}} X_{S_n}\,dP \\ &\geq \int_{\{X_{S_n}\leq\lambda\}} X_{S_k}\,dP + \int_{\{X_{S_n}<-\lambda\}} X_{S_k}\,dP \end{aligned}$$

and so, by rearrangement,

$$I(n,\lambda) \leq \int_{\{|X_{S_n}|>\lambda\}} |X_{S_k}|\,dP + \epsilon/2.$$

By Jensen's inequality, we also have that

$$E\big[|X_{S_n}|\big] = E[X_{S_n}] + 2E[X_{S_n}^-] \leq \alpha + 2E[X_{S_0}^-] =: \beta$$

and by Markov's inequality (Exercise 2.7.3),

$$P(|X_{S_n}| > \lambda) \leq \frac{E\big[|X_{S_n}|\big]}{\lambda} \leq \frac{\beta}{\lambda}.$$

By absolute continuity of the measure $\nu(A) = \int_A |X_{S_k}|\,dP$ with respect to P, from Lemma 1.6.2 we know that there exists a λ_0 such that

$$\int_{\{|X_{S_n}|>\lambda\}} |X_{S_k}|\,dP \leq \epsilon/2$$

for all $\lambda \geq \lambda_0$, $n \geq k$. That is, $I(n,k) \leq \epsilon$ for all $\lambda \geq \lambda_0$, $n \geq k$. For $n < k$, there exists a λ_1 such that $I(n,k) \leq \epsilon$ whenever $n < k$, $\lambda \geq \lambda_1$, and so if $\lambda \geq \lambda_0 \vee \lambda_1$, Definition 2.5.1 is satisfied. □

Theorem 5.1.8. *Suppose X is a supermartingale. Then*

(i) $X_t \geq E[X_{t+}|\mathcal{F}_t]$ *a.s. and* $X_{t-} \geq E[X_t|\mathcal{F}_{t-}]$ *a.s.*
(ii) The process $\{X_{t+}\}_{t\in\mathbb{T}}$ *is a supermartingale with respect to the right-continuous filtration* $\{\mathcal{F}_{t+}\}_{t\in\mathbb{T}}$.

(iii) *Suppose the filtration $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$ is right-continuous. Then the supermartingale X has a right-continuous modification if and only if $E[X_t]$ is right-continuous in t.*

Proof. (i) Consider a decreasing deterministic sequence $\{t_n\}_{n\in\mathbb{N}} \subset \mathbb{Q}$ converging to t. This sequence satisfies the requirements of Lemma 5.1.6, so the family $\{X_{t_n}\}_{n\in\mathbb{N}}$ is uniformly integrable. By Lemma 5.1.5, the limit $X_{t+} = \lim_n X_{t_n}$ exists a.s. and, as the family is uniformly integrable, by Theorem 2.5.8, for any $A \in \mathcal{F}_t$,

$$\lim_n E[I_A X_{t_n}] = E[I_A \lim_n X_{t_n}] = E[I_A X_{t+}],$$

that is, $E[X_{t+}|\mathcal{F}_t] = \lim_n E[X_{t_n}|\mathcal{F}_t]$. As $X_t \geq E[X_{t_n}|\mathcal{F}_t]$ a.s. for all n, this implies $X_t \geq E[X_{t+}|\mathcal{F}_t]$ a.s., as desired.

Next consider an increasing sequence $\{s_n\}_{n\in\mathbb{N}} \subset \mathbb{Q}$ such that $s_n < t$ for all n and $\{s_n\}_{n\in\mathbb{N}}$ converges to t. Write $Y_n = X_{s_n} - E[X_t|\mathcal{F}_{s_n}]$. Then $\{Y_n\}_{n\in\mathbb{N}}$ is a nonnegative $\{\mathcal{F}_{s_n}\}_{n\in\mathbb{N}}$ supermartingale such that $\lim_{n\to\infty} Y_n = Y_\infty = X_{t-} - E[X_t|\mathcal{F}_{t-}]$. Hence $Y_\infty \geq 0$ and the result follows.

(ii) Suppose $s < t$ and $\{s_n\}_{n\in\mathbb{N}}$, $\{t_n\}_{n\in\mathbb{N}}$ are decreasing sequences in $\mathbb{Q}$ with $\lim_{n\to\infty} s_n = s$ and $\lim_{n\to\infty} t_n = t$. Then, for any $A \in \mathcal{F}_{s+} \subseteq \mathcal{F}_{s_n}$, we have $E[I_A X_{s_n}] \geq E[I_A X_{t_n}]$. Letting $n \to \infty$, again by uniform integrability we see that $E[I_A X_{s+}] \geq E[I_A X_{t+}]$, that is, $X_{s+} \geq E[X_{t+}|\mathcal{F}_{s+}]$ a.s.

(iii) We now suppose that $\mathcal{F}_t = \mathcal{F}_{t+}$ for all $t \in \mathbb{T}$. Then by part (i), $X_t \geq X_{t+}$ a.s., so these random variables are equal if and only if they have the same expectation. Suppose $\{t_n\}_{n\in\mathbb{N}} \subset \mathbb{T}$ is a sequence decreasing to t. By uniform integrability (Lemma 5.1.6),

$$E[X_{t+}] = \lim_{n\to\infty} E[X_{t_n}].$$

Therefore, $X_t = X_{t+}$ a.s. if and only if this limit equals $E[X_t]$. Because $E[X_t]$ is monotonic in t, this is the same as saying it is right-continuous at every point $t \in \mathbb{T}$. Consequently, the process $\{X_{t+}\}_{t\in\mathbb{T}}$ is a right-continuous modification of $\{X_t\}_{t\in\mathbb{T}}$.

If $\{Y_t\}_{t\in\mathbb{T}}$ is a right-continuous modification of $\{X_t\}_{t\in\mathbb{T}}$, then $E[Y_t] = E[X_t]$ for all $t \in \mathbb{T}$, and so from the above argument $E[X_t]$ is right-continuous. □

Corollary 5.1.9. *If $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$ is right-continuous, then every martingale admits a càdlàg modification.*

5.2 Convergence Results

In this section, we establish continuous-time analogues to Theorem 4.4.1 and Corollaries 4.4.3–4.4.5.

Theorem 5.2.1. *Suppose X is a right-continuous supermartingale satisfying the bound $\sup_{t\in\mathbb{T}} E[X_t^-] < \infty$. Then X converges almost surely to an integrable random variable as $t \to \infty$.*

Proof. From Theorem 5.1.8(iii),

$$E[M(\omega, X, [0,n]; [\alpha,\beta])] \leq (\beta-\alpha)^{-1}(E[X_n^-] + |\alpha|),$$

so, letting $n \to \infty$,

$$E[M(\omega, X, \mathbb{T}; [\alpha,\beta])] \leq (\beta-\alpha)^{-1}\sup_n(E[X_n^-] + |\alpha|) < \infty.$$

The set $H_{\alpha,\beta} = \{\omega : M(\omega, X, \mathbb{T}; [\alpha,\beta]) = \infty\}$ then has measure zero. Writing

$$H = \bigcup_{\alpha,\beta\in\mathbb{Q},\alpha<\beta} H_{\alpha,\beta},$$

we see H also has measure zero. Consequently, the set $\{\omega : \limsup_{t\to\infty} X_t(\omega) > \liminf_{t\to\infty} X_t(\omega)\}$ has measure zero, that is,

$$X_\infty(\omega) := \liminf_{t\to\infty} X_t(\omega) = \lim_{t\to\infty} X_t(\omega) \quad a.s.$$

We know

$$\|X_t\|_1 = E[X_t^+ + X_t^-] = E[X_t] + 2E[X_t^-] \leq E[X_0] + 2E[X_t^-],$$

so, by Fatou's inequality,

$$E\big[|X_\infty|\big] \leq \sup_t \|X_t\|_1 < \infty,$$

and X_∞ is, therefore, integrable. □

Corollary 5.2.2. *Suppose $\{X_t\}_{t\in[0,\infty[}$ is a supermartingale such that $X_t \geq 0$ a.s. for every t. Clearly $X_t^- = 0$ a.s., so Theorem 5.2.1 applies and there exists a random variable X_∞ such that $\lim_{t\to\infty} X_t(\omega) = X_\infty(\omega)$ a.s. Furthermore, for $s > t$ and $A \in \mathcal{F}_t$, we know $E[I_A X_t] \geq E[I_A X_s]$. Letting $s \to \infty$, we have, by Fatou's inequality, that $X_t \geq E[X_\infty|\mathcal{F}_t]$ a.s., that is, $\{X_t\}_{t\in[0,\infty]}$ is a supermartingale.*

Corollary 5.2.3. *Suppose $\{X_t\}_{t\in[0,\infty[}$ is a uniformly integrable supermartingale. Then $\sup_t \|X_t\|_1 < \infty$, and, because $E[X_t^-] \leq \|X_t\|_1$, the condition of the theorem is satisfied and $\lim_{t\to\infty} X_t(\omega) = X_\infty(\omega)$ a.s. By uniform integrability the convergence also takes place in L^1, and $\{X_t\}_{t\in[0,\infty]}$ is a supermartingale.*

Corollary 5.2.4. *Suppose $\{X_t\}_{t\in[0,\infty[}$ is a uniformly integrable martingale. Then $\{X_t\}_{t\in[0,\infty]}$ is a uniformly integrable martingale and $X_t = E[X_\infty|\mathcal{F}_t]$ a.s. for each t.*

Corollary 5.2.5. *If Y is an integrable $\mathcal{F}_{\infty-}$-measurable random variable, then $\{E[Y|\mathcal{F}_t]\}_{t\in[0,\infty[}$ is a uniformly integrable martingale. We can take a right-continuous modification $\{Y_t\}_{t\in[0,\infty[}$ of this martingale, and $\lim_{t\to\infty} Y_t(\omega) = Y(\omega)$ a.s. and in $L^1(\Omega, \mathcal{F}, P)$.*

5.3 Optional Stopping

We now extend the optional stopping theorem to continuous-time supermartingales.

Theorem 5.3.1 (Doob's Optional Stopping Theorem). *Suppose X is a uniformly integrable or nonnegative right-continuous supermartingale with respect to the filtration $\{\mathcal{F}_t\}_{t\in[0,\infty]}$. If S and T are two stopping times such that $S \le T$ a.s., then the random variables X_S and X_T are integrable and $X_S \ge E[X_T|\mathcal{F}_S]$ a.s.*

Proof. Suppose n is a positive integer and write D_n for the set of all rationals of the form $2^{-n}k$ for $k \in \overline{\mathbb{Z}}^+$. Then $\{X_t^T\}_{t\in D_n}$, is a discrete-time supermartingale with respect to the filtration $\{\mathcal{F}_t\}_{t\in D_n}$.

For any number $\rho \in [0,\infty[$, define $\rho^{(n)}$ to be the unique number $2^{-n}k$ in D_n such that $2^{-n}(k-1) < \rho \le 2^{-n}k$. Furthermore, define $\infty^{(n)} = \infty$. Then for the stopping time S, the random times $S^{(n)}(\omega) = (S(\omega))^{(n)}$ are stopping times. Indeed, $S^{(n)}$ are discrete-valued stopping times with respect to the filtration $\{\mathcal{F}_t\}_{t\in D_n}$ and $\lim_n S^{(n)} = S$ a.s.

Now $S \le S^{(n+1)} \le S^{(n)}$ a.s., so $\mathcal{F}_{S^{(n)}} \supseteq \mathcal{F}_{S^{(n+1)}} \supseteq \ldots \supseteq \mathcal{F}_S$, and working with the discrete parameter set D_{n+1}, we see that, as $X_\infty^T = X_T$,

$$X^T_{S^{(n+1)}} \ge E[X^T_{S^{(n)}}|\mathcal{F}_{S^{(n+1)}}] \ge E[X_T|\mathcal{F}_{S^{(n+1)}}] \quad a.s.$$

Hence, for any $A \in \mathcal{F}_S$,

$$0 \le I_A X^T_{S^{(n)}} - I_A E[X_T|\mathcal{F}_{S^{(n)}}].$$

So, by Fatou's inequality,

$$\begin{aligned} 0 &\le E\big[\liminf_n \big\{I_A X^T_{S^{(n)}} - I_A E[X_T|\mathcal{F}_{S^{(n)}}]\big\}\big] \\ &\le \liminf_n \big\{E\big[I_A X^T_{S^{(n)}} - I_A E[X_T|\mathcal{F}_{S^{(n)}}]\big]\big\} \\ &= \liminf_n \big\{E[I_A X^T_{S^{(n)}}]\big\} - E[I_A X_T]. \end{aligned}$$

We know $\{X^T_{S^{(n)}}\}_{n\in\mathbb{N}}$ is a uniformly integrable family (Lemma 5.1.6), $E[I_A X^T_{S^{(n)}}]$ is nondecreasing in n, $S^{(n)} \downarrow S$ a.s. and X_s is right-continuous a.s. Therefore,

$$\liminf_n \{E[I_A X^T_{S^{(n)}}]\} = \lim_n \big\{E[I_A X^T_{S^{(n)}}]\big\} = E\big[\lim_n \{I_A X^T_{S^{(n)}}\}\big] = E[I_A X^T_S].$$

As $S \le T$, X_S is $\mathcal{F}_S$ measurable and $A \in \mathcal{F}_S$ was arbitrary, we see that

$$X_S \ge E[X_T|\mathcal{F}_S].$$

$\square$

Corollary 5.3.2. *If $\{X_t\}_{t\in[0,\infty]}$ is a uniformly integrable right-continuous martingale, and S and T are two stopping times such that $S \le T$ a.s. then $X_S = E[X_T|\mathcal{F}_S]$ a.s.*

Corollary 5.3.3. *If $\{X_t\}_{t\in[0,\infty]}$ is a uniformly integrable right-continuous supermartingale and T is a stopping time, then the stopped process X^T is also a right-continuous supermartingale.*

For supermartingales defined only on the time interval $[0,\infty[$, we also have the following version of Corollary 5.3.3.

Lemma 5.3.4. *If $\{X_t\}_{t\in[0,\infty[}$ is a right-continuous supermartingale and T is a finite valued stopping time, then the stopped process X^T is also a right-continuous supermartingale.*

Proof. For T a stopping time, we know $X_t^T = X_{t\wedge T} = X_T^t$, and that X_t is integrable. Hence, by the optional stopping theorem applied to the stopped process X^t, we see that X_t^T is integrable for every t. The stopped process is also adapted (Theorem 3.2.29). Furthermore, for any $s < t$, as $T \wedge s$ and $T \wedge t$ are bounded stopping times, by the optional stopping theorem and Lemma 3.1.16,

$$\begin{aligned} X_s^T &= X_{T\wedge s} \ge E[X_{T\wedge t}|\mathcal{F}_{T\wedge s}] = I_{\{T<s\}}X_T + I_{\{T\ge s\}}E[X_{T\wedge t}|\mathcal{F}_{T\wedge s}] \\ &= I_{\{T<s\}}X_t^T + I_{\{T\ge s\}}E[X_t^T|\mathcal{F}_s] = E[X_t^T|\mathcal{F}_s]. \end{aligned}$$

Therefore X^T is a supermartingale. □

Clearly, this result also extends to submartingales and martingales.

5.4 Decomposition of Supermartingales

The "Riesz decomposition" for a supermartingale in continuous-time is established next (cf. Theorem 4.6.5).

Definition 5.4.1. *As in the discrete-time case, a nonnegative, right-continuous supermartingale X is called a* potential *if* $\lim_{t\to\infty} E[X_t] = 0$.

Remark 5.4.2. As in the discrete time case (Remark 4.6.4), a right-continuous supermartingale X is a potential if and only if

(i) $X_t(\omega) \ge 0$ a.s.
(ii) $\lim_{t\to\infty} X_t(\omega) = 0$ a.s.

Theorem 5.4.3 (Riesz Decomposition). *Suppose X is a right-continuous, uniformly integrable supermartingale. Then there is a right-continuous, uniformly integrable martingale Y and a potential Z such that $X = Y + Z$ up to indistinguishability. This decomposition is unique (up to indistinguishability).*

Proof. Let Y be a right-continuous modification of the martingale $\{E[X_\infty|\mathcal{F}_t]\}_{t\in\mathbb{T}}$, and put $Z = X - Y$. Then Z is a right-continuous, uniformly integrable supermartingale, and

$$\lim_{t\to\infty} Z_t = \lim_{t\to\infty} (X_t - E[X_\infty|\mathcal{F}_t]) = 0 \quad a.s.$$

Therefore, Z is a potential. If $X = Y' + Z'$, then $Y'_\infty = X_\infty = Y_\infty$, so $Y'_t = E[X_\infty|\mathcal{F}_t] = Y_t$ a.s. for all $t \in \mathbb{T}$. By right-continuity Y' is indistinguishable from Y. □

Theorem 5.4.4. *Suppose X is a càdlàg nonnegative supermartingale adapted to a right-continuous filtration. Write*

$$T(\omega) = \inf\big\{t : (X_t(\omega) = 0) \text{ or } (t > 0 \text{ and } X_{t-}(\omega) = 0)\big\}.$$

Then, for almost every $\omega \in \Omega$, $X_s(\omega) = 0$ on $[T(\omega), \infty[$, that is, $I_{[T(\omega),\infty[}X_t = 0$ up to indistinguishability.

Proof. Consider, for $n \in \mathbb{N}$, the stopping times

$$T_n(\omega) = \inf\{t : X_t(\omega) \le 1/n\}.$$

Clearly $T_{n-1} \le T_n \le T$, and $X_{T_n} \le 1/n$ on $\{T < \infty\}$. The process $\{I_{\{t<\infty\}}X_t\}_{t\in\mathbb{T}}$ is a nonnegative supermartingale, so without loss of generality we can assume that $X_\infty = 0$. By the optional stopping theorem, for any rational $t \ge 0$,

$$E[X_{T+t}] \le E[X_{T_n}] \le 1/n.$$

Taking $n \to \infty$, at least for a subsequence we see $X_{T+t} = 0$ a.s. for each t, and the result follows by right-continuity. □

Remark 5.4.5. The above argument also shows that $\lim_n T_n = T$. Furthermore, if $S(\omega) = \inf\{t : X_t(\omega) = 0\}$, then $X_t(\omega)$ is strictly positive on $[0, S(\omega)[$, and for every random time $R < S$ we know $\inf_{t\le R} X_t(\omega) \ge 0$.

Finally, we give a characterization of a uniformly integrable martingale. Note that we assume M_∞ is already defined.

Theorem 5.4.6. *Suppose $\{M_t\}_{t\in[0,\infty]}$, is a right-continuous process adapted to a right-continuous filtration $\{\mathcal{F}_t\}_{t\in[0,\infty]}$. Then M is a uniformly integrable martingale if, and only if, for every stopping time T we know $E\big[|M_T|\big] < \infty$ and $E[M_T] = E[M_0]$.*

Proof. By considering the process $\{M_t - M_0\}_{t\ge0}$, we can assume without loss of generality that $E[M_T] = E[M_0] = 0$. If M is a u.i. martingale, then $M_t = E[M_\infty|\mathcal{F}_t]$, and the result follows by Theorem 5.3.1 and Jensen's inequality. Conversely, consider any time $t \in [0, \infty]$ and any $A \in \mathcal{F}_t$. Define a random time T by putting $T(\omega) = t$ if $\omega \in A$ and $T(\omega) = \infty$ if $\omega \notin A$. Then T

is a stopping time. (In the notation of Definition 6.2.7 in the next chapter, $T = t_A$.) By hypothesis,

$$E[M_T] = E[I_A M_t] + E[I_{A^c} M_\infty] = 0 = E[M_\infty] = E[I_A M_\infty] + E[I_{A^c} M_\infty].$$

Therefore,

$$E[I_A M_t] = E[I_A M_\infty]$$

for all $A \in \mathcal{F}_t$, so $M_t = E[M_\infty|\mathcal{F}_t]$ almost surely. From Theorem 5.1.8, the martingale $\{E[M_\infty|\mathcal{F}_t]\}_{t\in[0,\infty]}$ has a càdlàg modification and, for this modification, $M_t = E[M_\infty|\mathcal{F}_t]$ up to indistinguishability (Lemma 3.2.10). □

Remark 5.4.7. If the above property only holds for all *bounded* stopping times, then we see that M is a martingale (by considering the stopped process at an increasing sequence of deterministic times), and we do not need to assume M_∞ is well defined.

5.5 Examples of Martingales

We now give two fundamental examples of martingales – Brownian motion (Fig. 5.1) and compensated Poisson processes (Fig. 5.4).

5.5.1 Brownian Motion

We recall that $N(\mu, \Sigma)$ denotes the multivariate normal distribution with mean μ and positive definite covariance Σ, and has probability density given by the Radon–Nikodym derivative (with respect to Lebesgue measure on $\mathbb{R}^d$)

$$\frac{dP}{dx} = \phi_{(\mu,\Sigma)}(x) = (2\pi)^{-d/2} \det(\Sigma)^{-1/2} \exp\Big\{ -\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\Big\}.$$

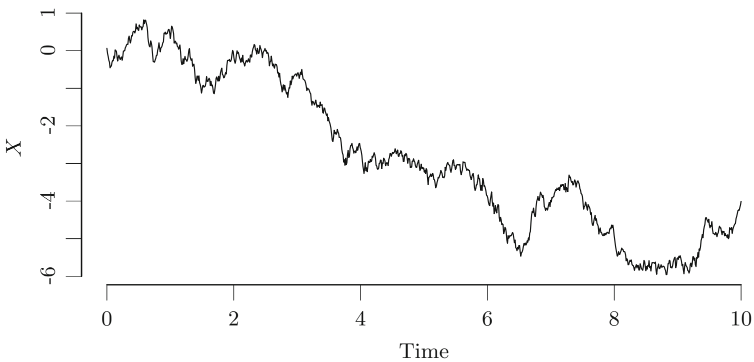

Fig. 5.1. A typical path of a one-dimensional Brownian motion

Definition 5.5.1. *A continuous adapted process X taking values in $\mathbb{R}^d$ will be called a (d-dimensional)* Brownian motion *if, for any increasing sequence $\{t_i\}_{i=1}^N \subset [0,\infty[$, the increments $\{X_{t_{i+1}} - X_{t_i}\}_{i=1}^{N-1}$ are normally distributed*

$$X_{t_{i+1}} - X_{t_i} \sim N(0, (t_{i+1} - t_i)I_d),$$

and $X_{t_{i+1}} - X_{t_i}$ is independent of $\mathcal{F}_{t_i}$. We say that X is a Brownian motion starting at zero if $X_0 = 0$ a.s.

Remark 5.5.2. We shall see in Corollary 5.5.10 that the assumption of continuity of a Brownian motion can be established from the assumptions on the law of the increments, simply by taking a modification of the process concerned.

It is clear, given its independent zero-mean increments, that a Brownian motion is a martingale. We now give a direct construction of a Brownian motion, using the Kolmogorov extension theorem (Theorem A.2.7).

The following lemma establishes some classical and useful properties of normal distributions. Its proof is left as an exercise.

Lemma 5.5.3. *(i) Let Z, Z' be independent random variables with $Z \sim N(\mu, \Sigma)$, $Z' \sim N(\mu', \Sigma')$. Then $Z + Z' \sim N(\mu + \mu', \Sigma + \Sigma')$. Equivalently, their densities satisfy the convolution property*

$$\int_{\mathbb{R}^d} \phi_{(\mu,\Sigma)}(y)\phi_{(\mu',\Sigma')}(x-y)dy = \phi_{(\mu+\mu',\Sigma+\Sigma')}(x).$$

(ii) If $Z_i \sim N(\mu_i, \Sigma_i)$ is a sequence of independent normal random variables such that $\mu^* = \sum_{i\in\mathbb{N}} \mu_i$ and $\Sigma^* = \sum_{i\in\mathbb{N}} \Sigma_i$ exist (i.e. the sums converge), then $\sum_{i\in\mathbb{N}} Z_i$ exists almost surely, and has distribution

$$\sum_{i\in\mathbb{N}} Z_i \sim N(\mu^*, \Sigma^*).$$

(iii) If the pair (Z, Z') is a multivariate normal random variable, then Z and Z' are normal, and are independent if and only if their covariance is zero, that is, $E[(Z-\mu)(Z'-\mu')^\top] = 0$.

Theorem 5.5.4. *Let $(\Omega, \mathcal{F}) = ((\mathbb{R}^d)^{\mathbb{T}}, \mathcal{B}((\mathbb{R}^d)^{\mathbb{T}}))$, that is Ω is the space of d-dimensional real paths $\omega : \mathbb{T} = [0,\infty[\to \mathbb{R}^d$, with $\mathcal{F}$ its Borel σ-algebra (see Definition A.2.1). Then there exists a measure P such that the canonical process $X_t(\omega) := \omega_t$ is (a modification of) a Brownian motion starting at zero in its natural filtration (cf. Definition 3.2.18).*

Proof. This is a direct application of Kolmogorov's extension theorem (Theorem A.2.7). For any ordered set $T = \{t_1 < t_2 < \ldots < t_n\}$, let $t_0 = 0$ and define

$$P_T(B) := \int_B \Big(\prod_{i=1}^n \phi_{(0,(t_i - t_{i-1})I_d)}(x_i)\Big) d(\otimes_{i=1}^n x_i)$$

with the convention, if $t_1 = 0$, that $\phi_{(0,0)}(x) = \delta_{x=0}$, the Dirac delta at zero[1]. By Lemma 5.5.3 and Remark A.2.5, these measures are consistent. They clearly also satisfy "$X_{t_{i+1}} - X_{t_i}$ normally distributed", and "$X_{t_{i+1}} - X_{t_i}$ is independent of $\mathcal{F}_{t_i}$", by an application of Exercise 2.7.12. Hence Theorem A.2.7 yields the existence of a measure P such that $X_t(\omega) = \omega_t$ has the desired law. We shall see in Corollary 5.5.11 that the process constructed has a continuous modification, and hence this modification is a Brownian motion. □

Remark 5.5.5. It is clear that this argument is not unique to constructing Brownian motion, but can be used to construct any process with independent increments, provided we have a convolution property of the densities, similar to that established in Lemma 5.5.3, so as to ensure consistency of the measures.

This existence theorem is elegant, but is not an explicit construction of a Brownian motion – one simply invokes the extension theorem and states that a measure exists such that the canonical paths are a Brownian motion. A method which focusses less on constructing the measure and more on constructing paths with the desired property is given below. This construction also directly shows that the paths of (this construction of) Brownian motion are almost surely continuous.

Lévy's Construction of Brownian Motion

We begin with a countable family $\{Z_m\}$ of identically distributed random variables with $Z_m \sim N(0, I_d)$ for all m. For $n \in \mathbb{Z}^+$, let $D_n = \{k2^{-n} : k \in \mathbb{Z}^+\}$, so that $D_n \subset D_{n+1}$, $D_0 = \mathbb{Z}^+$ and $\cup_n D_n$ is the set of Dyadic rationals. For simplicity of notation, let $\{Z_m\}$ be indexed by $m \in \cup_n D_n$ and $Z_0 := 0$.

We proceed as follows (Fig. 5.2): First, we determine the value of the nth approximation X^n on the points D_n. Second, we use linear interpolation to define X^n_t for all values of t. This gives us a sequence of paths which we shall show converge.

To fix the values of X^n_t for $t \in D_n$, we define

$$X^0_t = \sum_{\{k \in D_0 : k < t\}} Z_k.$$

Next, for every $n > 0$, define $X^n_t = X^{n-1}_t$ for all $t \in D_{n-1}$. For $t \in D_n \setminus D_{n-1}$, let

$$X^n_t = X^{n-1}_t + 2^{-(n+1)} Z_t. \tag{5.1}$$

[1] The Dirac delta function at y, denoted $\delta_{x=y}$, is the 'function' with the defining property that $\int_{\mathbb{R}^d} f(x)\delta_{x=y} dx = f(y)$ for any bounded measurable function y. Hence, $\delta_{x=y}$ can be thought of as an infinitely tall spike at y. While this is not a function, but a linear map $L^\infty \to \mathbb{R}$ (so an element of the dual of L^∞), its integral is still well defined, and this serves as convenient notation.

We now linearly interpolate between these points $\{X_t^n\}_{t\in D_n}$. Formally, we can write the interpolation step as

$$X_t^n = X_{\lfloor t\rfloor_n}^n + \frac{t-\lfloor t\rfloor_n}{\lceil t\rceil_n - \lfloor t\rfloor_n}(X_{\lceil t\rceil_n}^n - X_{\lfloor t\rfloor_n}^n),$$

where $\lfloor t\rfloor_n = \max\{s\in D_n : s\leq t\}$, $\lceil t\rceil_n = \min\{s\in D_n : s\geq t\}$. The use of linear interpolation is not vital to the construction, as we shall see (taking right-continuous step functions $X_t^n := X_{\lfloor t\rfloor_n}^n$ would work just as well for proving the existence of a limit, but would not immediately give continuity). We now seek to show that these paths converge, in a sufficiently strong sense, to a Brownian motion.

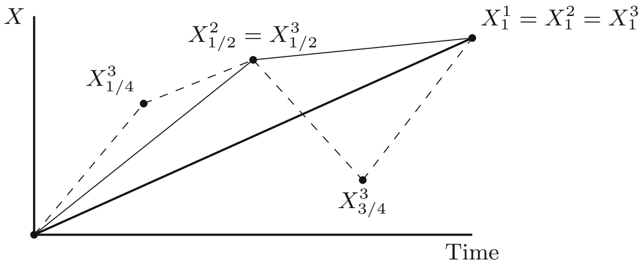

Fig. 5.2. Three steps in Lévy's construction

Lemma 5.5.6. *Let $\{X^n\}_{n\in\mathbb{N}}$ be a sequence of a.s. continuous (resp. càdlàg, càg) functions which converge uniformly on compacts in probability to a process X, that is, for any $\epsilon > 0$,*

$$\lim_n P\Big(\sup_{s\in[0,t]}\|X_s^n - X_s\| < \epsilon\Big) = 1$$

for all t. Then X is also continuous (resp. càdlàg, càg).

Proof. For fixed t, by taking a subsequence in n and using Lemma 1.3.38, we can assume that the convergence is almost sure, that is,

$$P\Big(\lim_n\big(\sup_{s\in[0,t]}\|X_s^n - X_s\|\big) = 0\Big) = 1$$

Fixing ω, this is a statement of uniform convergence of $X^{n_j}\to X$, and the continuity of the limit is classical, as for any $\epsilon>0$, we can find $\delta, m > 0$ such that

$$\begin{aligned}\|X_s - X_{s+\delta}\| &\leq \|X_s^{n_m} - X_s\| + \|X_{s+\delta}^{n_m} - X_{s+\delta}\| + \|X_s^{n_m} - X_{s+\delta}^{n_m}\|\\ &\leq 2\sup_{s\in[0,t]}\{\|X_s^{n_j} - X_s\|\} + \|X_s^{n_m} - X_{s+\delta}^{n_m}\|\\ &\leq 3\epsilon.\end{aligned}$$

□

Remark 5.5.7. The uncountable supremum in the statement of Lemma 5.5.6 is measurable, as our functions are continuous (so the supremum could equally be taken over the rationals, and suprema over countable sets are always measurable).

Theorem 5.5.8. *The processes X^n defined in (5.1) converge a.s. uniformly on compacts to a process X. In its natural filtration, the limit is a Brownian motion starting at zero.*

Proof. Convergence. We first show that the processes converge. We consider the case $d = 2$, as this implies all other cases by the triangle inequality, and is notationally simpler. From our construction, we can see that

$$\sup_{s\in[0,t]} \|X_s^n - X_s^{n+1}\| = \max_{\{s\in D_{n+1}\setminus D_n : s<t\}} \|2^{-(n/2+1)} Z_s\|.$$

The set $\{s \in D_{n+1} \setminus D_n : s < t\}$ contains at most $t2^n$ elements, and the Z_s are independent $N(0, I_d)$ random variables. It is standard that $\|Z_s\|^2$ has a χ^2-distribution with $d = 2$ degrees of freedom, so if $F(x) := P(\|Z_s\|^2 \le x)$ is the distribution function of $\|Z_s\|^2$ we have

$$\begin{aligned} &P\big(\sup_{s\in[0,t]} \|X_s^n - X_s^{n+1}\| > \epsilon\big) = P\big(\max_{\{s\in D_{n+1}\setminus D_n : s<t\}} \|Z_s\| > 2^{n+1}\epsilon\big) \\ &\le \sum_{\{s\in D_{n+1}\setminus D_n, s<t\}} P\big(\|Z_s\| > 2^{n+1}\epsilon\big) = t2^n(1 - F(2^{2n+2}\epsilon^2)). \end{aligned}$$

By changing into polar coordinates, it is easy to show that $F(x) = 1 - e^{-x/2}$ (this simple form is the reason we chose $d = 2$). Therefore,

$$P\big(\sup_{s\in[0,t]} \|X_s^n - X_s^{n+1}\| > \epsilon\big) \le t2^n \exp(-2^{2n+1}\epsilon^2)$$

Taking N large enough that $N \log(2) - 2^{2N+1}\epsilon^2 < -N$, for all $n > N$ we have

$$P\big(\sup_{s\in[0,t]} \|X_s^n - X_s^{n+1}\| > \epsilon\big) \le te^{-n}.$$

By the Borel–Cantelli Lemma (Theorem 2.1.13), as this sequence is summable we have

$$P\big(\sup_{s\in[0,t]} \|X_s^n - X_s^{n+1}\| > \epsilon \text{ for infinitely many } n\big) = 0.$$

Therefore, with probability one, the sequence of processes X^n converges uniformly on the interval $[0, t]$. By Lemma 5.5.6, X is a continuous process.

X is a Brownian motion. We now need to show that X is a Brownian motion in its natural filtration, that is, that the increment $X_t - X_s$ is normally distributed and independent of $\mathcal{F}_s = \sigma(X_u, u \le s)$. First note that for s, t with $t \in D_n \setminus D_{n+1}$ and $\lceil s \rceil_n < t$, the random variable Z_t is not involved in the

construction of X_s. Hence, as X generates the filtration and the $\{Z_u\}_{u\in\cup_n D_n}$ are independent, we see that Z_t is independent of $\mathcal{F}_s$.

It is clear that if s, t are integers with $s < t$, then

$$X_t - X_s = X_t^0 - X_s^0 = \sum_{\{k\in D_0: s<k<t\}} Z_k \sim N(0, (t-s)I_d),$$

by Lemma 5.5.3. Furthermore, in this case $X_t - X_s$ is independent of $\mathcal{F}_s$, as $Z_k = Z_{\lceil k\rceil_0}$ is independent of $\mathcal{F}_s$ for all $s < k$.

Now suppose that the result holds for $s, t \in D_n$. Then we see that for any $u \in D_{n+1} \setminus D_n$,

$$\begin{aligned} X_u - X_{\lfloor u\rfloor_n} &= \frac{X_{\lceil u\rceil_n} + X_{\lfloor u\rfloor_n}}{2} + 2^{-(n+2)} Z_u \\ &= \frac{2^{-(n+1)} Z_{\lceil u\rceil_n}}{2} + 2^{-(n+2)} Z_u \sim N(0, 2^{-(n+1)} I_d), \end{aligned}$$

which is independent of $\mathcal{F}_{\lfloor u\rfloor_n}$. Similarly,

$$X_{\lceil u\rceil_n} - X_u = \frac{2^{-(n+1)} Z_{\lceil u\rceil_n}}{2} - 2^{-(n+2)} Z_u \sim N(0, 2^{-(n+1)} I_d),$$

which is independent of $\mathcal{F}_{\lfloor u\rfloor_n}$. Therefore, for any $s, t \in D_{n+1}$,

$$X_t - X_s = (X_t - X_{\lfloor t\rfloor_n}) + (X_{\lfloor t\rfloor_n} - X_{\lceil s\rceil_n}) + (X_{\lceil s\rceil_n} - X_s),$$

which is the sum of three independent normal random variables, so by Lemma 5.5.3,

$$X_t - X_s \sim N(0, (t-s)I_d).$$

The first two terms are independent of $\mathcal{F}_{\lceil s\rceil_n} \supseteq \mathcal{F}_s$. We know the last term is directly independent of $\mathcal{F}_{\lfloor s\rfloor_n}$, and

$$E\big[(X_{\lceil s\rceil_n} - X_s)(X_s - X_{\lfloor s\rfloor_n})^\top\big] = 2^{-2(n+2)} E\big[(Z_{\lceil s\rceil_n} - Z_s)(Z_{\lceil s\rceil_n} + Z_s)^\top\big] = 0$$

so it is independent of the increment $X_s - X_{\lfloor s\rfloor_n}$, by Lemma 5.5.3(iii). As we can write

$$\mathcal{F}_s = \mathcal{F}_{\lfloor s\rfloor_n} \vee \sigma(X_s - X_{\lfloor s\rfloor_n}) \vee \sigma\big(Z_u; u \in \,]\lfloor s\rfloor_n, s[\,\big),$$

we see that $X_{\lceil s\rceil_n} - X_s$ is independent of $\mathcal{F}_s$. Therefore $X_t - X_s$ is normally distributed and independent of $\mathcal{F}_s$, as desired, for $s, t \in D_{n+1}$. Induction yields the result for $\cup_n D_n$.

Finally, for any $s < t$ we can find sequences $s_n \downarrow s$, $t_n \uparrow t$ with $s_n, t_n \in D_n$ and $s_k \leq t_k$ for some $k \geq 0$. Then $X_{t_n} - X_{s_n} \sim N(0, (t_n - s_n)I_d)$ and, by Lemma 5.5.3(ii) and continuity of X, we see

$$X_t - X_s = X_{t_k} - X_{s_k} + \sum_{n=k+1}^{\infty} \big(X_{t_n} - X_{t_{n-1}} - X_{s_n} + X_{s_{n-1}}\big) \sim N(0, (t-s)I_d).$$

All the terms in this sum are independent of $\mathcal{F}_s$, as required. As $X_0 = 0$ by construction, we see that X is a Brownian motion starting at zero, in its natural filtration. □

Some Properties of Brownian Motion

Given the significance of Brownian motion to stochastic calculus, we now give some of its basic properties. First, we note that we have given two constructions of Brownian motion, one of which has almost surely continuous paths, the other only satisfying the desired law of the increments. We now show that the existence of a continuous modification can be deduced directly from these laws.

Recall that a function $g : \mathbb{R}^+ \to E$, where E is a Banach space, is called locally Hölder γ-continuous if for all T

$$\sup_{s<t<T} \left\{ \frac{\|g(s) - g(t)\|}{|t-s|^\gamma} \right\} < \infty.$$

If g is locally Hölder continuous, it is clearly continuous.

Theorem 5.5.9 (Kolmogorov–Čentsov Theorem). *Let X be a Banach space valued measurable process such that, for some positive α, β, c, for all $s < t$,*

$$E\big[\|X_t - X_s\|^\alpha\big] \leq c|t-s|^{1+\beta}.$$

Then there exists a modification $\tilde{X}$ of X which is almost surely locally Hölder γ-continuous for all $\gamma \in]0, \beta/\alpha[$. In particular, for each T, there exists a constant $k > 0$ such that for all $\delta > 0$,

$$P\Big(\sup_{\{s<t<T\}} \left\{ \frac{\|\tilde{X}_t - \tilde{X}_s\|}{|t-s|^\gamma} \right\} > \delta \Big) \leq k\delta^{-\alpha}.$$

Note that this supremum is measurable as $\tilde{X}$ is continuous, so the statement is meaningful.

Proof. See Appendix A.4. □

Corollary 5.5.10. *A Brownian motion is almost surely locally γ-Hölder continuous for all $\gamma < 1/2$.*

Proof. As the increments of Brownian motion are normal random variables, for every $s < t$ we can write $X_t - X_s = |t-s|^{1/2} Z$ for some $Z \sim N(0, I_d)$. Hence, for any $\alpha \in \mathbb{R}^+$,

$$E[\|X_t - X_s\|^\alpha] = |t-s|^{\alpha/2} E[\|Z\|^\alpha].$$

As $E[\|Z\|^\alpha] < \infty$ and X is continuous, we see that X is γ-Hölder continuous for all $\gamma < (\alpha/2 - 1)/\alpha = 1/2 - 1/\alpha$. As α can be arbitrarily large, we have the result. □

Corollary 5.5.11. *The process constructed in Theorem 5.5.4 admits a continuous modification.*

Proof. This follows from the law of the process, as in Corollary 5.5.10. Note that, as the continuous modification of a process is unique up to indistinguishability (Lemma 3.2.10), this process will also be locally Hölder γ-continuous for all $\gamma < 1/2$.

Remark 5.5.12. Let $\Omega = C_d([0,\infty[)$ be the space of continuous functions $\omega : [0,\infty[\to \mathbb{R}$. Let $\mathcal{F}$ denote the σ-algebra generated by the canonical process $X_t(\omega) = \omega_t$, that is, $\mathcal{F}_s = \sigma(X_s : s < \infty)$, so that the canonical process $\{X_t\}_{t\in[0,\infty[}$ is measurable. By continuity, this is the same as the σ-algebra $\{A \cap \Omega : A \in \mathcal{B}(\mathbb{R}^{\mathbb{R}})\}$, with $\mathcal{B}(\mathbb{R}^{\mathbb{R}})$ as defined in Definition A.2.1. Given Corollary 5.5.10, after removing the null set on which the continuous limit does not exist, Theorem 5.5.4 gives a measure on the space $(\Omega, \mathcal{F})$ such that $X_t = \omega_t$ is a Brownian motion under its natural filtration. This particular setting is called the (classical) Wiener space, and the measure is called Wiener measure. In light of Example 3.4.15, the natural filtration given by $\mathcal{F}_t = \sigma(X_s : s < t)$ is not right continuous. However, one can show that this is remedied by simply completing the space (see Theorem 17.3.8).

One might ask whether paths of Brownian motion are Hölder γ-continuous for $\gamma \geq 1/2$. This is not the case, as can be shown using the following theorem which we state without proof (a proof can be found in [155, p.30]).

Theorem 5.5.13 (Lévy's Modulus of Continuity). *For a one-dimensional Brownian motion X,*

$$\limsup_{\epsilon\to 0} \left\{ \sup_{\{s<t<1:|t-s|<\epsilon\}} \frac{\|X_t - X_s\|}{(2\epsilon \log(1/\epsilon))^{1/2}} \right\} = 1 \qquad a.s.$$

A final inequality for Brownian motion is the 'law of the iterated logarithm'. This describes the behaviour of a Brownian motion near time $t = 0$.

Theorem 5.5.14 (Law of the Iterated Logarithm). *For a X one-dimensional Brownian motion,*

$$\limsup_{t\to 0} \frac{X_t}{(2t \log\log(1/t))^{1/2}} = 1 \qquad a.s.$$

Proof. By Exercise 5.7.6, we know that for any $\alpha \in \mathbb{R}$, $\exp(\alpha X_t - \alpha^2 t/2)$ is a nonnegative martingale with expected value 1. Applying Doob's maximal inequality (Theorem 5.1.2(i)),

$$P\Big(\max_{s\in[0,t]} \Big\{ X_s - \frac{\alpha s}{2} \Big\} > \lambda \Big) = P\Big(\max_{s\in[0,t]} \Big\{ \exp\Big(\alpha X_s - \frac{\alpha^2 s}{2}\Big) \Big\} > e^{\alpha\lambda} \Big) \leq e^{-\alpha\lambda},$$

Write $h(x) = (2x \log\log(1/x))^{1/2}$. Taking fixed $\theta, \delta \in]0,1[$, for any $n > 1$ we can set $\alpha_n = (1+\delta)\theta^{-n} h(\theta^n)$, $\lambda_n = h(\theta^n)/2$ and $t_n = \theta^n$, and so obtain

$$P\Big(\max_{s\in[0,t_n]} \Big\{ X_s - \frac{\alpha_n s}{2} \Big\} > \lambda_n \Big) \leq (n \log(1/\theta))^{-(1+\delta)}.$$

By the Borel–Cantelli lemma (Theorem 2.1.13(i)), as $\sum_{n\in\mathbb{N}} n^{-(1+\delta)} < \infty$,

$$P\Big(\liminf_{n\to\infty}\Big\{\max_{s\in[0,t_n]}\Big\{X_s - \frac{\alpha_n s}{2}\Big\}\Big\} > \lambda_n\Big) = 0.$$

Therefore, there exists an almost surely finite random variable N such that, whenever $n > N$ and $s \in [0, \theta^n]$,

$$X_s < \frac{\alpha_n s}{2} + \lambda_n = \frac{s}{2}(1+\delta)\theta^{-n}h(\theta^n) + \frac{h(\theta^n)}{2} \leq \Big(1 + \frac{\delta}{2}\Big)h(\theta^n).$$

As h is increasing near zero, this implies that for n sufficiently large, $m > 0$ and $t \in [\theta^{n+m}, \theta^n]$,

$$\frac{X_t}{h(t)} < \Big(1 + \frac{\delta}{2}\Big)\frac{h(\theta^n)}{h(\theta^{n+m})} < \Big(1 + \frac{\delta}{2}\Big)\theta^{-m/2}.$$

Taking $\theta \to 1$ and $\delta \to 0$, we see

$$\limsup_{t\to 0}\frac{X_t}{h(t)} \leq \Big(1 + \frac{\delta}{2}\Big)\theta^{-m/2} \to 1.$$

To show the reverse inequality, for any $\theta \in]0, 1[$ we define the event

$$A_n = \Big\{X_{\theta^n} - X_{\theta^{n+1}} \geq h(\theta^n)\sqrt{1-\theta}\Big\}.$$

As $X_{\theta^n} - X_{\theta^{n+1}} \sim N(0, \theta^n - \theta^{n+1})$, we can calculate

$$P(A_n) = (2\pi)^{-1/2}\int_y^\infty e^{-u^2/2}du$$

where

$$y = \frac{h(\theta^n)\sqrt{1-\theta}}{\sqrt{\theta^n - \theta^{n+1}}} = \sqrt{2\log\log(\theta^{-n})}.$$

So, as $\int_y^\infty e^{-u^2/2}du \geq \frac{y}{1+y^2}e^{-y^2/2}$, we know

$$P(A_n) \geq K(n^2 \log n)^{-1/2}$$

for some constant $K > 0$. However, this implies that $\sum_{n\in\mathbb{N}} P(A_n) = \infty$, and by independence and the second part of the Borel–Cantelli lemma (Theorem 2.1.13(ii)) we have $P(\cap_k \cup_{n\geq k} A_n) = 1$, that is,

$$X_{\theta^n} \geq X_{\theta^{n+1}} + h(\theta^n)\sqrt{1-\theta} \quad \text{for infinitely many } n\text{, a.s.}$$

By Exercise 5.7.5, $-X$ is also a Brownian motion, so by the first result, $X_{\theta^{n+1}} > -2h(\theta^{n+1})$ for all n sufficiently large. Therefore,

$$\frac{X_{\theta^n}}{h(\theta^n)} \geq -2\frac{h(\theta^{n+1})}{h(\theta^n)} + \sqrt{1-\theta}$$
$$= -2\sqrt{\theta\,\frac{\log(n+1)+\log\log(1/\theta)}{\log(n)+\log\log(1/\theta)}}+\sqrt{1-\theta} \quad \text{for infinitely many } n, \text{ a.s.}$$

Therefore, $\limsup_{n\to\infty} X_{\theta^n}/h(\theta^n) \geq -2\sqrt{\theta}+\sqrt{1-\theta}$. Finally, we note that

$$\limsup_{t\downarrow 0} \frac{X_t}{h(t)} \geq \limsup_{\theta\to 0}\limsup_{n\to\infty} \frac{X_{\theta^n}}{h(\theta^n)} \geq 1 \qquad \text{a.s.}$$

□

Remark 5.5.15. By Exercise 5.7.5, as $-X$ is also a Brownian motion we obtain the related bound

$$\liminf_{t\to 0} \frac{X_t}{(2t\log\log(1/t))^{1/2}} = -1$$

and by Exercise 5.7.7, we obtain (Fig. 5.3)

$$\limsup_{t\to\infty} \frac{X_t}{(2t\log\log(t))^{1/2}} = 1, \qquad \liminf_{t\to\infty} \frac{X_t}{(2t\log\log(t))^{1/2}} = -1. \tag{5.2}$$

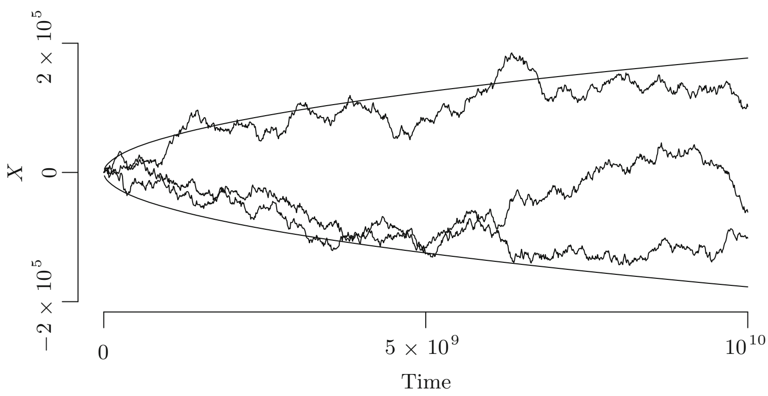

Fig. 5.3. Three paths of a Brownian motion over a long horizon, along with the bounds implied by (5.2).

5.5.2 Poisson Process Martingales

More briefly, we now present another class of martingales, based on processes with jumps.

Definition 5.5.16 (Poisson Process). *A process N will be called a* counting process *if it is nondecreasing, adapted, càdlàg and takes values in the integers $\mathbb{Z}^+$.*

An integrable counting process N will be called a (one-dimensional) Poisson process *if, for any $t, \delta > 0$,*

(i) the initial value is $N_0 = 0$,
(ii) N increases by at most one at every point, i.e. $N_t - N_{t-} \in \{0, 1\}$ up to indistinguishability,
(iii) $N_{t+\delta} - N_t$ is independent of $\mathcal{F}_t$, and
(iv) $N_{t+\delta} - N_t$ has a distribution which does not depend on t.

An explicit construction of such a process will be given in Theorem 5.5.22.

Lemma 5.5.17. *If N is a Poisson process, then for some $\lambda \geq 0$,*

$$E[N_{t+\delta} - N_t] = E[N_{t+\delta} - N_t|\mathcal{F}_t] = \lambda\delta$$

for all t.

Proof. By integrability and properties (iii) and (iv) of the Poisson process, there exists a function g such that

$$E[N_{t+\delta} - N_t] = E[N_{t+\delta} - N_t|\mathcal{F}_t] = g(\delta) \qquad a.s.$$

Furthermore, for any $\epsilon > 0$ we also have

$$\begin{aligned} g(\delta + \epsilon) &= E[N_{t+\delta+\epsilon} - N_t] = E[N_{t+\delta+\epsilon} - N_{t+\delta} + N_{t+\delta} - N_t] \\ &= E[N_{t+\delta+\epsilon} - N_{t+\delta}] + E[N_{t+\delta} - N_t|\mathcal{F}_t] = g(\epsilon) + g(\delta). \end{aligned}$$

Therefore, by a straightforward argument, for some λ we have $g(\delta) = \lambda\delta$, for all rational δ. As N is an nondecreasing process, g is nondecreasing, and so $\lambda \geq 0$ and $g(\delta) = \lambda\delta$ for all $\delta \in \mathbb{R}$. □

Theorem 5.5.18. *The process $\tilde{N}$ defined by $\tilde{N}_t = N_t - \lambda t$ is a martingale, and is called the* compensated Poisson Process*.*

Proof. By the above lemma, for any $t, \delta > 0$,

$$E[N_{t+\delta} - \lambda(t + \delta)|\mathcal{F}_t] = N_t + \delta\lambda - \lambda(t + \delta) = N_t - \lambda t.$$

The integrability condition is also guaranteed, so $N_t - \lambda t$ is a martingale. □

Remark 5.5.19. We call λ the parameter of the Poisson process. We shall see later that λt is the 'compensator' of the increasing process N (Example 8.2.17) and that the existence of such a compensator is generally true for integrable increasing processes. This gives us a large class of martingales to work with, as the jumps of $\tilde{N}$ are the same as those of N (Fig. 5.4).

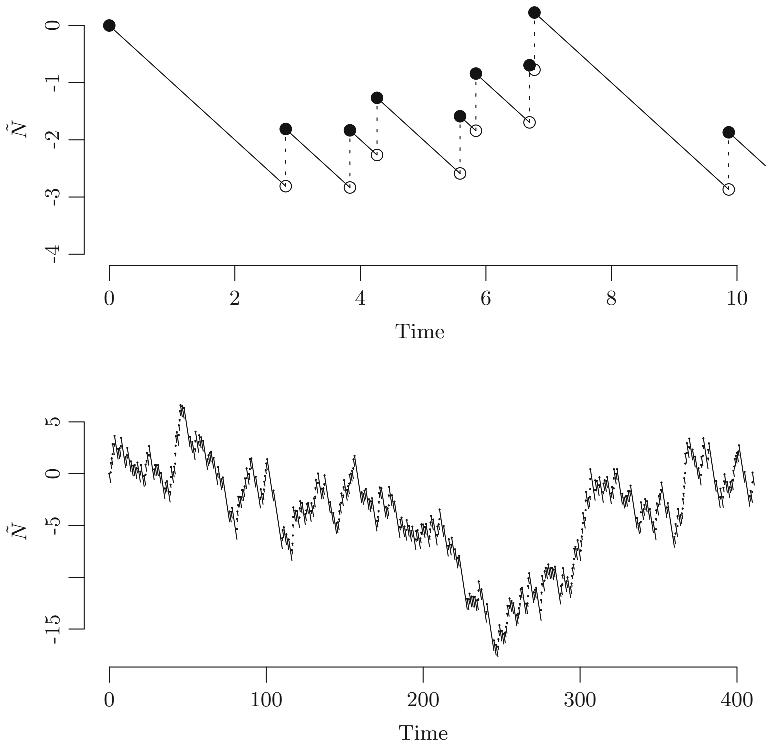

Fig. 5.4. Typical paths of the compensated Poisson process $\tilde{N}$ with $\lambda = 1$ (over different periods).

Some other useful properties of Poisson processes follow.

Lemma 5.5.20. *For N a Poisson process, $S \leq T$ stopping times,*

$$\lim_{\delta \to 0} \frac{P(N_{T+\delta} > N_T | \mathcal{F}_S)}{\lambda \delta} = 1 \quad a.s.,$$

or equivalently, $P(N_{T+\delta} > N_T | \mathcal{F}_S) = \lambda\delta + o(\delta)$ a.s.

Proof. By the optional stopping theorem applied to the martingale $\{N_t - N_{T \wedge t} - \lambda(t - t \wedge T)\}_{t \geq 0}$,

$$E[N_{T+\delta} - N_T | \mathcal{F}_S] = \lambda\delta.$$

Therefore,

$$\begin{aligned} P(N_{T+\delta} > N_T | \mathcal{F}_S) &= P(N_{T+\delta} = N_T + 1 | \mathcal{F}_S) + P(N_{T+\delta} > N_T + 1 | \mathcal{F}_S) \\ &= E[I_{\{N_{T+\delta} = N_T + 1\}} | \mathcal{F}_S] + E[I_{\{N_{T+\delta} > N_T + 1\}} | \mathcal{F}_S] \\ &= E[N_{T+\delta} - N_T | \mathcal{F}_S] \\ &\qquad - E[(N_{T+\delta} - N_T - 1) I_{\{N_{T+\delta} > N_T + 1\}} | \mathcal{F}_S] \\ &= \lambda\delta - E[(N_{T+\delta} - N_T - 1)^+ | \mathcal{F}_S], \end{aligned}$$

and hence

$$P(N_{T+\delta} > N_T|\mathcal{F}_S) \geq \lambda\delta.$$

By Fatou's inequality (Lemma 2.4.6) and property (ii) of the Poisson process,

$$\liminf_{\delta\to 0} \frac{E[(N_{T+\delta} - N_T - 1)^+|\mathcal{F}_S]}{\lambda\delta} = 0 \quad \text{a.s.}$$

Hence

$$1 - \limsup_{\delta\to 0} \frac{P(N_{T+\delta} > N_T|\mathcal{F}_S)}{\lambda\delta} = 0 \quad \text{a.s.},$$

and rearrangement yields the result. □

Theorem 5.5.21. *The times between the jumps of a Poisson process are independent and exponentially distributed with parameter λ.*

Proof. Let S be a stopping time and $T = \inf\{t > 0 : N_{t+S} - N_S > 0\}$, the time from S to the first jump of N following S. For any $t \geq s \geq 0$, we know that

$$\{T > t\} = \{T > t\} \cap \{T > s\} = \{N_{S+t} = N_{S+s}\} \cap \{T > s\}.$$

Write $G(t) = P(T > t|\mathcal{F}_S)$, so by Lemma 5.5.20,

$$\begin{aligned} G(t) &= E[I_{\{T>t\}}|\mathcal{F}_S] = E\big[E[I_{\{T>t\}}|\mathcal{F}_{S+s}]I_{\{T>s\}}\big|\mathcal{F}_S\big] \\ &= E\big[E[I_{\{N_{S+t}=N_{S+s}\}}|\mathcal{F}_{S+s}]I_{\{T>s\}}\big|\mathcal{F}_S\big] \\ &= E[(1 + \lambda(t-s) + o(t-s))I_{\{T>s\}}|\mathcal{F}_S]. \end{aligned}$$

Differentiating in t, and exchanging the order of expectation and differentiation by the conditional dominated convergence theorem,

$$\frac{d}{dt}G(t)|_{t=s} = \lambda E[I_{T>s}|\mathcal{F}_S] = \lambda G(s) \quad \text{a.s.}$$

We have the differential equation $G'(s) = \lambda G(s)$ with initial condition $G(0) = 1$, which has unique solution $G(s) = e^{-\lambda s}$. That is, the law of the waiting time to the first jump of N following S is exponential, given $\mathcal{F}_S$.

Therefore, the time of the first jump of N is exponential, and by induction the time from one jump to the next is conditionally exponential and independent of the past. Therefore, the times between jumps are independent and exponential. □

We also have a simple construction of Poisson processes.

Theorem 5.5.22. *Suppose we have a family $\{Z_n\}_{n\in\mathbb{N}}$ of iid exponentially distributed random variables with parameter λ. Let $X_n = \sum_{k\leq n} Z_n$, and $N_t = \max\{n \in \mathbb{Z}^+ : X_n \leq t\}$. Then N is a Poisson process.*

Proof. Clearly N is a counting process, starts at zero and increases by at most one at every point. It is easy to check that $X_n = \inf\{t : N_t \geq n\}$. Therefore, there is a bijection between the paths of X and N, and so it is enough to prove that X has the desired law (as the law of X implies the law of N). The law of X is determined by the law of $\{Z_n\}_{n\in N}$, which is as uniquely determined by Theorem 5.5.22. The result follows. □

Theorem 5.5.23. *A Poisson process is a strong Markov process, that is, for any stopping time T, the process*

$$N'_t = N_{T+t} - N_T$$

is a Poisson process on the filtration defined by $\tilde{\mathcal{F}}_t = \mathcal{F}_{T+t}$.

Proof. We have seen in the proof of Theorem 5.5.21 that if T is any stopping time and S the first jump of N after T, then $S - T$ has an exponential distribution with rate parameter λ independent of $\mathcal{F}_S$. Therefore $N'_t = N_{S+t} - N_S$ is another process with exponentially distributed times between jumps, with the same rate parameter λ. By Theorem 5.5.22, N' is a Poisson process. □

5.6 Local Martingales

Equipped with these examples, we now consider those processes which are *local* martingales, that is, where there exists a sequence of stopping times $\{T_n\}_{n\in\mathbb{N}}$ such that $T_n \to \infty$ a.s. and the stopped process X^{T_n} is a martingale. These processes are fundamental to understanding the general theory of stochastic processes.

Example 5.6.1. Consider a filtration $\{\mathcal{F}_t\}_{t\geq 0}$ such that there is an $\mathcal{F}_1$-measurable nonnegative random variable ξ with infinite expectation, and also a random variable η taking values ± 1 with equal probability, which is $\mathcal{F}_2$-measurable and independent of $\mathcal{F}_{2-}$. Then the process $X_t = I_{\{t\geq 2\}}\eta\xi$ is a *local* martingale, but is not a martingale.

To see this, consider the sequence of stopping times

$$T_n = \begin{cases} n & \text{if } \xi \leq n, \\ 1 & \text{if } \xi > n. \end{cases}$$

As ξ is finite valued, $T_n \uparrow \infty$ a.s., and it is easy to check that X^{T_n} is a (bounded) martingale for every n. However, as $E\big[|X_2|\big] = E\big[|\xi|\big] = \infty$, X is *not* a martingale.

When X is a uniformly integrable martingale, the set of random variables $\{X_T\}_{T\in\mathcal{T}}$, for $\mathcal{T}$ the set of stopping times, is uniformly integrable.

This follows from Doob's optional stopping theorem (Theorem 5.3.1) because $X_T = E[X_\infty|\mathcal{F}_T]$ a.s. However, this is not true in general, even when X is a uniformly integrable supermartingale or local martingale. (An example of this is below, Example 5.6.9.)

Definition 5.6.2. *A right-continuous uniformly integrable supermartingale X is said to be of* class (D) *if the set of random variables $\{X_T\}_{T\in\mathcal{T}}$ is uniformly integrable (where $\mathcal{T}$ is the set of all stopping times).*

The term 'class (D)' is in reference to J.L. Doob, who developed many of the results for this class of processes.

Definition 5.6.3. *$\mathcal{M}$ will denote the set of càdlàg uniformly integrable martingales on $(\Omega, \mathcal{F}, P)$ with respect to the filtration $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$. That is, if $M \in \mathcal{M}$, then M is a martingale and the set of random variables $\{M_t\}_{t\in\mathbb{T}}$, is uniformly integrable.*

As in Section 3.3, $\mathcal{M}_{\text{loc}}$ will denote the set of processes which are locally in $\mathcal{M}$. If $\mathcal{C}$ is any class of processes, $\mathcal{C}_0$ will denote the set of $X \in \mathcal{C}$ such that $X_0 = 0$ a.s. Therefore, $\mathcal{M}_0$ will denote the set of martingales $M \in \mathcal{M}$ such that $M_0 = 0$ a.s. For simplicity, we write $\mathcal{M}_{0,\text{loc}} = (\mathcal{M}_0)_{\text{loc}} = (\mathcal{M}_{\text{loc}})_0$.

Note that $M_{\infty-}$ is always well defined for $M \in \mathcal{M}$, and if $\mathbb{T} = [0, \infty[$, we write $M_\infty := M_{\infty-}$ for notational simplicity.

Lemma 5.6.4. *Every càdlàg local martingale is locally uniformly integrable (that is, all càdlàg local martingales are in $\mathcal{M}_{\text{loc}}$).*

Proof. From the definition of a local martingale, we have an nondecreasing sequence $\{T_n\}_{n\in\mathbb{N}}$ of stopping times, with $T_n \to \infty$ a.s., such that the stopped processes M^{T_n} are martingales for every n. We can always replace the sequence $\{T_n\}_{n\in\mathbb{N}}$ by the sequence $\{T_n \wedge n\}_{n\in\mathbb{N}}$, so without loss of generality, each T_n is bounded by n. As M^n is a (uniformly integrable) martingale, by Doob's optional stopping theorem (Theorem 5.3.1) we know $M_{T_n} = M^n_{T_n} = E[M_n|\mathcal{F}_{T_n}]$ is integrable. Again by the optional stopping theorem, $M_t^{T_n} = E[M_{T_n}|\mathcal{F}_t]$. Therefore $\{M_t^{T_n}\}_{t\in[0,\infty]}$ is uniformly integrable. □

Lemma 5.6.5. *Every martingale is a local martingale.*

Proof. Suppose M is a martingale. For each positive integer $n \in \mathbb{N}$, consider the stopped value $M^n_s = E[M_n|\mathcal{F}_s]$. Clearly $M^n \in \mathcal{M}$ and we see $M \in \mathcal{M}_{\text{loc}}$. □

Lemma 5.6.6. *A càdlàg local martingale is in $\mathcal{M}$ if and only if it is of class (D).*

Proof. If $M \in \mathcal{M}$, then, from the optional stopping theorem, for every stopping time T,

$$M_T = E[M_\infty|\mathcal{F}_T].$$

By Theorem 2.5.10, the set of random variables $\{M_T\}_{T\in\mathcal{T}}$ is uniformly integrable, and so M is of class (D).

Conversely, suppose $M \in \mathcal{M}_{\rm loc}$ is of class (D). Then M is càdlàg and the set of random variables $\{M_t\}_{t\geq 0}$ is uniformly integrable. To show that $M \in \mathcal{M}$ we must establish that $M_s = E[M_t|\mathcal{F}_s]$ whenever $s \leq t < \infty$.

Let $\{T_n\}_{n\in\mathbb{N}}$ be a localizing sequence for M in $\mathcal{M}$, that is, an increasing sequence of stopping times such that $M^{T_n} \in \mathcal{M}$ for each n. Then, for each n, $M_s^{T_n} = E[M_t^{T_n}|\mathcal{F}_s]$. However, because M is of class (D) and $M_s^{T_n} = M_{T_n}^s$, the sequences $\{M_s^{T_n}\}_{n\in\mathbb{N}}$ and $\{M_t^{T_n}\}_{n\in\mathbb{N}}$ converge, both almost surely and in L^1, to M_s and M_t respectively. The result follows from the result of Exercise 2.7.8. $\square$

Remark 5.6.7. In a similar way, one can show that a local supermartingale of class (D) is a true supermartingale (Exercise 5.7.14).

The following result can naturally be extended to all supermartingales bounded below.

Lemma 5.6.8. *A nonnegative local supermartingale M is a supermartingale, in particular, a nonnegative local martingale is a supermartingale.*

Proof. Let $\{T_n\}_{n\in\mathbb{N}}$ be a localizing sequence for our local martingale M. We know $E[M_0] = E[M_0^{T_n}] < \infty$. Then, by the supermartingale property for M^{T_n} and Fatou's inequality,

$$M_s = \lim_n M_s^{T_n} \geq \lim_n E[M_t^{T_n}|\mathcal{F}_s] \geq E[\lim_n M_t^{T_n}|\mathcal{F}_s] = E[M_t|\mathcal{F}_s].$$

By nonnegativity, this also shows $E\big[|M_t|\big] = E[M_t] \leq E[M_0] < \infty$ for all t, so M is a supermartingale. $\square$

We have already seen one example of a local martingale which is not a martingale (Example 5.6.1). We now give an example of a continuous process which is a nonnegative local martingale (and hence a supermartingale), but is not of class (D).

Example 5.6.9. Let W be a one-dimensional Brownian motion and define $T = \inf\{t : W_t = -1\}$. Then let

$$X_t := \begin{cases} 1 + W_{\min\{\frac{t}{1-t}, T\}} & \text{for } t < 1, \\ 0 & \text{for } t \geq 1. \end{cases}$$

Then X is a local martingale in the filtration $\tilde{\mathcal{F}}_t := \mathcal{F}_{t/(1-t)}$, but is not a martingale. To see this, first note that, as a consequence of the law of the iterated logarithm (Remark 5.5.15, Exercise 5.7.12), T is almost surely finite. Therefore, our process X is well defined for all times and has almost surely continuous paths. Note that X is clearly not a martingale, as $E[X_1] = 0 \neq 1 = E[X_0]$.

Let

$$S_n = \Big(\frac{n}{n+1}\Big) I_{\{T\geq n\}} + \Big(\frac{T}{T+1} + n\Big) I_{\{T<n\}}.$$

One can verify that S_n is an $\{\tilde{\mathcal{F}}_t\}_{t\in[0,\infty]}$-stopping time and, as T is almost surely finite, $S_n \to \infty$ almost surely. Furthermore, for every n we have

$$X_t^{S_n} = 1 + W_{t/(1-t)}^{n\wedge T}.$$

By Exercise 5.7.11 and Lemma 5.3.4, we see that X^{S_n} is a martingale. Therefore, X is a local martingale, but not a martingale, and so is not of class (D).

See Exercise 14.7.11 for another classic example, based on a three-dimensional Brownian motion. Note that (as this example will show) even if a local martingale is uniformly integrable (that is, the set $\{X_t\}_{t\in[0,\infty[}$ is uniformly integrable) or if $\sup_t E[X_t^2] < \infty$, it is not guaranteed to be of class (D). Conversely, if $E[\sup_t X_t^2] < \infty$, then X is of class (D), by Example 2.5.3.

Definition 5.6.10. *Suppose M is an adapted process. A stopping time T is said to* reduce *M if M^T is a uniformly integrable martingale.*

From Lemma 5.6.6, note that if T reduces M then M^T is of class (D). Note also that $M \in \mathcal{M}_{\mathrm{loc}}$ if there is a sequence $\{T_n\}_{n\in\mathbb{N}}$ of stopping times such that $\lim T_n = \infty$ a.s. and each T_n reduces M.

Lemma 5.6.11. *(i) Let M be a local martingale. If the stopping time T reduces M and S is a stopping time such that $S \leq T$ then S reduces M.*
(ii) The sum of two local martingales is a local martingale.

Proof. Doob's optional stopping theorem (Theorem 5.3.1) immediately gives (i). To show (ii), if M and N are local martingales, then suppose $\{S_n\}_{n\in\mathbb{N}}$ is a sequence of stopping times reducing M with $S_n \uparrow \infty$ a.s., and $\{T_n\}_{n\in\mathbb{N}}$ is a sequence of stopping times reducing N with $T_n \uparrow \infty$ a.s. Then, by (i), $S_n \wedge T_n$ is a sequence of stopping times reducing $M+N$ and we see that $S_n \wedge T_n \to \infty$ a.s. Hence $M+N$ is a local martingale. □

Lemma 5.6.12. *(i) If $M \in \mathcal{M}_{0,\mathrm{loc}}$ and S, T are stopping times which reduce M, then $S \vee T$ reduces M.*
(ii) Suppose M is an arbitrary process and there is a nondecreasing sequence of stopping times T_n such that $\lim_n T_n = \infty$ and each M^{T_n} is a local martingale. Then M is a local martingale.

Proof. (i) The process $M^{S\vee T} = M^S + M^T - M^{S\wedge T}$ is a uniformly integrable martingale, by Lemma 5.6.11(i).
(ii) Suppose that for each n, $\{R_{n,m}\}_{m\in\mathbb{N}}$ is a sequence of stopping times reducing M^{T_n} with $\lim_m R_{n,m} = \infty$. Write $S_{n,m} = R_{n,m} \wedge T_n$, so $\lim_m S_{n,m} = T_n$. Index the stopping times $S_{n,m}$ in a single sequence

$\{S_k\}_{k \in \mathbb{N}}$ and write $V_k = S_1 \vee S_2 \vee \cdots \vee S_k$, so $\lim_k V_k = \infty$. We show the V_k reduce M.

Suppose $S_1 = S_{n_1 m_1}, \cdots, S_k = S_{n_k m_k}$ and write $r = n_1 \vee \cdots \vee n_k$. Now $S_{n_i m_i}$ reduces $M^{T_{n_i}}$, but $M^{T_{n_i}}$ and M^{T_r} are the same process up to the stopping time $S_{n_i m_i}$. Therefore, $S_{n_i m_i}$ reduces M^{T_r}. By part (i), we see that V_k reduces M^{T_r}. However, M and M^{T_r} are the same process up to time V_k. Therefore, V_k reduces M and $M \in \mathcal{M}_{\text{loc}}$. □

Theorem 5.6.13. *Suppose M is a right-continuous uniformly integrable martingale and $\{T_n\}_{n \in \mathbb{N}}$ is an increasing sequence of stopping times. Then*

$$\lim_n M_{T_n} = E\Big[M_{(\lim_n T_n)}\Big| \bigvee_n \mathcal{F}_{T_n}\Big]$$

almost surely and in L^1.

Proof. Considering the discrete time martingale M_{T_n}, this follows from Corollary 4.4.5 and Theorem 4.6.7. □

A useful version of this result can be expressed in terms of *predictable* stopping times, which are the subject of the next chapter (see Theorem 6.2.18).

5.7 Exercises

Exercise 5.7.1. Consider a probability space $(\Omega, \mathcal{F}, P)$ with filtration $\{\mathcal{F}_t\}_{t \geq 0}$.

(i) Let Q be a probability measure absolutely continuous with respect to P. Define $\Lambda_t = E[dQ/dP|\mathcal{F}_t]$. Show that Λ is a uniformly integrable nonnegative martingale.

(ii) Conversely, let Λ be a uniformly integrable nonnegative martingale with $\Lambda_0 = 1$. Show that a new probability measure Q can be defined by $dQ/dP = \Lambda_\infty$.

(iii) Let Q and Λ be as in parts (i) and (ii). For a given t, show that, on the measurable space $(\Omega, \mathcal{F}_t)$, the measure Q agrees with the measure R given by $dR/dP = \Lambda_t$.

(iv) Let Q and Λ be as above. For any t and any Q-integrable random variable X, show that the conditional expectation satisfies a version of *Bayes' rule*:

$$E^Q[X|\mathcal{F}_t] = \frac{E^P[X\Lambda_\infty|\mathcal{F}_t]}{\Lambda_t}.$$

(Hint: X is Q-integrable if and only if $X\Lambda$ is P-integrable)

Exercise 5.7.2. Let M be a martingale in a probability space with filtration $\{\mathcal{F}_t\}_{t \geq 0}$. Let $\{\mathcal{G}_t\}_{t \geq 0}$ be a right-continuous and complete filtration on the same space, with $\mathcal{G}_t \subseteq \mathcal{F}_t$ for all t, and such that M is $\{\mathcal{G}_t\}_{t \geq 0}$-adapted. Show that M is a $\{\mathcal{G}_t\}_{t \geq 0}$-martingale.

Exercise 5.7.3. Prove the basic convolution properties of normal distributions, as stated in Lemma 5.5.3.

Exercise 5.7.4. Let X be a continuous process in $\mathbb{R}^d$ such that $X_0 = 0$ and, for every $s < t$, (X_s, X_t) is jointly normally distributed, $E[X_t] = 0$ and $E[X_s X_t^\top] = sI_d$. Show that X is a Brownian motion.

Exercise 5.7.5. Let X be a Brownian motion in a probability space with filtration $\{\mathcal{F}_t\}_{t\geq 0}$. Show that for any $c \in \mathbb{R}$, the process $\tilde{X}_t = cX_{t/c^2}$ is a Brownian motion under the filtration $\{\tilde{\mathcal{F}}_t = \mathcal{F}_{c^2 t}\}_{t\geq 0}$. (In particular, $-X$ is a Brownian motion under $\{\mathcal{F}_t\}_{t\geq 0}$.)

Exercise 5.7.6. Let X be a one-dimensional Brownian motion. Show that

(i) $\{X_t^2 - t\}_{t\geq 0}$ is a continuous martingale,
(ii) $\{\exp(X_t - t/2)\}_{t\geq 0}$ is a nonnegative continuous martingale.

Using the law of the iterated logarithm or otherwise, discuss the behaviour of these martingales as $t \to \infty$.

Exercise 5.7.7. Let X be a Brownian motion in its completed natural filtration, that is $\mathcal{F}_t = \sigma(X_s : s \leq t) \vee \{\text{null sets}\}$. Show that the time-reversed process $\tilde{X}_t := tX_{1/t}$ is also a Brownian motion in its completed natural filtration. (Hint: Use Exercise 5.7.4.)

Exercise 5.7.8. Show that a Brownian motion starting at zero changes sign almost surely infinitely many times on the interval $[0, \epsilon]$ for any $\epsilon > 0$.

Exercise 5.7.9. Let N be a Poisson process. Given the times between jumps of N are exponentially distributed and independent, derive the distribution of $N_t - N_s$. Hence or otherwise, show that $\exp(\mu N_t - \lambda t(e^\mu - 1))$ is a martingale, for any $\mu \in \mathbb{R}$.

Exercise 5.7.10. Let N be a Poisson process, and X be the associated martingale $X_t = N_t - \lambda t$. Let T_n be the time of the nth jump of N and $\{\mathcal{F}_t\}_{t\geq 0}$ the completed filtration generated by N.

Show that a random variable Y is in $\mathcal{F}_{T_n}$ if and only if there is a Borel function $g : \mathbb{R}^n \to \mathbb{R}$ with $Y = g(T_1, T_2, \ldots, T_n)$.

Exercise 5.7.11. Let M be a martingale and C a nondecreasing measurable process starting at zero such that C_t is a stopping time for every t. If M is uniformly integrable or C_t is bounded for every t, show that $\tilde{M}_t := M_{C_t}$ is a martingale in the filtration $\tilde{\mathcal{F}}_t := \mathcal{F}_{C_t}$. We call C_t the 'time change' and $\tilde{M}$ the 'time changed martingale'.

Exercise 5.7.12. Using the law of the iterated logarithm or otherwise, for X a Brownian motion, show that $T := \inf\{t : X_t = a\}$ is finite almost surely, for any $a \in \mathbb{R}$. Show that T is a stopping time, and hence that the stopped process W^T is not uniformly integrable for any $a \neq 0$.

Exercise 5.7.13. Let X be a continuous local martingale and $T = \inf\{t : X_t \notin [a, b]\}$, for some $a, b \in \mathbb{R}$. Suppose $T < \infty$ a.s. Show that X^T is a uniformly integrable martingale. Writing $P_a = P(X_T = a) = P(X \text{ hits } a \text{ before } b)$, use the optional stopping theorem applied to X^T to give a formula for P_a in terms of a, b and X_0.

Exercise 5.7.14. Let X be a local supermartingale of class (D). Show that X is a supermartingale.

6

The Classification of Stopping Times

We now wish to classify stochastic processes more finely, which will be invaluable for the development of the theory of stochastic integration. In order to do this, we first study classes of stopping times.

Throughout this chapter, and the remainder of the book (unless otherwise stated) we shall suppose that $(\Omega, \mathcal{F}, P)$ is a complete probability space with a right-continuous complete filtration $\{\mathcal{F}_t\}_{t\in[0,\infty[}$ (so that, in particular $\mathcal{F}_0$ contains all P-null sets of $\mathcal{F}$). These assumptions are collectively known as the *usual conditions* on the filtration. All (in)equalities should be read as 'up to indistinguishability'.

Recall that a stopping time was defined in Definition 3.1.6 as a random variable T with values in $[0,\infty]$ such that $\{T \leq t\} \in \mathcal{F}_t$ for all $t \in [0,\infty[$. The σ-algebra $\mathcal{F}_T$ of events known at the stopping time T is given by

$$\mathcal{F}_T = \big\{A \in \mathcal{F} : A \cap \{T \leq t\} \in \mathcal{F}_t \text{ for all } t\big\}.$$

Definition 6.0.1. *Suppose S and T are maps $\Omega \to [0,\infty]$ and $S \leq T$ a.s. The* (half open) stochastic interval *denoted by $[\![S,T[\![$ is the set*

$$\big\{(t,\omega) \in [0,\infty[\times\Omega : S(\omega) \leq t < T(\omega)\big\}.$$

The stochastic intervals $[\![S,T]\!]$, $]\!]S,T]\!]$ and $]\!]S,T[\![$ are defined similarly. The stochastic interval

$$[\![T,T]\!] = \{(t,\omega) \in [0,\infty[\times\Omega : T(\omega) = t\}$$

is denoted by $[\![T]\!]$, and is called the graph *of T.*

S.N. Cohen, R.J. Elliott, *Stochastic Calculus and Applications*,
Probability and Its Applications, DOI 10.1007/978-1-4939-2867-5_6

Note that the stochastic intervals and graphs are defined to be subsets of $[0,\infty[\,\times\,\Omega$ (and not $[0,\infty]\times\Omega$). Of particular interest will be the graphs of stopping times (and less frequently, random times), as these have useful measurability properties. For s,t deterministic, we use this notation to succinctly write $[\![s,t[\![$ for the set $[s,t[\,\times\,\Omega$.

Example 6.0.2. Let Ω be the space of triples (a_1,a_2,a_3), where $a_i\in\{0,1\}$, endowed with the σ-algebra of all subsets of Ω. We can think of this as a model for the outcomes of throwing a coin three times, where 1 represents a head. The filtration is the natural one for the sequence of coin throws.

Let T be the stopping time corresponding to "T is the time of the first tail, or $T=3$ if there are no tails". Then the graph of T is the set

$$[\![T]\!]=\left\{\begin{matrix}(1,(0,0,0)),(1,(0,0,1)),(1,(0,1,0)),(1,(0,1,1)),\\(2,(1,0,0)),(2,(1,0,1)),(3,(1,1,0)),(3,(1,1,1))\end{matrix}\right\}.$$

Note that each outcome ω appears precisely once here (as $S(\omega)<\infty$ for all ω).

Example 6.0.3. Let $\Omega=[0,\infty]$ with the Lebesgue σ-algebra, and let $\mathcal{F}_t$ be the continuous-time filtration of the form $\mathcal{F}_t=\sigma(\{I_{\{\omega\leq s\}}\}_{s\leq t})$, and P some probability measure on Ω equivalent to Lebesgue measure. Then for any measurable function $T:[0,\infty]\to[0,\infty]$ with $T(\omega)\geq\omega$, the random variable $T(\omega)$ is a stopping time. The graph of T is the set

$$[\![T]\!]=\{(T(\omega),\omega):\omega<\infty,T(\omega)<\infty\}\subset[0,\infty[\times\Omega,$$

that is, the (Cartesian) graph of the function T (excluding ∞).

6.1 Events Before a Stopping Time

Definition 6.1.1. *The σ-algebra $\mathcal{F}_{T-}$ of events strictly prior to the stopping time T is the σ-algebra generated by $\mathcal{F}_0$ and all sets of the form $A\cap\{t<T\}$, where $t\in[0,\infty[$ and $A\in\mathcal{F}_t$, that is,*

$$\mathcal{F}_{T-}=\mathcal{F}_0\vee\sigma\big(A\cap\{t<T\}:A\in\mathcal{F}_t,t\in[0,\infty[\big).$$

Note that $\mathcal{F}_{T-}$ is a σ-algebra by definition.

Remark 6.1.2. We describe $\mathcal{F}_{T-}$ as the σ-algebra of events known strictly prior to the stopping time T, but intuition can deceive us here. For example, as we shall see in the following theorem, we always have that T is $\mathcal{F}_{T-}$-measurable, that is, the event $\{T\leq t\}$ is known 'strictly prior to the stopping time T'. This holds even if the event $\{T\leq t\}\notin\mathcal{F}_{t-\epsilon}$ for any $\epsilon>0$ (so at no time before t do we know if we will have stopped at t).

Intuitively, one can think of $\mathcal{F}_{T-}$ as the σ-algebra of events known at some time before T, together with knowledge of the value of T, but not of any other events which will be known at time T.

In particular, it is important to note that, if X is a random variable and $M_t = E[X|\mathcal{F}_t]$ up to indistinguishability, then it does *not* follow that $M_{T-} = E[X|\mathcal{F}_{T-}]$ (cf. Exercise 6.5.4).

Remark 6.1.3. In general, we cannot assume that there exist *any* stopping times $S < T$, so we cannot try to define $\mathcal{F}_{T-}$ in terms of $\bigvee_{S<T} \mathcal{F}_S$.

Theorem 6.1.4. *Suppose S and T are stopping times. Then*

(i) $\mathcal{F}_{T-} \subseteq \mathcal{F}_T$,
(ii) T *is* $\mathcal{F}_{T-}$*-measurable,*
(iii) if $T \leq S$, *then* $\mathcal{F}_{T-} \subseteq \mathcal{F}_{S-}$,
(iv) for every $A \in \mathcal{F}_S$, $A \cap \{S < T\} \in \mathcal{F}_{T-}$.

Proof. (i) It is enough to show that for any $t \in [0, \infty[$ and $A \in \mathcal{F}_t$ the set $A \cap \{t < T\}$ is in $\mathcal{F}_T$. Now for any $r \in [0, \infty[$,

$$A \cap \{t < T\} \cap \{T \leq r\} = \begin{cases} \emptyset, & \text{if } t \geq r, \\ A \cap \{t < T \leq r\}, & \text{if } t < r. \end{cases}$$

In either case, $A \cap \{t < T\} \cap \{T \leq r\} \in \mathcal{F}_r$. Hence $A \cap \{t < T\} \in \mathcal{F}_T$ by definition of $\mathcal{F}_T$.

(ii) Simply note $\{T \leq t\} = \{t < T\}^c \in \mathcal{F}_{T-}$ for all t.

(iii) We know $A \cap \{t < T\} = (A \cap \{t < T\}) \cap \{t < S\}$, so the sets generating $\mathcal{F}_{T-}$ are all in $\mathcal{F}_{S-}$.

(iv) One can write $A \cap \{S < T\} = \bigcup_{r \in \mathbb{Q}} \big\{A \cap \{S \leq r\} \cap \{r < T\}\big\}$. As $A \in \mathcal{F}_S$, we have that $A \cap \{S \leq r\} \in \mathcal{F}_r$.

□

Lemma 6.1.5. *Suppose T is a stopping time and $A \in \mathcal{F}_{\infty-}$. Then*

$$A \cap \{T = \infty\} \in \mathcal{F}_{T-}.$$

Proof. The sets $B \in \mathcal{F}_{\infty-}$ such that $B \cap \{T = \infty\} \in \mathcal{F}_T$ certainly form a σ-algebra. As $\mathcal{F}_{\infty-} = \bigvee_n \mathcal{F}_n$, it suffices to show that $A \cap \{T = \infty\} \in \mathcal{F}_{T-}$, whenever $A \in \mathcal{F}_n$ for some $n \in \mathbb{N}$. However,

$$A \cap \{T = \infty\} = \bigcap_{m=n+1}^{\infty} A \cap \{T > m\},$$

and by definition each set $A \cap \{T > m\} \in \mathcal{F}_{T-}$. □

Lemma 6.1.6. *Suppose S and T are stopping times and $S \leq T$. If $S < T$ on $\{0 < T < \infty\}$ then $\mathcal{F}_S \subset \mathcal{F}_{T-}$.*

Proof. By definition,

$$\mathcal{F}_S = \big\{A \in \mathcal{F} : A \cap \{S \leq t\} \in \mathcal{F}_t \text{ for all } t \in [0, \infty[\big\}.$$

Any $A \in \mathcal{F}_S$ can be written

$$A = \big(A \cap \{T = 0\}\big) \cup \big(A \cap \{S < T\}\big) \cup \big(A \cap \{T = \infty\}\big).$$

Now $A \cap \{T = 0\} \in \mathcal{F}_0 \subseteq \mathcal{F}_{T-}$ by definition, $A \cap \{S < T\} \in \mathcal{F}_{T-}$ by Theorem 6.1.4(iv) and $A \cap \{T = \infty\} \in \mathcal{F}_{T-}$ by Lemma 6.1.5. The result follows. $\square$

6.2 Predictable, Accessible and Totally Inaccessible Stopping Times

We now give the definition of a predictable stopping time. These are stopping times T where, in some sense, we are warned beforehand that we are about to reach T. Our definition is standard under the assumption that our filtration is complete (i.e. $\mathcal{F}_0$ contains all subsets of nullsets of $\mathcal{F}$); in the case where this does not hold, an alternative definition is preferable (see Remark 7.2.8), and the theory becomes more subtle and less intuitive.

Definition 6.2.1. *A stopping time T is said to be* predictable *(or previsible) if there is a sequence $\{T_n\}_{n\in\mathbb{N}}$, of stopping times such that*

(i) $\{T_n(\omega)\}_{n\in\mathbb{N}}$ is almost surely a nondecreasing sequence in $[0, \infty[$ and $\lim_n T_n(\omega) = T(\omega)$ a.s. and
(ii) on the set $\{T > 0\}$, $T_n(\omega) < T(\omega)$ a.s. for all n.

The sequence $\{T_n\}_{n\in\mathbb{N}}$ is said to announce *T.*

Remark 6.2.2. Predictable stopping times occur naturally in the real world. For example, consider a ship being driven onto a rocky coastline. The time T when the ship is wrecked is a stopping time announced by the family $\{T_n\}_{n\in\mathbb{N}}$, where T_n is the first time it is $1/n$ km. from the shore.

Example 6.2.3. For any stopping time $T < \infty$ and any real number $r > 0$, the random variable $T + r$ is a stopping time. Indeed, $T + r$ is a predictable stopping time because it is announced by the sequence $\{T_n\}_{n\in\mathbb{N}}$ where $T_n = T + r(1 - 1/n)$.

Because of this, for any stopping time T, the sequence $T_n = T + n^{-1}$ is a sequence of predictable stopping times such that $T_n \downarrow T$.

We now introduce other classes of stopping times, using definitions which involve their relation with predictable stopping times.

Definition 6.2.4. *A stopping time T is said to be* accessible *if there is a countable set $\{T_n\}_{n\in\mathbb{N}}$, of predictable stopping times such that*

$$[\![T]\!] \subset \bigcup_n [\![T_n]\!].$$

A stopping time T is said to be totally inaccessible *if for every predictable stopping time S,*

$$[\![T]\!] \cap [\![S]\!] = \emptyset,$$

up to an evanescent set, that is $P\big(\{\omega : T(\omega) = S(\omega) < \infty\}\big) = 0$.

Intuitively, an accessible stopping time is one which must equal one of a countable number of predictable stopping times, but we don't know which beforehand. For example, any stopping time in discrete time, or which must take a rational value, is accessible. (To see this, observe that such a stopping time must equal one of the rational numbers, which are stopping times as they are deterministic, and there are countably many such numbers.) Conversely, a totally inaccessible stopping time is one which has zero probability of equalling *any* predictable stopping time.

Example 6.2.5. Let $\Omega = [0, \infty]$ with the Borel σ-algebra, and let T be a random variable with values in $[0, \infty[$ such that $F(t) := P(T \le t)$ is continuous in t. (For example, T could be exponentially distributed.) Consider the filtration $\mathcal{F}_t = \sigma\big(\{T(\omega) \le s\}_{s\le t}\big)$, so the *only* information available at time t is whether $T \le t$, and if so, the value of T. Then T is a totally inaccessible stopping time.

To prove this, note that from the definition of the filtration, any stopping time S_n must be constant on $\{S_n < T\}$. Hence any predictable time S must be announced by a sequence of constants, and hence must also be a constant on $\{S \le T\}$. As $t \mapsto F(t)$ is continuous, the probability that T equals any constant is zero and therefore $[\![T]\!] \cap [\![S]\!] = \emptyset$ up to an evanescent set.

Lemma 6.2.6. *If the stopping time T is both accessible and totally inaccessible, then $T = \infty$ a.s.*

Proof. Note that

$$\{T < \infty\} = \big\{\omega \in \Omega : (t, \omega) \in [\![T]\!] \text{ for some } t \in [0, \infty[\big\}.$$

Since T is accessible, there exists a countable set of stopping times $\{T_n\}_{n\in\mathbb{N}}$ such that

$$[\![T]\!] \subset \bigcup_n [\![T_n]\!].$$

But since T is also totally inaccessible,

$$[\![T]\!] \cap [\![T_n]\!] = E_n$$

where E_n is an evanescent set for each n. Thus

$$[\![T]\!] = [\![T]\!] \cap \Big(\bigcup_n [\![T_n]\!]\Big) = \bigcup_n E_n,$$

which implies that $[\![T]\!]$ is evanescent, so $\{T < \infty\}$ has measure zero. $\square$

Definition 6.2.7. *If T is a stopping time and $A \in \mathcal{F}$, the restriction of T to A is the random variable T_A defined by*

$$T_A(\omega) = \begin{cases} T(\omega) & \text{for } \omega \in A, \\ \infty & \text{for } \omega \notin A. \end{cases}$$

Lemma 6.2.8. *T_A is a stopping time if and only if $A \in \mathcal{F}_T$.*

Proof. The result follows from the fact that

$$\{T_A \le t\} = A \cap \{T \le t\}. \qquad \square$$

Having defined accessible and totally inaccessible stopping times without much motivation, we now illustrate how these two concepts are actually the basic elements which constitute general stopping times. In particular, we show that every stopping time T admits an essentially unique decomposition in terms of an accessible stopping time and a totally inaccessible stopping time.

Theorem 6.2.9. *Suppose T is a stopping time. Then there is a partition of Ω into two elements A and B of $\mathcal{F}_T$, such that T_A is accessible and T_B is totally inaccessible. This partition is unique up to a set of measure zero.*

Proof. For any stopping times S and S', we shall say that S and S' are disjoint if $[\![S]\!] \cap [\![S']\!]$ is evanescent, that is, if $P(S = S' < \infty) = 0$. For any predictable stopping times S and S', by Definition 6.2.7 we can define a predictable stopping time $S'' = S'_{\{S \neq S'\}}$. Then S'' is disjoint from S, and $[\![S]\!] \cup [\![S']\!] = [\![S]\!] \cup [\![S'']\!]$.

For T a given stopping time and any stopping time S, we define $A(S) := \{S = T < \infty\} \in \mathcal{F}_T$. Clearly, for any S, S' disjoint, we know $A(S) \cap A(S')$ is a null set. Consider the set $\mathcal{A}_n(T)$ of all predictable times S_n with $P(S_n = T < \infty) = P(A(S_n)) > n^{-1}$. As $P(\Omega) = 1$, for each $n \in \mathbb{N}$ there is a collection of at most n disjoint predictable stopping times $\{S_{n,1}, S_{n,2}, \dots S_{n,n}\} \subset \mathcal{A}_n(T)$, such that $A(S') \cap \big(\cup_i A(S_{i,n})\big) \neq \emptyset$ for every $S' \in \mathcal{A}_n(T)$. Write $A = \bigcup_{n\in\mathbb{N}} \bigcup_{i=1}^n A(S_{n,i})$. It is clear that the induced stopping time T_A is accessible.

Define $B = A^c$, we claim T_B is totally inaccessible. To see this, for any predictable stopping time R, suppose $P(T_B = R < \infty) = P(T_B = R_B < \infty) > n^{-1}$ for some n. As $R_B = \infty$ on A, R_B is disjoint to $S_{n,i}$ for every i, but as $P(T = R < \infty) \ge P(T_B = R < \infty)$ this is a contradiction with the construction of $\{S_{n,i}\}_{i=1}^n$. Hence $P(T_B = R < \infty) = 0$.

The uniqueness is immediate from Lemma 6.2.6 and the following lemma. $\square$

Lemma 6.2.10. *Suppose T is a stopping time and $A \in \mathcal{F}_T$. If T is accessible (resp. totally inaccessible), then T_A is accessible (resp. totally inaccessible).*

Proof. The result follows easily from the fact that $[\![T_A]\!] \subset [\![T]\!]$. □

6.2.1 Limits of Stopping Times

We now present results on the predictability (or accessibility) of the limit of a sequence of stopping times.

Lemma 6.2.11. *Suppose S and T are two stopping times. If both S and T are predictable (resp. accessible, totally inaccessible), then $S \vee T$ and $S \wedge T$ are predictable (resp. accessible, totally inaccessible).*

Proof. Suppose $\{S_n\}_{n\in\mathbb{N}}$ announces S and $\{T_n\}_{n\in\mathbb{N}}$ announces T. Then $\{S_n \wedge T_n\}_{n\in\mathbb{N}}$ announces $S \wedge T$ and $\{S_n \vee T_n\}_{n\in\mathbb{N}}$ announces $S \vee T$. As $[\![S \wedge T]\!] \subseteq [\![S]\!] \cup [\![T]\!]$, the accessible and totally inaccessible cases follow directly. □

Theorem 6.2.12. *Suppose $\{T_n\}_{n\in\mathbb{N}}$ is a nondecreasing sequence of predictable (resp. accessible) stopping times. Then $T = \lim T_n$ is predictable (resp. accessible).*

Proof. Predictable Case. Suppose $\{S_{n,p}\}_{p\in\mathbb{N}}$ is a sequence of stopping times which announce T_n. For each $n \in \mathbb{N}$ write

$$S_n = \sup_{k\leq n, p\leq n} S_{k,p}.$$

Then $\{S_n\}_{n\in\mathbb{N}}$ announces T.

Accessible Case. Write $A = \{\omega : T_n(\omega) < T(\omega) \text{ for all } n\}$ and let S_n be the restriction of T_n to A. If $R_n := S_n \wedge n$, then $\{R_n\}_{n\in\mathbb{N}}$ announces T_A and so T_A is predictable. For $\omega \in A^c$, $T(\omega) = T_n(\omega)$ for some n, so

$$[\![T_{A^c}]\!] \subset \bigcup_n [\![T_n]\!].$$

Therefore, T_{A^c} is accessible and, by Lemma 6.2.11, we see that $T = T_A \wedge T_{A^c}$ is accessible. □

Theorem 6.2.13. *Suppose $\{T_n\}_{n\in\mathbb{N}}$ is a nonincreasing sequence of predictable (resp. accessible) stopping times and $T = \lim_n T_n$ a.s. If, for almost every $\omega \in \Omega$, $T(\omega) = T_n(\omega)$ for some n, then T is predictable (resp. accessible).*

Proof. Predictable Case. Clearly we can restrict our attention to $\{T < \infty\}$. Suppose $\{S_{n,p}\}_{p\in\mathbb{N}}$, is a sequence of stopping times announcing T_n such that

$$P(\{\omega : d(S_{n,p}(\omega), T_n(\omega)) > 2^{-p}\}) \leq 2^{-(n+p)},$$

for every p, where d is the metric on $[0,\infty]$ given by $d(s,t) = (|s-t|)/(1+|s-t|)$, $d(s,\infty) = d(\infty,t) = 1$ and $d(\infty,\infty) = 0$, for $s,t \in [0,\infty[$. Such a sequence can be obtained by selecting a subsequence from any sequence announcing T_n.

Put $S_p = \inf_n S_{n,p}$. This is a stopping time by Lemma 3.1.9. Then the sequence $\{S_p\}_{p\in\mathbb{N}}$, is increasing and, by hypothesis, $S_p < T$ on $\{T > 0\}$ for each p. Write $S = \lim_p S_p$. Then, for each $p \in \mathbb{N}$,

$$\begin{aligned} P(\{\omega : d(S(\omega),T(\omega)) > 2^{-p}\}) &\leq P(\{\omega : d(S_p(\omega),T(\omega)) > 2^{-p}\}) \\ &\leq \sum\nolimits_n P(\{\omega : d(S_{n,p}(\omega),T(\omega)) > 2^{-p}\}) \\ &\leq \sum\nolimits_n P(\{\omega : d(S_{n,p}(\omega),T_n(\omega)) > 2^{-p}\}) \leq 2^{-p}. \end{aligned}$$

Letting $p \to \infty$, we see that $P(S < T) = 0$ and so $S = T$ a.s.

Accessible Case. We have $[\![T]\!] \subset \bigcup_n [\![T_n]\!]$. Because the T_n are accessible the result is immediate from the definition. $\square$

Lemma 6.2.14. *If $\{T_n\}_{n\in\mathbb{N}}$ is a nondecreasing sequence of stopping times and $T = \lim_n T_n$, then $\mathcal{F}_{T-} = \bigvee_n \mathcal{F}_{T_n-}$.*

Proof. From Theorem 6.1.4 (iii), $\mathcal{F}_{T-} \supseteq \bigvee_n \mathcal{F}_{T_n-}$. Conversely, for any $t \in [0,\infty[$ and $A \in \mathcal{F}_t$,

$$A \cap \{t < T\} = \bigcup_n A \cap \{t < T_n\} \in \bigvee_n \mathcal{F}_{T_n-},$$

and so $\mathcal{F}_{T-} \subseteq \bigvee_n \mathcal{F}_{T_n-}$. $\square$

Corollary 6.2.15. *Suppose $\{T_n\}_{n\in\mathbb{N}}$ is a nondecreasing sequence of stopping times and $T = \lim_n T_n$. If $T_n < T$ on $\{0 < T < \infty\}$ then $\mathcal{F}_{T-} = \bigvee_n \mathcal{F}_{T_n}$.*

Proof. From Theorem 6.1.4(i) and Lemma 6.1.6, $\mathcal{F}_{T_n-} \subseteq \mathcal{F}_{T_n} \subseteq \mathcal{F}_{T-}$, so the result follows from Lemma 6.2.14. $\square$

Definition 6.2.16. *Let X be a càdlàg adapted (and hence progressive) process and T be a random time. Let $Y_t = X_{t-}$ be the left-limit process of X (which is also progressive). We shall write $X_{T-} = Y_T$ for the value of X immediately before the stopping time T. It is easy to check that X_{T-} is $\mathcal{F}_{T-}$ measurable.*

Remark 6.2.17. One can also define X_{T-} as the limit of the random variables $X_{T-\epsilon}$ as $\epsilon \downarrow 0$. However, one needs to be aware that $T-\epsilon$ is not a stopping time (it is a random time), so $X_{T-\epsilon}$ is generally only $\mathcal{F}_{T-}$-measurable, not $\mathcal{F}_{T-\epsilon}$-measurable.

The following extension of Theorem 5.6.13 gives one reason for the importance of predictable times.

Theorem 6.2.18. *If T is a predictable stopping time with announcing sequence $\{T_n\}_{n\in\mathbb{N}}$ and M is a càdlàg uniformly integrable martingale, then*

$$M_{T-} = \lim_n M_{T_n} = E[M_T|\mathcal{F}_{T-}] \quad \text{a.s.}$$

Proof. The statement $M_{T-} = \lim_n M_{T_n}$ is trivial. We know T is a predictable stopping time, so announced by a sequence $\{T_n\}_{n\in\mathbb{N}}$. Corollary 6.2.15 states that $\mathcal{F}_{T-} = \bigvee_n \mathcal{F}_{T_n}$, so Theorem 5.6.13 gives the result. □

6.3 Characterization of Predictable Stopping Times

Theorem 6.3.1. *Suppose T is a stopping time and $A \in \mathcal{F}_T \cap \mathcal{F}_{\infty-}$. If T_A is predictable then $A \in \mathcal{F}_{T-}$. Conversely, if T is predictable and $A \in \mathcal{F}_{T-}$ then T_A is predictable.*

Proof. Consider the family of sets

$$\mathcal{A} = \{A \in \mathcal{F}_T : T_A \text{ is predictable}\}.$$

We wish to show that $\mathcal{A} \subseteq \mathcal{F}_{T-}$, with $\mathcal{A} = \mathcal{F}_{T-}$ when T is predictable.

Suppose $A \in \mathcal{A}$, that is, T_A is predictable, and let $\{S_n\}_{n\in\mathbb{N}}$ be a sequence of stopping times announcing T_A. We can write

$$A = \{T_A \leq T\} \setminus (A^c \cap \{T = \infty\}).$$

Now $\{T_A \leq T\} = \{T = 0\} \cup \big(\bigcap_n \{S_n < T\}\big)$, and $\{S_n < T\} \in \mathcal{F}_{T-}$ from Theorem 6.1.4(iv). Also $A \in \mathcal{F}_T \cap \mathcal{F}_{\infty-}$, so $A^c \in \mathcal{F}_{\infty-}$ and, by Lemma 6.1.5, $A^c \cap \{T = \infty\} \in \mathcal{F}_{T-}$. Consequently, $A \in \mathcal{F}_{T-}$.

Conversely, suppose T is predictable. Then $\Omega \in \mathcal{A}$, and from Theorems 6.2.12 and 6.2.13, $\mathcal{A}$ is closed under countable unions and intersections. Hence $\mathcal{A}$ is a monotone class, so to show $\mathcal{F}_{T-} \subseteq \mathcal{A}$, it suffices to prove that an algebra generating $\mathcal{F}_{T-}$ is in $\mathcal{A}$ (by Theorem 1.1.14).

If $\{T_n\}_{n\in\mathbb{N}}$ is a sequence of stopping times announcing T, from Corollary 6.2.15, $\mathcal{F}_{T-} = \bigvee_n \mathcal{F}_{T_n}$. Therefore, $\bigcup_n \mathcal{F}_{T_n}$ generates $\mathcal{F}_{T-}$, and so we can suppose $A \in \mathcal{F}_{T_n}$ for some n. Defining $S_m = (T_{n+m})_A \wedge m$, the sequence of stopping times $\{S_m\}_{m\in\mathbb{N}}$ announces T_A, so T_A is predictable, and the result follows. □

Theorem 6.3.2. *Suppose S is a predictable stopping time and T an arbitrary stopping time. For any $A \in \mathcal{F}_{S-}$ the set $A \cap \{S \leq T\}$ belongs to $\mathcal{F}_{T-}$. In particular, the sets $\{S \leq T\}$ and $\{S = T\}$ belong to $\mathcal{F}_{T-}$.*

Proof. Let $\{S_n\}_{n\in\mathbb{N}}$ be a sequence of stopping times announcing S and suppose $A \in \mathcal{F}_{S_m}$ for some m. Then

$$A \cap \{S \leq T\} = (A \cap \{S = 0\}) \cup \Big(\bigcap_n (A \cap \{S_{m+n} < T\})\Big).$$

But $A\cap\{S=0\}\in\mathcal{F}_0\subset\mathcal{F}_{T-}$ and from Theorem 6.1.4(iv) $A\cap\{S_{m+n}<T\}\in\mathcal{F}_{T-}$. Therefore, for $A\in\mathcal{F}_{S_m}$, we have $A\cap\{S\le T\}\in\mathcal{F}_{T-}$. However, the family of sets

$$\{A : A\cap\{S\le T\}\in\mathcal{F}_{T^-}\}$$

is clearly a σ-algebra, and from Corollary 6.2.15, $\mathcal{F}_{S-}=\bigvee_m\mathcal{F}_{S_m}$, so $A\cap\{S\le T\}\in\mathcal{F}_{T-}$ for any $A\in\mathcal{F}_{S-}$. In particular $\{S\le T\}\in\mathcal{F}_{T-}$.

From Theorem 6.1.4(iv), for general stopping times S and T we know that $\{S<T\}\in\mathcal{F}_{T-}$. Therefore, $\{S=T\}=\{S\le T\}\setminus\{S<T\}\in\mathcal{F}_{T-}$. $\square$

Remark 6.3.3. From Theorem 3.1.13 and Theorem 6.1.4(iv) we know that the sets

$$\{S\le T\},\ \{S<T\},\ \{S=T\},\ \{S>T\},\ \{S\ge T\}$$

belong to both $\mathcal{F}_S$ and $\mathcal{F}_T$. Reversing the roles of S and T in Theorem 6.1.4(iv) we see that the event $\{T<S\}$ (and so $\{S\le T\}$) belongs to $\mathcal{F}_{S-}$, but not, in general, to $\mathcal{F}_{T-}$. However, the above result shows that when S is predictable, the events $\{T<S\}$ and $\{S\le T\}$ also belong to $\mathcal{F}_{T-}$.

Lemma 6.3.4. *Suppose S and T are predictable stopping times. Then T_A is predictable, where $A=\{T<S\}$.*

Proof. From Theorem 6.3.2 and Remark 6.3.3 we see that $A\in\mathcal{F}_{T-}$. Therefore the result follows from Theorem 6.3.1. $\square$

Theorem 6.3.5. *Suppose S is an accessible stopping time. Then S is predictable if and only if the set $\{S=T\}\in\mathcal{F}_{T-}$ for every predictable stopping time T.*

Proof. If S is predictable then $\{S=T\}\in\mathcal{F}_{T-}$ by Theorem 6.3.1. Conversely, suppose S is accessible and $\{S_n\}_{n\in\mathbb{N}}$ is a sequence of predictable stopping times such that

$$[\![S]\!]\subset\bigcup_n[\![S_n]\!].$$

Then, for each n, by Theorem 6.3.1 and Remark 6.3.3,

$$\{S\le S_n\}\in\mathcal{F}_{S_n-}.$$

Write T_n for the restriction of S_n to $\{S\le S_n\}$, so that T_n is predictable (Theorem 6.3.1). For each n, the stopping time

$$R_n=T_1\wedge T_2\wedge\cdots\wedge T_n$$

is predictable. The sequence $\{R_n\}_{n\in\mathbb{N}}$ is decreasing, and for each $\omega\in\Omega$ there is an n such that $R_n(\omega)=S(\omega)$. Therefore, by Theorem 6.2.13, S is predictable. $\square$

6.4 Quasi-Left Continuity

Finally, we give a useful notion of 'predictable' left-continuity for the filtration. This notion is sufficient to guarantee that many of the processes we study (in particular, martingales) will not jump at predictable times.

Definition 6.4.1. *The filtration $\{\mathcal{F}_t\}_{t\in[0,\infty[}$ is said to be quasi-left continuous if for every predictable stopping time T*

$$\mathcal{F}_{T-} = \mathcal{F}_T.$$

The word "quasi" is used because equality of $\mathcal{F}_{T-}$ and $\mathcal{F}_T$ is only required for predictable stopping times. Note that quasi-left continuity is a stronger assumption than stating '$\mathcal{F}_t = \mathcal{F}_{t-}$ for all $t > 0$', but weaker than '$\mathcal{F}_T = \mathcal{F}_{T-}$ for all stopping times T.'

Theorem 6.4.2. *Suppose $\mathcal{F} = \mathcal{F}_{\infty-}$. Then the following three properties are equivalent:*

(i) the filtration $\{\mathcal{F}_t\}_{t\in[0,\infty[}$ is quasi-left continuous,
(ii) if $\{T_n\}_{n\in\mathbb{N}}$ is any nondecreasing sequence of stopping times then

$$\mathcal{F}_{(\lim T_n)} = \bigvee_n \mathcal{F}_{T_n},$$

that is, the filtration has no predictable times of discontinuity,
(iii) the accessible stopping times are predictable.

Proof. (ii)⇒(i). For any predictable T, take $\{T_n\}_{n\in\mathbb{N}}$ to be an announcing sequence for T. By Corollary 6.2.15 we have $\mathcal{F}_T = \bigvee_n \mathcal{F}_{T_n} = \mathcal{F}_{T-}$.

(i)⇒(iii). By Remark 6.3.3 and the quasi-left continuity of $\{\mathcal{F}_t\}_{t\in[0,\infty[}$ we have $\{S = T\} \in \mathcal{F}_T = \mathcal{F}_{T-}$. The result follows from Theorem 6.3.5.

(iii)⇒(ii). Suppose $\{T_n\}_{n\in\mathbb{N}}$ is a nondecreasing sequence of stopping times, and write $T = \lim T_n$. Let R (resp. S) denote the accessible (resp. totally inaccessible) part of T. For any set $A \in \mathcal{F}_T$ we can write

$$A = \big(\{R_A = T\} \setminus (A^c \cap \{T = \infty\})\big) \cup \{S_A < \infty\} \cup (A \cap \{T = \infty\}).$$

From Lemma 6.1.5, the sets $A^c \cap \{T = \infty\}$ and $A \cap \{T = \infty\}$ belong to $\mathcal{F}_{T-}$, and so to $\bigvee_n \mathcal{F}_{T_n-} \subseteq \bigvee_n \mathcal{F}_{T_n}$ by Lemma 6.2.14. By hypothesis, R_A is predictable, so, by Theorem 6.3.2 and Lemma 6.2.14,

$$\{R_A = T\} \in \mathcal{F}_{T-} = \bigvee_n \mathcal{F}_{T_n-} \subseteq \bigvee_n \mathcal{F}_{T_n}.$$

Consider the set $\{S_A < \infty\}$. Because S_A is totally inaccessible, $\{T_n\}_{n\in\mathbb{N}}$ cannot announce S_A, that is,

$$\{S_A < \infty\} = \bigcap_n \{S_A = T_n < \infty\} \in \bigvee_n \mathcal{F}_{T_n}.$$

Therefore, $A \in \bigvee_n \mathcal{F}_{T_n}$ and the result is proven. □

In a similar way, we can also define a notion of quasi-left continuity for a particular process.

Definition 6.4.3. *A càdlàg adapted process $\{X_t\}_{t\in[0,\infty[}$ is said to be quasi-left continuous if, for every predictable stopping time T,*

$$X_{T-} = X_T \qquad \text{a.s.}$$

Theorem 6.4.4. *Any local martingale adapted to a quasi-left continuous filtration is a quasi-left continuous process.*

Proof. For M a local martingale, we know from Theorem 5.6.4 that M is locally uniformly integrable. Let $\{S_n\}_{n\in\mathbb{N}}$ be a localizing sequence, so M^{S_n} is uniformly integrable. Then for any predictable stopping time T, we know $\mathcal{F}_T = \mathcal{F}_{T-}$, and Theorem 6.2.18 states that

$$M^{S_n}_{T-} = E[M^{S_n}_T|\mathcal{F}_{T-}] = E[M^{S_n}_T|\mathcal{F}_T] = M^{S_n}_T.$$

As $S_n \to \infty$ a.s., we know that $P(S_n < T < \infty) \to 0$ as $n \to \infty$. Therefore, $M^{S_n}_{T-} \to M_{T-}$ and $M^{S_n}_T \to M_T$ a.s., and the result follows. □

Remark 6.4.5. One can show that the filtration generated by a Poisson process is quasi-left continuous (Exercise 6.5.7) and similarly for a Brownian motion (Lemma 14.5.2). Example 6.5.4 then demonstrates that, even in the simple Poisson setting, there are some stopping times T such that $\mathcal{F}_T \neq \mathcal{F}_{T-}$.

The following example shows that left continuity at deterministic times does not imply quasi-left continuity.

Example 6.4.6. Let $S, T \sim U([0,1])$ be independent uniformly distributed random variables. Consider the filtration $\{\mathcal{F}^0_t\}_{t\geq 0}$ generated by the processes $I_{\{T\leq t\}}$ and $SI_{\{T\leq t-1\}}$, and its completion $\{\mathcal{F}_t\}_{t\geq 0}$. For any deterministic time t, $\mathcal{F}^0_t = \mathcal{F}^0_{t-} \vee \{T = t\} \vee \sigma(SI_{\{T=t-1\}})$. However the events $\{T = t\}$ and $\{T = t-1\}$ both have probability zero, so by completeness $\mathcal{F}_t = \mathcal{F}_{t-}$. However, the time $T+1$ is predictable, and S is $\mathcal{F}^0_{T+1}$-measurable but not $\mathcal{F}^0_{(T+1)-}$-measurable. As the value of S is not trivial, it follows that $\mathcal{F}_{T+1} \neq \mathcal{F}_{(T+1)-}$.

6.5 Exercises

Exercise 6.5.1. Show that any stopping time T can be written as the decreasing limit of a sequence of predictable stopping times T_k, where T_k takes at most finitely many values.

Exercise 6.5.2. Consider Example 6.2.5, but where $F_T(t)$ may be discontinuous. Describe the accessible and totally inaccessible parts of T in terms of F_T.

Exercise 6.5.3. Give a counterexample to Theorem 6.2.13 when T_n is strictly decreasing in n. (You may assume the existence of a non-predictable stopping time.)

Exercise 6.5.4. Let X be a Poisson process and $M_t = E[X_1|\mathcal{F}_t] = X_{t\wedge 1} + \lambda(1-t)^+$. Show that $M_{T-} \neq E[X_1|\mathcal{F}_{T-}]$, for T the time of the first jump of X.

Exercise 6.5.5. Show that a filtration in discrete time can be quasi-left continuous if and only if it is trivial (that is, $\mathcal{F}_t = \mathcal{F}_0$ for all t).

Exercise 6.5.6. Let N be a Poisson process under its completed natural filtration. Show that the jump times of N are totally inaccessible.

Exercise 6.5.7. Let N be a Poisson process under its completed natural filtraion, and for each $k \in \mathbb{N}$ let T_k denote the kth jump of N.

(i) For each k, describe the stopping times S such that $P(\{T_k \leq S < T_{k+1}\} \cup \{S = \infty\}) = 1$.
(ii) Show that the completed natural filtration of the Poisson process is quasi-left continuous.

Exercise 6.5.8. Let X be a càdlàg adapted process in space with a right-continuous filtration. Show that for any $k \in \mathbb{R}$, the stopping time $T = \inf\{t : X_t > k\}$ is predictable if X is a.s. continuous.

Exercise 6.5.9. Give an example of a filtered probability space with right-continuous filtration $\{\mathcal{F}_t\}_{t\in[0,\infty[}$ (so $\mathcal{F}_t = \mathcal{F}_{t+}$ for all $t \in [0,\infty[$) and a stopping time T such that

$$\mathcal{F}_T \neq \bigcap_{\epsilon>0} \mathcal{F}_{T+\epsilon}.$$

Hint: Start with a setting where $\mathcal{F}_t \neq \mathcal{F}_{t+}$ for some t, then randomize the point of discontinuity as in Example 6.4.6.

Exercise 6.5.10. Let $\{T_n\}_{n\in\mathbb{N}}$ and $\{S_n\}_{n\in\mathbb{N}}$ be sequences of nondecreasing stopping times. Show that there exists a sequence of stopping times $\{R_n\}_{n\in\mathbb{N}}$ such that $R_n < R_{n+1}$ a.s. on the set $\{R_n < \infty\}$, and

$$\bigcup_n [\![R_n]\!] = \bigcup_n \big([\![T_n]\!] \cup [\![S_n]\!]\big).$$

Show that if $\{T_n, S_n\}_{n\in\mathbb{N}}$ are predictable, then one can choose $\{R_n\}_{n\in\mathbb{N}}$ predictable.

7

The Progressive, Optional and Predictable σ-Algebras

We now move from looking at different types of stopping times to different types of processes. Recall that we defined a real-valued process Y to be progressive (Definition 3.2.25) if, for every t, the map $(s,\omega) \mapsto X_s(\omega)$ of $[0,t] \times \Omega$ into $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is measurable, when $[0,t] \times \Omega$ is given the product σ-algebra $\mathcal{B}([0,t]) \otimes \mathcal{F}_t$. Essentially, this states that the process X is adapted *and* is Borel measurable with respect to time.

However, for many problems, this space is too broad. Instead, we wish to consider those processes which are not only measurable through time, but which have some continuity properties, for example, càdlàg processes or left-continuous processes. Restricting our attention to these spaces yields different σ-algebras on the product space $[0,\infty] \times \Omega$, which we will use extensively. This chapter is devoted to exploring these σ-algebras and the associated processes. This approach may seem unduly abstract, and this chapter is one of the most abstract in the book, but we shall see that understanding these σ-algebras enables our understanding of a wide range of processes of interest.

Again we shall work with a probability space $(\Omega, \mathcal{F}, P)$ with a filtration $\{\mathcal{F}_t\}_{t\in[0,\infty[}$ satisfying the usual conditions (completeness and right-continuity). For notational simplicity, we recall that $\mathcal{F}_{\infty-} = \bigvee_t \mathcal{F}_t$.

7.1 Progressive, Optional and Predictable σ-Algebras

We begin by defining the σ-algebras in question in terms of stochastic intervals. When we do this with generic stopping times, we obtain the 'optional' σ-algebra, when we work with predictable stopping times we obtain the 'predictable' σ-algebra. One can also define, in the same way, the accessible σ-algebra; however this will not be of interest here.

S.N. Cohen, R.J. Elliott, *Stochastic Calculus and Applications*,
Probability and Its Applications, DOI 10.1007/978-1-4939-2867-5_7

Definition 7.1.1. *We shall denote the set of all stopping times by $\mathcal{T}$, and all predictable stopping times by $\mathcal{T}_p$. For notational consistency with what follows, we shall also write $\mathcal{T} = \mathcal{T}_o$.*

Definition 7.1.2. *The optional (resp. predictable) σ-algebra Σ_o (resp. Σ_p) on $[0,\infty[\times\Omega$ is the σ-algebra generated by the evanescent sets and all stochastic intervals of the form $[\![T,\infty[\![$ for T an arbitrary (resp. predictable) stopping time.*

Remark 7.1.3. Note that this is equivalent to saying Σ_x is generated by the intervals $[\![T,S[\![$, where $T,S\in\mathcal{T}_x$, for $(x=o,p)$. Also, we immediately have that $\Sigma_p\subseteq\Sigma_o$, and the inclusion is typically strict.

As we shall see, these σ-algebras have alternative equivalent definitions, either in terms of different stochastic intervals (Lemma 7.1.7, Theorem 7.1.9 and Corollary 7.2.5), or in terms of processes (Theorems 7.2.4 and 7.2.7).

One should always remember that Σ_x $(x=o,p)$ is a σ-algebra on the product space $[0,\infty[\times\Omega$, and so Σ_x-measurable functions correspond to *stochastic processes*, rather than random variables. Intuitively, (once we have seen Theorems 7.2.4 and 7.2.7), we can think of Σ_o-measurable processes as being approximable by adapted *càdlàg* processes, and Σ_p-measurable processes as being approximable by adapted *left-continuous* processes.

Definition 7.1.4. *A set $A\subseteq[0,\infty[\times\Omega$ is said to be* progressively measurable*, or* progressive*, if its indicator function I_A is a progressive process. The family of all progressive sets forms a σ-algebra Σ_π, and it is easy to see that a process $\{X_t\}_{t\in[0,\infty[}$ is progressive if and only if the map $(t,\omega)\to X_t(\omega)$ is measurable with respect to Σ_π.*

Exercise 7.7.2 provides a direct characterization of progressive sets.

Remark 7.1.5. As $I_{[\![T,S[\![}$ is adapted and right-continuous, the optional σ-algebra is contained in the progressive σ-algebra, that is, $\Sigma_o\subseteq\Sigma_\pi$.

Remark 7.1.6. In many settings, it is possible, but nontrivial, to construct sets which are in Σ_π but not Σ_o, so the inclusion $\Sigma_o\subseteq\Sigma_\pi$ is typically strict. See Appendix A.5 for a construction of such a set in a space containing a Brownian motion.

We have the following alternative descriptions of Σ_o and Σ_p, based on different stochastic intervals.

Lemma 7.1.7. *For $x=o,p$, the σ-algebra Σ_x is generated by the intervals $[\![T,S]\!]$, for $T\in\mathcal{T}_x$ and $S\in\mathcal{T}_o$.*

Proof. That Σ_o is generated by the stated intervals is true by definition. For Σ_p, we know that an arbitrary stopping time $S\in\mathcal{T}_o$ is the limit, from above, of the sequence of predictable stopping times $S_n=S+1/n$. Therefore, $[\![T,S]\!]=\bigcap_n[\![T,S_n[\![$ and $[\![T,S]\!]\in\Sigma_x$ if $T\in\mathcal{T}_x(x=o,p)$.

Conversely for $x=o,p$, if $S,T\in\mathcal{T}_x$, then $[\![S]\!]=\bigcap_n[\![S,S+1/n]\!]$ and $[\![T,S[\![=[\![T,S]\!]\setminus[\![S]\!]$. Therefore the generating intervals $[\![T,S[\![$ for Σ_x are in the σ-algebra generated by the intervals $[\![T,S]\!]$. $\square$

Remark 7.1.8. Taking $S = T \in \mathcal{T}_x$ immediately shows that $[\![S]\!] \in \Sigma_x$. A converse result can be found in Lemma 7.3.6.

Theorem 7.1.9. *For $A \in \mathcal{F}_0$, write 0_A for the stopping time which is 0 on $A \in \mathcal{F}_0$ and ∞ on A^c (cf. Definition 6.2.7). Then Σ_p is generated by stochastic intervals of the form $\{[\![0_A]\!] \cup]\!]S,T]\!]\}_{A\in\mathcal{F}_0, S,T\in\mathcal{T}_o}$.*

Proof. Let $\mathcal{E}$ be the σ-algebra generated by these intervals. To show $\mathcal{E} \subseteq \Sigma_p$, first note that $[\![0_A]\!] = \{\{0\}\times A\} \in \Sigma_p$ as $0_A \in \mathcal{T}_p$. By Lemma 7.1.7, $[\![0,T]\!] \in \Sigma_p$ and $[\![S+1/n,\infty[\![\in \Sigma_p$, because $S+1/n \in \mathcal{T}_p$. Therefore,

$$]\!]S,T]\!] = \bigcup_n [\![S+1/n,\infty[\![\cap [\![0,T]\!] \in \Sigma_p.$$

As $\mathcal{E}$ is generated by elements of Σ_p, we have $\mathcal{E} \subseteq \Sigma_p$.

Conversely, consider a stochastic interval $[\![S,\infty[\![$ for S predictable. As these intervals generate Σ_p, we must show that $[\![S,\infty[\![$ is in $\mathcal{E}$. Now $B = \{S = 0\} \in \mathcal{F}_0$ and, if $\{S_n\}_{n\in\mathbb{N}}$ is a sequence of stopping times announcing S on $\{S > 0\}$, we can write $[\![S,\infty[\![= [\![0_B]\!] \cup (\bigcap_n]\!]S_n,\infty[\![) \in \mathcal{E}$. Therefore, $\mathcal{E} = \Sigma_p$. □

7.2 Optional and Predictable Processes

We now consider the optional and predictable processes, that is, the measurable functions on $[0,\infty[\times\Omega$ with the σ-algebra Σ_x (for $x = o, p$). These form a very broad and useful class of processes to study.

Definition 7.2.1. *A stochastic process $\{X_t\}_{t\in[0,\infty[}$ defined on $(\Omega,\mathcal{F})$, with values in the measurable space $(E,\mathcal{E})$, is said to be optional (resp. predictable) if the map $X : [0,\infty[\times\Omega \to E$ is measurable, when $[0,\infty[\times\Omega$ is given the optional (resp. predictable) σ-algebra.*

Remark 7.2.2. We should compare these definitions with the definition of a progressive process (Definition 3.2.25). By inspecting the definitions, it is clear that the progressive processes are those which are Σ_π measurable. As $\Sigma_p \subseteq \Sigma_o \subseteq \Sigma_\pi$, any predictable process is optional, and any optional process is also progressive.

Example 7.2.3. (i) Suppose $[\![S,T]\!]$ is a stochastic interval and Z is an $\mathcal{F}_S$-measurable random variable. Then the process X defined by

$$X_t(\omega) = Z(\omega) I_{[\![S,T[\![}(t,\omega)$$

is optional. If Z is $\mathcal{F}_{S-}$-measurable and S and T are predictable, then X is predictable. These statements are easily verified by first considering the case $Z = I_A$, where $A \in \mathcal{F}_S$ (resp. $\mathcal{F}_{S-}$). The process X is then the indicator function of the interval $[\![S_A,T_A]\!]$, and the result follows from Lemma 6.2.8 and Theorem 6.3.1. The general case follows by approximation.

(ii) In part (i), $[\![S,T[\![$ can be replaced by $[\![S,T]\!]$, where $T \in \mathcal{T}_o$ in all cases (because $[\![S,T]\!] = \bigcap_n [\![S,T+1/n[\![$).
(iii) For $S,T \in \mathcal{T}_o$ we know from Theorem 7.1.9 that $]\!]S,T]\!] \in \Sigma_p$. A pointwise approximation argument then shows that, if Z is $\mathcal{F}_S$-measurable, then the process $ZI_{]\!]S,T]\!]}$ is predictable.

We can now give another description of the σ-algebras Σ_o and Σ_p.

Theorem 7.2.4. *The predictable σ-algebra Σ_p is generated by the family of left-continuous adapted stochastic processes.*

Proof. From Theorem 7.1.9, we see that Σ_p is generated by the processes of the form $I_{[\![0_A]\!]}$ for $A \in \mathcal{F}_0$, and $I_{]\!]S,T]\!]}$ for $S,T \in \mathcal{T}_o$. These processes are left-continuous and adapted, so it remains to be shown that every left-continuous adapted process X is predictable. For any such X, write

$$X^n = X_0 I_{[\![0]\!]} + \sum_{k\geq 1} X_{k/n} I_{]\!]k/n,(k+1)/n]\!]}, \tag{7.1}$$

so X^n is a left-continuous step function and $X^n \to X$ pointwise. For each n, the process X^n is predictable because it is the countable sum of predictable processes (by Example 7.2.3(iii)). Therefore, by Lemma 1.3.28, we see their limit X is Σ_p-measurable, that is, X is predictable. □

Corollary 7.2.5. *The predictable σ-algebra Σ_p is generated by the family of stochastic intervals of the form $[\![0_A]\!]$ for $A \in \mathcal{F}_0$, and $[\![s_B, t_B]\!]$ for $B \in \mathcal{F}_{s-}$, where s and t are constants (cf. Definition 6.2.7).*

Proof. These intervals are predictable. Conversely, because the random variables $X_{k/n}$ in (7.1) are $\mathcal{F}_{(k/n)-}$-measurable, one can see that any left continuous adapted process is the pointwise limit of linear combinations of the indicator functions of these intervals. Hence, the intervals must generate Σ_p. □

Corollary 7.2.6. *Σ_p is generated by the family of all continuous adapted processes.*

Proof. An adapted continuous process is certainly left-continuous and so predictable by Theorem 7.2.4. Conversely, note that an interval $]\!]S,\infty[\![$ is equal to the set $\{X > 0\}$, where X is the continuous adapted process $X_t(\omega) = t - S(\omega) \wedge t$. As $\{tI_{[\![0_A,\infty[\![}\}_{t\geq 0}$ is continuous and adapted if $A \in \mathcal{F}_0$, we see that the σ-algebra generated by continuous processes contains $[\![0_A]\!] = [\![0_A,\infty[\![\setminus]\!]0_A,\infty[\![$ and $]\!]S,T]\!] =]\!]S,\infty[\![\setminus]\!]T,\infty[\![$ for all stopping times S,T and all $A \in \mathcal{F}_0$. Hence it contains Σ_p, by Theorem 7.1.9. □

Theorem 7.2.7. *Σ_o is generated by the family of all adapted processes which are continuous on the right and have limits on the left, that is, all adapted càdlàg processes.*

Proof. By definition, Σ_o is generated by the càdlàg adapted processes of the form $\{I_{[\![S,T[\![}\}_{S,T\in\mathcal{T}_o}$. Consequently, it remains to be shown that every adapted càdlàg process is optional.

For an adapted càdlàg process X and for each integer $k > 0$ define an increasing sequence of stopping times as follows:

$$\begin{aligned} T_1^k(\omega) &= 0 \quad \text{for all } \omega \text{ and all } k, \\ T_{n+1}^k(\omega) &= \inf\{t > T_n^k(\omega) : |X_t(\omega) - X_{T_n^k}(\omega)| \geq 1/k\}, \end{aligned}$$

with the convention $\inf\{\emptyset\} = \infty$. Because X is right-continuous, we have $|X_{T_{n+1}^k} - X_{T_n^k}| \geq 1/k$ on the set $\{T_{n+1}^k < \infty\}$, and the existence of left limits implies that T_n^k converges to infinity for each integer k. For $k > 0$, write

$$X^k = \sum_{n=1}^{\infty} X_{T_n^k} I_{[\![T_n^k, T_{n+1}^k[\![}.$$

Then X^k is optional, because it is a countable sum of elementary optional processes. Letting k tend to infinity, by right-continuity, $X_t^k(\omega)$ approaches $X_t(\omega)$, except possibly on an evanescent set. Therefore X is optional. $\square$

It can also be shown that a general right-continuous adapted process is optional, but this result will not be needed in the sequel. See Dellacherie [53, Chapter IV, Theorem 27].

Remark 7.2.8. Our definition of predictable stopping times and the predictable σ-algebra has, up to this point, been based on the assumption that the filtration is complete. In some applications, this is a problematic assumption, and so a different approach is needed.

In the absence of completeness, we define the *predictable σ-algebra* first, using the result of Theorem 7.2.4 as a definition, that is, Σ_p is the sigma algebra generated by the left-continuous adapted processes. We then define a predictable stopping time to be a stopping time T such that $[\![T]\!] \in \Sigma_p$. With this definition, many of the results we have established remain valid; however one needs to be much more careful with sets of measure zero. For a rigorous and careful approach which covers this situation, see the first chapter of Jacod and Shiryaev [110] or the classic text of Dellacherie and Meyer [54].

7.3 The Debut and Sections of a Set

For stochastic intervals $[\![S,T]\!]$, we can think of S as the start (in French *début*) of the interval. By construction, this is a stopping time. For general sets $A \subseteq [0,\infty[\times\Omega$, we now define a similar notion, which we call the debut D_A of the set. These will often appear as the 'first time' some event occurs. This section discusses when D_A is a stopping time, and the relationship of the graph $[\![D_A]\!]$ with the original set A (Fig. 7.1).

Definition 7.3.1. *Suppose* $A \subseteq [0,\infty[\times\Omega$. *The map* $D_A(\omega) = \inf\{t \in [0,\infty[: (t,\omega) \in A\}$ *is called the* debut *of* A. *Here we again use the convention that the infimum of the empty set is* ∞.

Note that without some restrictions on the set A, there is no guarantee that D_A is measurable.

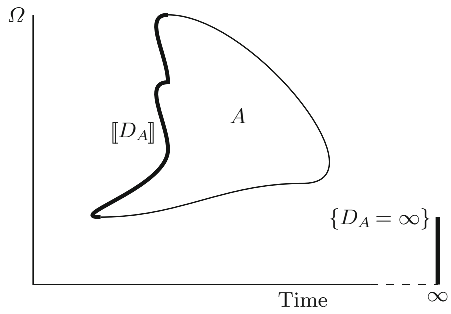

Fig. 7.1. A depiction of a set $A \subset [0,\infty[\times \Omega$, with the graph of its debut (and the set $\{D_A = \infty\}$) in bold. Recall that Ω is the set of all outcomes, not the state space of some process.

Example 7.3.2. Let Ω be the space of sequences of heads and tails from an infinite series of coin tosses, with the usual filtration (obtained by observing the coin tosses sequentially at integer times). We write $\omega = \omega_1\omega_2\ldots$ for $\omega_i \in \{H,T\}$.

If A is the set of 'all heads that were immediately preceded by a tail' (that is, $(t,\omega) \in A$ if $\omega_{t-1} = T$ and $\omega_t = H$), then the debut D_A is given by

$$D_A(\omega) = \min\{t : \omega_{t-1} = T, \omega_t = H\}$$

and is a stopping time.

If B is the set of 'all heads which were followed by a head', then the debut D_B is given by

$$D_B(\omega) = \min\{t : \omega_t = H, \omega_{t+1} = H\}$$

and is not a stopping time.

We quote the following theorem without proof from Dellacherie [53, Chapter I, Theorem 32]. While the statement of the theorem is straightforward, its proof is remarkably difficult, depending on fine analysis of measurable sets (in particular, the results of 'capacity theory').

Theorem 7.3.3 (Measurable Projection Theorem). *Suppose* $(\Omega, \mathcal{F}, P)$ *is a complete probability space and* $\mathcal{B}([0,t[)$ *is the Borel* σ*-algebra on* $[0,t[$*. Write* π *for the projection map of* $[0,t[\times\Omega$ *onto* Ω*, that is,*

$$\pi(A) = \{\omega : (s,\omega) \in A \text{ for some } s < t\} \subseteq \Omega.$$

If A *is a measurable set in the product* σ*-algebra* $\mathcal{B}([0,t[) \otimes \mathcal{F}$*, then the projection* $\pi(A)$ *is in* $\mathcal{F}$*.*

Using Theorem 7.3.3 we now show that the debut of a progressive set is a stopping time.

Theorem 7.3.4. *For a filtered probability space with a complete filtration* $\{\mathcal{F}_t\}_{t\in[0,\infty[}$*, the debut* D_A *of a progressively measurable set* $A \in \Sigma_\pi$ *is a stopping time.*

Proof. For each $u > 0$, $A \cap [\![0,u]\!]$ is a measurable subset of $[0,u]\times\Omega$ given the σ-algebra $\mathcal{B}([0,u])\otimes\mathcal{F}_u$. For each $t \in [0,\infty[$, we define the sets $A_t := A\cap[\![0,t[\![$. Note that A_t is a measurable subset of $([0,t[\times\Omega, \mathcal{B}([0,t[)\otimes\mathcal{F}_t)$.

Because each $\mathcal{F}_t$ is complete, from Theorem 7.3.3, $\pi(A_t) = \{D_A < t\}$ is $\mathcal{F}_t$-measurable. Therefore, we know that D_A is an ($\mathcal{F}$-measurable) random variable. For each $s > t$,

$$\{D_A \le t\} = \bigcap_{n\in\mathbb{N}} \{D_A < t + (s-t)n^{-1}\} \in \mathcal{F}_s,$$

and so $\{D_A \le t\} \in \mathcal{F}_{t+} = \mathcal{F}_t$, as the filtration is right-continuous. □

Remark 7.3.5. This clearly implies that D_A is a stopping time, for $A \in \Sigma_x$, where $x = o, p$. However, it is *not* the case that D_A is predictable for $A \in \Sigma_p$. For example, take $S \in \mathcal{T}_o$ and $A =]\!]S,\infty[\![\in \Sigma_p$. Then $D_A = S$, but S is not necessarily a predictable stopping time. On the other hand, Lemma 7.3.7 will give a result in this direction.

Lemma 7.3.6. *Suppose* T *is a random time with values in* $[0,\infty]$ *(that is, a measurable map* $\Omega \to [0,\infty]$*). Then* $T \in \mathcal{T}_x$ $(x = o,p)$ *if and only if* $[\![T]\!] \in \Sigma_x$*.*

Proof. The necessity is immediate, from Lemma 7.1.7 with $S = T$.

If $[\![T]\!] \in \Sigma_o$ we know from Theorem 7.3.4 that $T \equiv D_{[\![T]\!]} \in \mathcal{T}_o$, giving the result. If $[\![T]\!]$ is a predictable set, then $[\![T, \infty[\![= [\![T]\!] \cup]\!]T, \infty[\![$ is a predictable set, that is, it is in the σ-algebra generated by stochastic intervals of the form $[\![S, \infty[\![$ for $S \in \mathcal{T}_p$. Hence there exists a sequence of predictable stopping times S_n with $S_n \uparrow T$, and so, by Theorem 6.2.12, T is a predictable stopping time. □

Lemma 7.3.7. *If $A \in \Sigma_p$ and $[\![D_A]\!] \subseteq A$, then $D_A \in \mathcal{T}_p$.*

Proof. As $[\![D_A]\!] \subseteq A$ and $[\![0, D_A]\!]$ is predictable, we see that

$$[\![D_A]\!] = A \cap [\![0, D_A]\!]$$

is a predictable set, and hence, by Lemma 7.3.6, D_A is a predictable stopping time. □

Corollary 7.3.8. *Suppose $\{X_t\}_{t\in[0,\infty[}$ is a progressive process with values in the measurable space $(E, \mathcal{E})$. Then for any set $B \in \mathcal{E}$ the random time Z defined by*

$$Z(\omega) = \inf\{t > 0 : X_t(\omega) \in B\}$$

is a stopping time.

Proof. The set $A = \{(t, \omega) : X_t(\omega) \in B\}$ is progressively measurable, as is $C = A \cap]\!]0, \infty[\![$. Then $Z = D_C$ and is thus a stopping time by Theorem 7.3.4. □

Definition 7.3.9. *The stopping time Z defined in Corollary 7.3.8 is called the first hitting time of B by X.*

Remark 7.3.10. When X is càdlàg and $B = [-\infty, c[$, we have already come across this stopping time directly in Remark 3.2.16.

7.3.1 The Section Theorem

We think of the debut D_A of a set A as giving us, for each ω, the infimum of the times t such that $(t, \omega) \in A$. In the same way, we can define a *section* of A, which tells us, for each ω, an *arbitrary* t such that $(t, \omega) \in A$ (Fig. 7.2). Particularly if A is not closed on the left, so $[\![D_A]\!] \not\subseteq A$, this is a useful object to work with.

Definition 7.3.11. *For $A \subset [0, \infty[\times \Omega$ a* section *of A is a map $T : \Omega \to [0, \infty]$ such that $[\![T]\!] \subseteq A$ and $\{T < \infty\} = \{\omega : ([0, \infty[\times\{\omega\}) \cap A \neq \emptyset\}$, that is,*

$$(T(\omega), \omega) \in A \text{ whenever } (t, \omega) \in A \text{ for some } t < \infty.$$

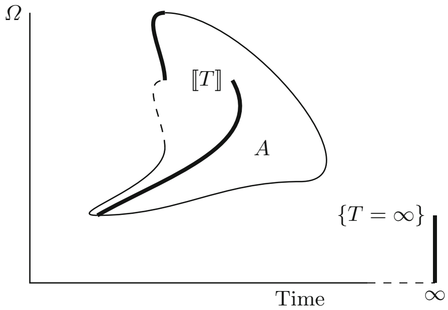

Fig. 7.2. A depiction of a set $A \subset [0,\infty[\times \Omega$, with the graph of a section T (and the set $\{T=\infty\}$) in bold. Note that as A is not closed on the left, the debut of A is not a section of A.

Example 7.3.12. If $A = [\![S, S']\!]$ for S, S' stopping times, then

(i) $T = S$ is a stopping time which is a section of A, as is $T = S'$,
(ii) $T = (S+S')/2$ is a random time which is a section of A (but usually not a stopping time),
(iii) for any $B \subseteq \Omega$, $T = I_B S + I_{B^c} S'$ is a section of A, (but is usually only a random time if $B \in \mathcal{F}$, and only a stopping time if $B \in \mathcal{F}_S$).

We now give a powerful result regarding appropriately *measurable* sections of sets. In particular, we prove (Theorem 7.3.17) that for any optional set A, one can find a *stopping time* S with a graph lying in the set (that is, if $S < \infty$ then $S(\omega) \in A(\omega)$), such that S stops in A whenever $A(\omega)$ is nonempty, except on some set of arbitrarily small probability. That is, S takes values in a section of A with probability $1-\varepsilon$. A similar result holds for predictable sets A, where our stopping time is then also predictable.

This result will allow us to compare the behaviour of optional and predictable processes using the associated class of stopping times.

Definition 7.3.13. *Let $\mathcal{B}_x$ denote the algebra of sets generated by (finite unions, complements and intersections of) stochastic intervals of the form $[\![S, T[\![$, where $S, T \in \mathcal{T}_x$ (for $x = o, p$).*

By construction, we have that $\Sigma_x = \sigma(\mathcal{B}_x)$. Therefore, knowledge of the debuts of sets in $\mathcal{B}_x$ can be used to understand the debuts (or more generally, the sections) of sets in Σ_x.

Lemma 7.3.14. *If $B \in \mathcal{B}_x$ (for $x = o, p$), then B is a finite union of the form*

$$B = [\![S_1, T_1[\![\cup \dots \cup [\![S_m, T_m[\![, \quad S_i, T_i \in \mathcal{T}_x, \quad i = 1, \dots, m.$$

Furthermore, we clearly have $D_B = \bigwedge_i S_i \in \mathcal{T}_x$ and $[\![D_B]\!] \subset B$.

Proof. $\mathcal{B}_x$ is the family of complements and all finite unions of stochastic intervals $[\![S,T[\![$ for $S,T \in \mathcal{T}_x$. However, we can represent complements of intervals $[\![S,T[\![$ using $[\![S,T[\![^c = [\![0,S[\![\cup [\![T,\infty[\![$, which is of the desired form. The stated properties of D_B are trivial. □

Lemma 7.3.15. *Suppose $\{A_n\}_{n\in\mathbb{N}}$ is a sequence of elements of $\mathcal{B}_x$ (for $x = o,p$). Write $A = \bigcap_n A_n$. Then $[\![D_A]\!] \subseteq A$ and $D_A \in \mathcal{T}_x$.*

Proof. We know that $A \in \Sigma_x$ and so D_A is a stopping time. Define $B_n = A_n \cap [\![D_A,\infty[\![$, so that $B_n \in \mathcal{B}_o$ and $D_{B_n} \geq D_A$. However, $\bigcap_n B_n = A \cap [\![D_A,\infty[\![= A$, so $D_{B_n} \leq D_A$, which implies $D_{B_n} = D_A$. As $B_n \in \mathcal{B}_o$, we see $[\![D_A]\!] \subset B_n$ for all n, hence $[\![D_A]\!] \subset \bigcap_n B_n = A$. If $A \in \Sigma_p$, then it follows from Lemma 7.3.7 that D_A is predictable. □

The following "section theorem" is again quoted without proof from Dellacherie [53, Chapter I, Theorem 37] (see also [54, 94] for presentations).

Theorem 7.3.16 (Measurable Section Theorem). *Suppose $(\Omega,\mathcal{F},P)$ is a probability space and that B is an element of the product σ-algebra $\mathcal{B}\otimes\mathcal{F}$ on $[0,\infty[\times\Omega$. Then there is an $\mathcal{F}$-measurable random variable T, with values in $[0,\infty]$, such that*

(i) if $T(\omega) < \infty$ then $(T(\omega),\omega) \in B$, and
(ii) $P(T<\infty) = P(D_B < \infty)$.

That is, T is a section of B, up to an evanescent set.

Using the axiom of choice, it is easy to find an arbitrary section T of a set B. The key result of this theorem is that, when B is measurable, we can choose T to be measurable. It is unsurprising that the random time T is not, in general, a stopping time, as measurability does not involve the filtration.

We extend Theorem 7.3.16 to the σ-algebras Σ_x $(x = o,p)$ in the next result. Note that this does not include progressively measurable sets (see Appendix A.5 for a counterexample).

Theorem 7.3.17 (Optional and Predictable Section Theorem). *Suppose $(\Omega,\mathcal{F},P)$ is a probability space with a filtration $\{\mathcal{F}_t\}_{t\in[0,\infty[}$, and suppose $A \in \Sigma_x$ for $x = o,p$. For any $\varepsilon > 0$ there is a* stopping *time $S \in \mathcal{T}_x$ such that*

(i) $[\![S]\!] \subseteq A$, and
(ii) $P(S<\infty) \geq P(\pi(A)) - \varepsilon$.

Proof. From Theorem 7.3.16 there is an $\mathcal{F}$-measurable random time T such that

(i) if $T(\omega) < \infty$ then $(T(\omega),\omega) \in A$, and
(ii) $P(T<\infty) = P(\pi(A))$.

Define a measure μ on the product σ-algebra $\mathcal{B} \otimes \mathcal{F}$ by

$$\mu(E) = P(\{\omega : (T(\omega), \omega) \in E\} \cap \{T < \infty\}) \quad \text{for} \quad E \in \mathcal{B} \otimes \mathcal{F}.$$

This measure has support A and $\mu(A) = P(\pi(A))$. The Boolean algebra $\mathcal{B}_x$ generates Σ_x so, from Lemma A.1.19, for any $\varepsilon > 0$ there is a sequence $\{B_n\}_{n\in\mathbb{N}} \subset \mathcal{B}_x$ with $B := \bigcap_n B_n \subset A$ and $\mu(A \setminus B) < \varepsilon$. Note that as μ has support A, we have $\mu(B) = \mu(B \cap A)$.

Let $S = D_B$. By Lemma 7.3.15 we know $[\![S]\!] = [\![D_B]\!] \subset B \subset A$ and $S \in \mathcal{T}_x$. Finally, $\{S < \infty\} = \pi(B \cap A)$, and so $P(\{S < \infty\}) = \mu(B \cap A) \geq \mu(A) - \varepsilon$. □

The following corollary to the section theorem provides a way of checking the indistinguishability of stochastic processes.

Corollary 7.3.18. *Suppose X and Y are two Σ_x-measurable processes. Then X and Y are indistinguishable if and only if $X_T = Y_T$ a.s. for every $T \in \mathcal{T}_x$ ($x = o, p$).*

Proof. Necessity of the statement is clear, from the definition of indistinguishability. To show sufficiency, note that the set $A = \{(t, \omega) : X_t(\omega) \neq Y_t(\omega)\}$ is in Σ_x. If $P(\pi(A)) \neq 0$ there is, for any $\varepsilon > 0$, a stopping time $S \in \mathcal{T}_x$ such that $[\![S]\!] \subset A$ and $P(\{S < \infty\}) \geq P(\pi(A)) - \varepsilon$. Therefore, there is a $t \in [0, \infty[$ such that $X_{S\wedge t}$ differs from $Y_{S\wedge t}$ on a set of positive measure, so the processes cannot be indistinguishable. □

Theorem 7.3.19. *Suppose X and Y are two optional (resp. predictable) processes. Then the following two properties are equivalent:*

(i) $X_t(\omega) \geq Y_t(\omega)$, except on an evanescent set,
(ii) for any $T \in \mathcal{T}_o$ (resp. $T \in \mathcal{T}_p$) such that the random variables $X_T I_{\{T<\infty\}}$ and $Y_T I_{\{T<\infty\}}$ are integrable we have

$$E[X_T I_{\{T<\infty\}}] \geq E[Y_T I_{\{T<\infty\}}].$$

Proof. It is clear that (i) implies (ii). Suppose (ii) holds, but $\{X < Y\}$ is not evanescent. Then by Theorem 7.3.17 there is a $T \in \mathcal{T}_x$ ($x = o, p$) such that $P(T < \infty) > 0$ and $X_T I_{\{T<\infty\}} < Y_T I_{\{T<\infty\}}$. There is certainly a constant $\alpha > 0$ such that the measure of the set

$$B = \{T < \infty\} \cap \{|X_T| \leq \alpha\} \cap \{|Y_T| \leq \alpha\}$$

is nonzero. We can see that $B \in \mathcal{F}_T$, and in the predictable case $B \in \mathcal{F}_{T-}$, by Exercise 7.7.5. Therefore, $T_B \in \mathcal{T}_x$ ($x = o, p$) and

$$E[X_{T_B} I_{\{T_B<\infty\}}] < E[Y_{T_B} I_{\{T_B<\infty\}}],$$

contradicting (ii). □

Before concluding this section, we also observe the following result on the local boundedness of processes.

Lemma 7.3.20. *If H is a càdlàg adapted process, then $\{H_{t-}\}_{t\geq 0}$ is locally bounded. If H is a càdlàg predictable process, then H is locally bounded.*

Proof. Write $T_n = \inf\{t : |H_t| \geq n\}$. In the càdlàg adapted case (see also Remark 10.3.2), we clearly have

$$|H_{t-}^{T_n}| \leq n.$$

In the predictable case, each T_n is the debut of a predictable set and $H_{T_n} \geq n$. By Lemma 7.3.7, we see that T_n is a predictable stopping time, and so is announced by a sequence $\{S_{n,m}\}_{m\in\mathbb{N}}$. Writing

$$S_k = \sup_{\substack{n\leq k\\ m\leq k}} S_{n,m},$$

we have $\lim_k S_k = \infty$ a.s. and $|H_t^{S_k}| \leq k$. □

7.4 A Function-Space Monotone Class Theorem

In the sequel, we use the following version of the monotone class theorem (Theorem 1.1.14).

Theorem 7.4.1. *Suppose $\mathcal{H}$ is a real vector space of real-valued, bounded functions defined on Ω, such that*

(i) $\mathcal{H}$ contains the constant functions,
(ii) $\mathcal{H}$ is closed under uniform convergence and
(iii) for every uniformly bounded increasing sequence $\{f_n\}_{n\in\mathbb{N}} \subset \mathcal{H}$ of positive functions the limit $f = \lim_n f_n$ is in $\mathcal{H}$.

Suppose $\mathcal{C}$ is a subset of $\mathcal{H}$ which is closed under multiplication and, as usual, write $\sigma(\mathcal{C})$ for the σ-algebra on Ω generated by $\mathcal{C}$ (i.e. the smallest σ-algebra such that f is measurable for every $f \in \mathcal{C}$). Then $\mathcal{H}$ contains all bounded $\sigma(\mathcal{C})$-measurable functions.

Proof. As elements of $\mathcal{H}$ are bounded and $f/\sup_\omega |f(\omega)|$ is $\sigma(\mathcal{C})$-measurable for any $f \in \mathcal{C}$, we can assume without loss of generality that the functions in $\mathcal{C}$ take values in $[-1, +1]$.

For $N \in \mathbb{N}$, write Λ_N for the set of real-valued functions ϕ on $[-1, +1]^N$ such that $\phi \circ (f_1, f_2, \ldots, f_N) \in \mathcal{H}$ for all sequences $\{f_i\}_{i=1}^N \subset \mathcal{C}$. Then Λ_N contains the constant functions (as all constants are in $\mathcal{H}$) and the polynomial functions (as $\mathcal{C}$ is closed under multiplication), and Λ_N is closed under uniform convergence (also from $\mathcal{H}$). Consequently, by the Stone–Weierstrass approximation theorem (see [160, Chapter 12]), Λ_N contains all the continuous functions on $[-1, +1]^N$. Also, because Λ_N is closed under taking limits of monotone uniformly bounded sequences (by the monotone property), Λ_N contains all the Borel functions on $[-1,+1]^N$. In particular, for any $\{f_i\}_{i=1}^N \subset \mathcal{C}$, any $A \in \sigma(\{f_i\}_{i=1}^N)$ we know $I_A \in \mathcal{H}$.

Let $\mathcal{M} = \{A \in \sigma(\mathcal{C}) : I_A \in \mathcal{H}\}$. We know that

$$\left\{A \in \sigma(\{f_i\}_{i=1}^N) \text{ for some } \{f_i\}_{i=1}^N \subset \mathcal{C}\right\} \subset \mathcal{M},$$

and the left-hand side is an algebra of sets. By monotone convergence in $\mathcal{H}$ we see $\mathcal{M}$ is closed under increasing unions of sets, and as $\mathcal{H}$ is a real vector space, taking $I_{A^c} = 1 - I_A$ we see $\mathcal{M}$ is closed under decreasing intersections of sets. Therefore $\mathcal{M}$ is a monotone class, and so, by Theorem 1.1.14, $\mathcal{M}$ contains $\sigma(\mathcal{C})$.

Finally, we see that as $\mathcal{H}$ is a vector space closed under monotone convergence, an approximation argument shows that $\mathcal{H}$ contains all bounded $\sigma(\mathcal{C})$-measurable functions. □

Remark 7.4.2. A typical example where we apply Theorem 7.4.1 is when $\mathcal{C}$ is taken to be the indicator functions of a (Boolean) algebra B. Such functions are always closed under multiplication, as required. The theorem then states that $\mathcal{H}$ contains all bounded $\sigma(B)$-measurable functions, where $\sigma(B)$ is the smallest σ-algebra containing B.

Corollary 7.4.3. *Let X be a predictable process. There exists a sequence $\{X^n\}_{n\in\mathbb{N}}$ of left-continuous simple processes, that is, processes of the form*

$$X_t^n = X_0 + \sum_{i=1}^k I_{t\in]\!]S_i,T_i]\!]} x_i$$

where, for each i, $x_i \in \mathbb{R}$ and S_i and T_i are stopping times, such that, except possibly on an evanescent set, we have the pointwise convergence

$$X^n \to X.$$

Proof. Let $\mathcal{H}$ be the class of processes with the desired approximation property, and $\mathcal{C}$ the class of simple processes. Clearly $\mathcal{H}$ is closed under uniform convergence and taking increasing limits, while $\mathcal{C}$ is closed under multiplication (as the product of two simple functions is a simple function). As the left-continuous simple processes generate Σ_p (Corollary 7.2.5), we know that $\mathcal{H}$ contains all bounded predictable processes. By considering approximation of $\phi(X)$ for a measurable bijection $\phi : \mathbb{R} \to]-1,1[$ (e.g. $\phi(x) = \frac{\pi}{4}\arctan(x)$), we see that $\mathcal{H}$ must contain all predictable processes. □

7.5 Thin Sets

We now consider sets in $[0,\infty[\times\Omega$ defined by countable collections of stopping times.

Definition 7.5.1. *A set $A \in \Sigma_o$ will be called* thin *if there exists a sequence of stopping times $\{S_n\}_{n\in\mathbb{N}}$ such that $A \subseteq \bigcup_{n\in\mathbb{N}} [\![S_n]\!]$. If moreover $[\![S_n]\!]\cap[\![S_m]\!] = \emptyset$ whenever $n \neq m$, then $\{S_n\}_{n\in\mathbb{N}}$ is called an* exhausting sequence *for A.*

Theorem 7.5.2. *Any thin set admits an exhausting sequence. Moreover if $A \in \Sigma_p$, then there is an exhausting sequence in $\mathcal{T}_p$.*

Proof. Suppose $A \subset \bigcup_n [\![S_n]\!]$. Write T_n for the stopping time whose graph is given by

$$\Big(A \setminus \bigcup_{k<n} [\![S_k]\!]\Big) \cap [\![S_n]\!].$$

Then the T_n have disjoint graphs and $A = \bigcup_n [\![T_n]\!]$. This proves the result in the optional case.

If $A \in \Sigma_p$ then A is in Σ_o since $\Sigma_p \subset \Sigma_o$, so we can apply the above result to say $A = \bigcup_n [\![T_n]\!]$ for some sequence $\{T_n\}_{n\in\mathbb{N}} \subset \mathcal{T}_o$. Let U_n, V_n denote the accessible and totally inaccessible parts of T_n, as constructed in Theorem 6.2.9. Because U_n is accessible, its graph is contained in the union of the graphs of a countable number of predictable stopping times $\{R_{n,m}\}_{m\in\mathbb{N}}$. As V_n is totally inaccessible, $[\![V_n]\!] \cap [\![R_{n,m}]\!] = \emptyset$ for all m, therefore

$$\bigcup_n [\![V_n]\!] = A \setminus \Big(\bigcup_{n,m} [\![R_{n,m}]\!]\Big) \in \Sigma_p.$$

By the section theorem, there exists a predictable stopping time V^* with $[\![V^*]\!] \subseteq \bigcup_n [\![V_n]\!]$; however this is a contradiction with V_n being totally inaccessible, unless $V_n = \infty$ for all n.

Hence

$$A = \bigcup_n [\![T_n]\!] = \bigcup_n [\![U_n]\!] \subset \bigcup_{n,m} [\![R_{n,m}]\!].$$

And we can reorder $\{R_{n,m}\}_{n,m\in\mathbb{N}}$ to give a singly indexed sequence $\{R_n\}_{n\in\mathbb{N}}$, without loss of generality. Similarly to in the optional case, write W_n for the stopping time whose graph is

$$\Big(A \setminus \bigcup_{k<n} [\![R_k]\!]\Big) \cap [\![R_n]\!].$$

Then, from Lemma 7.3.6, W_n is predictable and $A = \bigcup_n [\![W_n]\!]$. By construction, $[\![W_n]\!] \cap [\![W_m]\!] = \emptyset$ if $n \neq m$. □

Remark 7.5.3. While thin sets have nice countability properties, they can also be dense, for example, the set $\mathbb{Q} \times \Omega$ is clearly thin.

We now show that the concepts of optional and predictable processes differ only by the behaviour of the process on a thin set.

Theorem 7.5.4. *Suppose X is an optional process. Then there is a predictable process Y such that $\{X \neq Y\}$ is a thin set.*

Proof. Write $\mathcal{H}$ for the set of optional processes X such that there is a predictable process Y^X with $\{X \neq Y^X\}$ a thin set. Then $\mathcal{H}$ is a vector space, and $\mathcal{H}$ contains all constant processes. Furthermore, $\mathcal{H}$ is closed under pointwise convergence. To see this suppose $\{X^n\}_{n\in\mathbb{N}}$ is a convergent sequence of optional processes in $\mathcal{H}$. For each process X^n, there is a predictable process Y^n such that $\{X^n \neq Y^n\}$ is contained in the union of the graphs of a sequence of stopping times. As X^n converges pointwise, so does Y^n, except possibly on the thin set $\cup_n\{X^n \neq Y^n\}$. If $Y = \limsup_n Y^n$, then we see that $\{X \neq Y\}$ is thin and Y is predictable. Therefore, in particular, $\mathcal{H}$ is closed under uniform convergence and monotone convergence.

Now for any stopping times $S, T \in \mathcal{T}_o$, the process $I_{[\![S,T[\![}$ is in $\mathcal{H}$, because from Theorem 7.1.9, $I_{]\!]S,T[\![}$ is predictable, and these functions differ only on $[\![S]\!]$. Consequently, from Theorem 7.4.1, $\mathcal{H}$ contains all optional processes. □

Remark 7.5.5. A typical example of this is when X is càdlàg and Y is the left-limit of X. See Exercise 7.7.1.

7.6 Optional and Predictable Projections

We now come to the extremely useful technique of projecting a measurable process into the optional or predictable σ-algebra. This technical operation will be shown to be of fundamental importance when constructing the stochastic integral.

Definition 7.6.1. *We say that a measurable process X is* (essentially) bounded *if there exists $k > 0$ such that $\{\omega : \sup_t |X_t(\omega)| > k\}$ is evanescent. We write $B(\mathcal{B}\otimes\mathcal{F})$ for the space of essentially bounded measurable processes, and $B(\Sigma_x)$ $(x = o, p)$ for the space of essentially bounded Σ_x-measurable processes.*

The following projection theorem shows there is a projection map which behaves similarly to the conditional expectation operator. Recall that a map Π is a projection if $\Pi \circ \Pi = \Pi$.

Theorem 7.6.2 (Projection Theorem). *For $x = o, p$, there is a unique linear order-preserving projection, $\Pi_x : B(\mathcal{B}\otimes\mathcal{F}) \to B(\Sigma_x)$, such that for $X \in B(\mathcal{B}\otimes\mathcal{F})$ and for every $T \in \mathcal{T}_x$,*

$$E[X_T I_{\{T<\infty\}}] = E[(\Pi_x X)_T I_{\{T<\infty\}}]. \tag{7.2}$$

The process $\Pi_x X$, $x = o$ (resp. $x = p$), is called the optional *(resp.* predictable*) projection of X.*

Proof. We first show existence, then uniqueness and the order preserving property, then that the map is a projection.

Existence. Write $\mathcal{H}_x$ for the set of all processes in $B(\mathcal{B}\otimes\mathcal{F})$ for which we can find a process $Y \in B(\Sigma_x)$ with $E[X_T I_{\{T<\infty\}}] = E[Y_T I_{\{T<\infty\}}]$. Then $\mathcal{H}$

is a vector space, contains constants and is closed under limits of increasing sequences, so Theorem 7.4.1 holds. Consequently, we need only to prove the existence of $\Pi_x X$ for X in a family of uniformly bounded processes which is closed under multiplication and which generates the σ-algebra $\mathcal{B} \otimes \mathcal{F}$. Such a family is given by processes of the form $X_t(\omega) = Z(\omega) I_{t \in [r,s]}$ for $r < s$ real numbers and Z in $L^\infty(\Omega, \mathcal{F}, P)$. The optional and predictable cases will be treated separately.

Existence: Optional Case. For $X = Z I_{[r,s]}$, let Y be the right-continuous modification of the martingale $E[Z|\mathcal{F}_t]$. By Theorem 5.1.4, Y has left-hand limits almost surely, so by Theorem 7.2.7, Y is optional.

Define $(\Pi_o X) := Y I_{[r,s]}$. Then, for $T \in \mathcal{T}_o$,

$$E[X_T I_{\{T<\infty\}}] = E[Z I_{\{r \le T \le s\}}],$$

and

$$\begin{aligned} E[(\Pi_o X)_T I_{\{T<\infty\}}] &= E[Y_T I_{\{r \le T \le s\}} I_{\{T<\infty\}}] \\ &= E\big[E[Z|\mathcal{F}_T] I_{\{r \le T \le s\}}\big] \\ &= E\big[E[Z I_{\{r \le T \le s\}} \big| \mathcal{F}_T]\big], \end{aligned}$$

where in passage from the first line to the second, we have used Theorem 5.3.1. Consequently, $\Pi_o X$ is an optional process satisfying (7.2).

Existence: Predictable Case. As above, suppose $X = Z I_{[r,s]}$ and write Y for the càdlàg version of the martingale $\{E[Z|\mathcal{F}_t]\}_{t \ge 0}$. Then the process W defined by $W_t = Y_{t-}$ is left-continuous, and so predictable by Theorem 7.1.9. Suppose $T \in \mathcal{T}_p$, so T is announced by an increasing sequence of stopping times $\{T_n\}_{n \in \mathbb{N}}$ and from Lemma 6.2.14, $\mathcal{F}_{T-} = \bigvee_n \mathcal{F}_{T_n}$. By discrete time martingale convergence (Corollary 4.4.5) applied to the uniformly integrable discrete time martingale $\{V_n\}_{n \in \mathbb{N}}$, where $V_n = Y_{T_n} = E[Z|\mathcal{F}_{T_n}]$, we have that $Y_{T-} = \lim_n Y_{T_n} = E[Z|\mathcal{F}_{T-}]$. That is, $W_T = E[Z|\mathcal{F}_{T-}]$ for $T \in \mathcal{T}_p$.

Define $(\Pi_p X) = W I_{[r,s]}$. Then $(\Pi_p X)$ is predictable and, for $T \in \mathcal{T}_p$,

$$E[X_T I_{\{T<\infty\}}] = E[Z I_{\{r \le T \le s\}}]$$

and

$$\begin{aligned} E[(\Pi_p X)_T I_{\{T<\infty\}}] &= E[W_T I_{\{r \le T \le s\}} I_{\{T<\infty\}}] \\ &= E[Y_{T-} I_{\{r \le T \le s\}}] \\ &= E\big[E[Z|\mathcal{F}_{T-}] I_{\{r \le T \le s\}}\big] \\ &= E\big[E[Z I_{\{r \le T \le s\}} | \mathcal{F}_T]\big], \end{aligned}$$

because $I_{\{r \le T \le s\}}$ is $\mathcal{F}_{T-}$-measurable.

Uniqueness and order preservation. First note that for $X, Y \in B(\mathcal{B} \otimes \mathcal{F})$, if $X \le Y$ (except possibly on an evanescent set), then the constructed $\Pi_x X \le \Pi_x Y$. To see this, observe that $A = \{\Pi_x X > \Pi_x Y\} \in \Sigma_x$ and if $P(\pi(A)) \neq 0$, by Theorem 7.3.17, there would be a stopping time $T \in \mathcal{T}_x$ such that $E[(\Pi_x X - \Pi_x Y)_T I_{\{T<\infty\}}] > 0$. So by (7.2), $E[(X - Y)_T I_{\{T<\infty\}}] > 0$, giving a contradiction. Furthermore, if $\Pi_x X$ and $\tilde{\Pi}_x X$ are two projections of the process $X \in B(\mathcal{B} \otimes \mathcal{F})$, both satisfying the above conditions, then for every $T \in \mathcal{T}_x$

$$E[(\Pi_x X)_T I_{\{T<\infty\}}] = E[X_T I_{\{T<\infty\}}] = E[(\tilde{\Pi}_x X)_T I_{\{T<\infty\}}].$$

Therefore, by Theorem 7.3.19, $\Pi_x X = \tilde{\Pi}_x X$ (up to evanescence) and so the projection is unique.

Projection property. Clearly, if $X \in B(\Sigma_x)$ then X is a projection of X. By uniqueness, it follows that $\Pi_x X = X$. □

Remark 7.6.3. From this construction, we note that $(\Pi_p X)_T$ is $\mathcal{F}_{T-}$-measurable for any $T \in \mathcal{T}_p$. By uniqueness of the projection, it follows that X_T is $\mathcal{F}_{T-}$-measurable for any predictable process X and $T \in \mathcal{T}_p$.

Remark 7.6.4. By the order preservation property, if $\{X^n\}_{n\in\mathbb{N}}$, is an increasing sequence of processes, that is $X_t^{n+1}(\omega) \geq X_t^n(\omega)$ except possibly on an evanescent set, then $\{(\Pi_x X^n)\}_{n\in\mathbb{N}}$ is an increasing sequence. Therefore, for a general positive measurable process X, the map $\Pi_x X$ can be defined as the limit of an increasing sequence of processes $\{\Pi_x(X \wedge n)\}_{n\in\mathbb{N}}$. From this, we can use linearity to define $\Pi_x X$ for general X, by writing

$$\Pi_x X = \begin{cases} \Pi_x X^+ - \Pi_x X^- & \text{if } (\Pi_x X^+) \wedge (\Pi_x X^-) < \infty \\ +\infty & \text{otherwise} \end{cases}$$

Considering (7.2) and the monotone convergence theorem, we know $(\Pi_x X)_t$ is integrable whenever X_t is integrable, and (7.2) continues to hold.

The following results show that Π_x behaves much like a conditional expectation. The main reason for the added formality of defining the projection is to ensure that these processes have good measurability properties through time.

Theorem 7.6.5. *Suppose X is a bounded or positive $\mathcal{B} \otimes \mathcal{F}$-measurable process. Then for $T \in \mathcal{T}_o$ such that $X_T I_{\{T<\infty\}}$ is integrable or nonnegative,*

$$(\Pi_o X)_T I_{\{T<\infty\}} = E[X_T I_{\{T<\infty\}} | \mathcal{F}_T],$$

and, if $T \in \mathcal{T}_p$,

$$(\Pi_p X)_T I_{\{T<\infty\}} = E[X_T I_{\{T<\infty\}} | \mathcal{F}_{T-}].$$

Proof. Suppose $A \in \mathcal{F}_T$ in the optional case, and $A \in \mathcal{F}_{T-}$ in the predictable case. If $T \in \mathcal{T}_x$ then $T_A \in \mathcal{T}_x$ (by Lemma 6.2.8 and Theorem 6.3.1). From the projection theorem (Theorem 7.6.2),

$$\begin{aligned} E[X_T I_{\{T<\infty\}} I_A] &= E[X_{T_A} I_{\{T_A<\infty\}}] = E[(\Pi_x X)_{T_A} I_{\{T_A<\infty\}}] \\ &= E[(\Pi_x X)_T I_{\{T<\infty\}} I_A]. \end{aligned}$$

As $(\Pi_x X)_T I_{\{T<\infty\}}$ is $\mathcal{F}_T$-measurable ($\mathcal{F}_{T-}$-measurable in the predictable case), the result follows from uniqueness of the conditional expectation. □

Corollary 7.6.6. *For any càdlàg martingale X, $(\Pi_p X)_t = X_{t-}$ up to indistinguishability.*

Proof. From Theorem 7.6.5, for every t,

$$(\Pi_p X)_t = E[X_t|\mathcal{F}_{t-}] = X_{t-}.$$

As $t \mapsto X_{t-}$ is left-continuous it is predictable, and hence $\Pi_p X = \{X_{t-}\}_{t\geq 0}$ by uniqueness of the projection. □

Corollary 7.6.7. *Suppose X and Y are bounded measurable processes. If $Y \in B(\Sigma_x)$, then $\Pi_x(XY) = (\Pi_x X)Y$.*

Proof. Consider the optional case. As $Y \in B(\Sigma_o)$, we know $Y_T I_{T<\infty}$ is $\mathcal{F}_T$-measurable. From Theorem 7.6.5,

$$Y_T I_{\{T<\infty\}} = (\Pi_o Y)_T I_{\{T<\infty\}} = E[Y_T I_{\{T<\infty\}}|\mathcal{F}_T]$$

for any $T \in \mathcal{T}_o$. Then

$$\begin{aligned}(\Pi_o XY)_T I_{\{T<\infty\}} &= E[X_T Y_T I_{\{T<\infty\}}|\mathcal{F}_T] \\ &= E[X_T I_{\{T<\infty\}}|\mathcal{F}_T] Y_T I_{\{T<\infty\}} \\ &= (\Pi_o X)_T Y_T I_{\{T<\infty\}}\end{aligned}$$

and the result follows by uniqueness. The predictable case is identical, with $\mathcal{F}_T$ replaced by $\mathcal{F}_{T-}$. □

Corollary 7.6.8. *Suppose X is a $\mathcal{B}\otimes\mathcal{F}$-measurable process with X_t integrable or nonnegative for all t. Then the process $\{E[X_t|\mathcal{F}_t]\}_{t\geq 0}$ admits an optional modification, and the process $\{E[X_t|\mathcal{F}_{t-}]\}_{t\geq 0}$ admits a predictable modification.*

Proof. The desired modifications are given by $\Pi_o X$ and $\Pi_p X$ respectively. □

Finally we show the projections differ on only a thin set.

Theorem 7.6.9. *Suppose X is a measurable process with X_t integrable or nonnegative for all t. Then the set $\{\Pi_o X \neq \Pi_p X\}$ is a thin set.*

Proof. Again using Theorem 7.4.1, we need only prove the result for the kind of processes for which the Π_x projections were originally defined. Consider, therefore, the process $X = ZI_{[r,s]}$, where Z is a bounded $\mathcal{F}$-measurable random variable. By construction, the set

$$\{\Pi_o X \neq \Pi_p X\} = \{E[Z|\mathcal{F}_t]I_{[r,s]}(t) \neq E[Z|\mathcal{F}_{t-}]I_{[r,s]}(t)\}$$

is contained in thin set of the jumps of the càdlàg martingale $E[Z|\mathcal{F}_t]$ (see Exercise 7.7.1), and hence is itself a thin set. □

7.7 Exercises

Exercise 7.7.1. Let X be an adapted càdlàg process. Show that the process Y defined by $Y_t := X_{t-}$, the left limits of X, is a predictable process satisfying the requirements of Theorem 7.5.4, in particular that $\{(t,\omega) : \Delta X_t(\omega) \neq 0\}$ is a thin set, where $\Delta X_t = X_t - X_{t-}$.

Exercise 7.7.2. Show that a set $A \subseteq [0,t[\times\Omega$ is progressive if and only if $A \cap [\![0,t]\!] \in \mathcal{B}([0,t]) \otimes \mathcal{F}_t$ for every $t \in [0,\infty[$.

Exercise 7.7.3. Let X be a progressive process (but not necessarily càdlàg). Show, using Theorem 7.4.1 or otherwise, that $\{\sup_{s\leq t}|X_s|\}_{t\geq 0}$ is a progressive process.

Exercise 7.7.4. Let A be a thin set. Show that the function X which counts, for each ω, the number of points in $[0,t]\cap A(\omega)$ is an optional process in $\overline{\mathbb{R}}$.

Exercise 7.7.5. Show that if X is predictable then $X_T I_{\{T<\infty\}}$ is $\mathcal{F}_{T-}$-measurable for every stopping time $T \in \mathcal{T}_o$. (Hint: First consider the case where X is the indicator function of a stochastic interval and use a monotone class argument).

Exercise 7.7.6. Show that if T is a totally inaccessible stopping time, then $\Pi_p I_{[\![T]\!]} = 0$.

Exercise 7.7.7.
(i) Show that if A_n is an increasing sequence of progressive sets $A_1 \subseteq A_2 \subseteq \ldots$, then $D_{A_n} \downarrow D_{\{\cup_n A_n\}}$.
(ii) Give an example of a decreasing sequence of optional sets $A_1 \supseteq A_2 \supseteq \ldots$ such that $\lim_n D_{A_n} \neq D_{\{\cap_n A_n\}}$.

Exercise 7.7.8. For a progressive set A, let $A(\omega) = \{t : (t,\omega) \in A\}$, and let D^n_A denote the 'n-debut' of A, that is,

$$D^n_A(\omega) = \inf\{t \in [0,\infty[: [0,t]\cap A(\omega) \text{ contains at least } n \text{ points}\}.$$

Show that D^n_A is a stopping time, for every $n \in \overline{\mathbb{N}}$.

Exercise 7.7.9. Let X be a càdlàg adapted process with bounded jumps. Show that there exists a sequence of stopping times $\{T_n\}_{n\in\mathbb{N}}$ such that $\sup_t |X^{T_n}_t|$ is almost surely bounded for each n (that is, X is 'locally bounded', see Section 3.3). Give a counterexample when the jumps of X are not bounded.

Part III

Stochastic Integration

8

Processes of Finite Variation

Given our understanding of general stochastic processes, we now set our sights on establishing a theory of stochastic integration. We do this in stages, beginning with the simple case where we take the integral with respect to a process which does not vary 'too much', that is, where its paths are of finite variation for almost all ω. This first step is deceptively simple, as we can establish our integral pathwise, simply by appealing to the Lebesgue–Stieltjes integral considered in Chapter 1. We then use this theory to establish the stochastic integral for more general processes, over the coming chapters.

As previously, we shall work with a probability space $(\Omega, \mathcal{F}, P)$ with a filtration $\{\mathcal{F}_t\}_{t\in[0,\infty[}$ satisfying the usual conditions (completeness and right-continuity), and we define $\mathcal{F}_{0-} = \mathcal{F}_0$ and $\mathcal{F}_{\infty-} = \bigvee_t \mathcal{F}_t$. Unless otherwise stated, all (in)equalities should be read as 'up to evanescence'. We begin by defining the relevant families of processes.

Definition 8.0.1. *A $\mathcal{B} \otimes \mathcal{F}$-measurable stochastic process $\{A_t\}_{t\in[0,\infty[}$ with values in $\mathbb{R}$ is called a* processes of finite variation *if it is càdlàg and, for every $T \in [0,\infty[$,*

$$\sup\Big(\sum_i |A_{t_{i+1}}(\omega) - A_{t_i}(\omega)|\Big) < \infty \quad a.s.$$

where the supremum is taken over the increasing deterministic sequences $\{t_i\}_{i\in\mathbb{N}}$ in $[0,T]$. Write $\mathcal{W}$ for the set of processes of finite variation. For any $A \in \mathcal{W}$, we define $A_{0-}(\omega) \equiv 0$.

Remark 8.0.2. Unlike some authors, we do not require $A_0(\omega) = 0$ a.s. However, unless otherwise indicated, we shall follow the convention that $A_{0-}(\omega) = 0$ a.s. Also note that if A is right-continuous and of finite variation, then it must be càdlàg (Exercise 1.8.17). In particular, if $A, B \in \mathcal{W}$ and $A_t = B_t$ a.s. for all t, then A and B are indistinguishable (Lemma 3.2.10).

S.N. Cohen, R.J. Elliott, *Stochastic Calculus and Applications*,
Probability and Its Applications, DOI 10.1007/978-1-4939-2867-5_8

Definition 8.0.3. *A $\mathcal{B} \otimes \mathcal{F}$-measurable stochastic process $\{A_t\}_{t\in[0,\infty[}$ with values in $[0,\infty[$, is called a* nondecreasing process *if almost every sample path $t \mapsto A_t(\omega)$ is right-continuous and nondecreasing. Write $\mathcal{W}^+$ for the set of nondecreasing processes.*

Remark 8.0.4. As nondecreasing functions have left limits, a right continuous nondecreasing process is càdlàg. Therefore, it is clear that $\mathcal{W}^+ \subset \mathcal{W}$. Furthermore, the $\mathcal{F}$-measurable random variable $A_{\infty-}(\omega) = \lim_{t\to\infty} A_t(\omega)$ exists for an increasing process (but may be infinite).

Remark 8.0.5. We shall see in Lemma 8.1.11 that any process in $\mathcal{W}$ has an essentially unique decomposition into the difference of elements of $\mathcal{W}^+$.

We now restrict our definitions to consider processes adapted to the filtration.

Definition 8.0.6. *We write $\mathcal{V}$ for the family of processes in $\mathcal{W}$ adapted to the filtration $\{\mathcal{F}_t\}_{t\in[0,\infty[}$. We denote by $\mathcal{V}_0$ the set of processes $A \in \mathcal{V}$ such that $A_0 = 0$. Similarly, we define $\mathcal{V}^+$ to be adapted processes in $\mathcal{W}^+$ and define $\mathcal{V}_0^+$ to be those $A \in \mathcal{V}^+$ with $A_0 = 0$.*

From Theorem 7.2.7, any process in V is optional, that is, $\mathcal{V} \subset \mathcal{L}^0(\Sigma_o)$.

Remark 8.0.7. Note that a progressive process in $\mathcal{W}$ is adapted, and hence in $\mathcal{V}$. Also, a process in $\mathcal{V}$ is adapted and càdlàg and is, therefore, optional.

Remark 8.0.8. As a process in $\mathcal{W}$ or $\mathcal{V}$ is only required to be of finite variation on the interval $[0,T]$ for each T, it is clear that $\mathcal{W} = \mathcal{W}_{\text{loc}}$ and $\mathcal{V} = \mathcal{V}_{\text{loc}}$.

Example 8.0.9. Let N be a Poisson process, as in Definition 5.5.16. Then $N \in \mathcal{V}^+$, as N is an increasing adapted process. The martingale X defined by $X_t = N_t - \lambda t$ is in $\mathcal{V}$. Similarly if W is a Brownian motion, then W^* defined by $W_t^* = \sup_{s\le t} |W_s|$ is in $\mathcal{V}^+$. If A is any process in $\mathcal{V}^+$ and ϕ is a nondecreasing function, then $\phi(A) \in \mathcal{V}^+$.

8.1 Integration with Respect to Processes in $\mathcal{W}$

We now define the integral with respect to a finite variation process.

Definition 8.1.1. *Let $A \in \mathcal{W}$. For almost all ω, we define the measure $dA(\omega)$ on $[0,\infty[$ by applying Theorem 1.7.22 to the function of bounded variation $t \mapsto A_t(\omega)$.*

Remark 8.1.2. Under this measure, the 'size' of an interval $]s,t]$ is the stochastic quantity $A_t(\omega) - A_s(\omega)$.

As we have defined the integral for each ω separately, we describe this integral as 'pathwise'. This definition is useful conceptually, but we need to ensure that it does not cause measurability problems.

Lemma 8.1.3. *Consider $A \in \mathcal{W}$ and X a real-valued $\mathcal{B} \otimes \mathcal{F}$-measurable process. For $t \in [0, \infty[$, consider the process defined pathwise by the Stieltjes integral*

$$(X \bullet A)_t(\omega) := \int_{[0,t]} X_s(\omega)\, dA_s(\omega)$$

whenever it exists. (The integral is permitted to take the values $+\infty$ or $-\infty$.) Then

(i) $(X \bullet A)_0 = X_0 A_0$ *and* $(X \bullet A)$ *is càdlàg,*
(ii) $(\lambda X + \tilde{X}) \bullet A = \lambda(X \bullet A) + \tilde{X} \bullet A$ *(for any random variable λ and any $X, \tilde{X}$, provided all terms are well defined), that is, the integral is linear with respect to the integrand,*
(iii) if $(X \bullet A)_t(\omega)$ exists for all t and almost all ω, then the process $X \bullet A$ is $\mathcal{B} \otimes \mathcal{F}$-measurable and
(iv) if A and X are both Σ_x-measurable ($x = o, p$) and the integrals exist, then $X \bullet A$ is Σ_x-measurable.

Proof. That $(X \bullet A)_0 = X_0 A_0$ is trivial, as $A_{0-} = 0$, so $A_0 = \Delta A_0$. That the integral is càdlàg follows from the definition of the integral with respect to a Stieltjes measure (Lemma 1.3.41). The linearity of (ii) follows directly from the linearity of the Lebesgue–Stieltjes integral.

To show (iii) and (iv), we shall prove $\mathcal{B} \otimes \mathcal{F}$-measurability using a monotone class argument, the Σ_x cases are similar. Let $\mathcal{H}$ be the class of bounded functions X such that $X \bullet A$ is $\mathcal{B} \otimes \mathcal{F}$-measurable. Then $\mathcal{H}$ contains the constant functions (as these integrate to multiples of A), and by the dominated convergence theorems, $\mathcal{H}$ is closed under uniform convergence and for uniformly bounded increasing sequences. Therefore, $\mathcal{H}$ satisfies the assumptions of Theorem 7.4.1. Therefore, we only need to show that $\mathcal{H}$ contains a collection $\mathcal{C}$ of functions generating $\mathcal{B} \otimes \mathcal{F}$ where $\mathcal{C}$ is closed under multiplication.

Define[1] $\mathcal{C} = \{I_B I_{\{t>s\}} : s \in [0, \infty[\cup \{0-\} \text{ and } B \in \mathcal{F}\}$. Then for any $X \in \mathcal{C}$, we know that for some $B \in \mathcal{F}$, $s < t$ we have $(X \bullet A)_t = I_B(A_t - A_{t \wedge s})$, so $X \bullet A$ is $\mathcal{B} \otimes \mathcal{F}$-measurable. Therefore $\mathcal{H}$ contains $\mathcal{C}$. As $\mathcal{C}$ generates $\mathcal{B} \otimes \mathcal{F}$, we know that $\mathcal{H}$ contains all bounded $\mathcal{B} \otimes \mathcal{F}$-measurable functions. By linearity of the integral and monotone convergence, this implies the desired measurability of $X \bullet A$ for all X such that the integral exists. □

Remark 8.1.4. It is worth pointing out the importance of this result that the pathwise integral of an optional finite variation process with respect to an optional process is itself an optional process. In particular, the pathwise integral is still adapted to the filtration.

Remark 8.1.5. From this definition, it is clear that, for any fixed random variable X and any $A \in \mathcal{W}$, as X does not vary through time, we have the identity $(X \bullet A)_t = \int_{[0,t]} X dA_s = X A_t$.

[1] We formally allow $s =$ "$0-$", so as to include the case $X_t = I_B$ without additional notation. For consistency, $A_{t \wedge s} := A_{0-} = 0$ whenever $s = 0-$.

Lemma 8.1.6. *If $A, \tilde{A} \in \mathcal{W}$ and λ is a random variable, then $X \bullet (\lambda A + \tilde{A}) = \lambda(X \bullet A) + (X \bullet \tilde{A})$ for all $\mathcal{B} \otimes \mathcal{F}$-measurable X, that is, the integral is linear with respect to the integrator.*

Proof. As in the Proof of Lemma 8.1.3, we use a monotone class argument. Let $\mathcal{H}$ denote those bounded functions for which the result holds. Then $\mathcal{H}$ contains the constants, and is closed under uniform convergence and for uniformly bounded monotone increasing sequences, by the dominated convergence theorem. As in the proof of Lemma 8.1.3, let $\mathcal{C} = \{I_B I_{\{t>s\}} : s \in [0, \infty[\cup \{0-\}$ and $B \in \mathcal{F}\}$, then for any $X \in \mathcal{C}$ and any t we have

$$\begin{aligned}(X \bullet (\lambda A + \tilde{A}))_t &= I_B(\lambda A_t - \lambda A_{t\wedge s} + \tilde{A}_t - \tilde{A}_{t\wedge s}) \\ &= \lambda I_B(A_t - A_{t\wedge s}) + I_B(\tilde{A}_t - \tilde{A}_{t\wedge s}) = \lambda (X \bullet A)_t + (X \bullet \tilde{A})_t.\end{aligned}$$

The Monotone Class Theorem (Theorem 7.4.1) yields the result. □

The following lemma shows us that if A is adapted, then a Fubini-type result holds, allowing us to interchange the order of the conditional expectation and stochastic integral.

Lemma 8.1.7. *If $A \in \mathcal{V}^+$ and X is $\mathcal{B} \otimes \mathcal{F}$-measurable, then for every $t \in [0, \infty[$ and almost all ω, if all terms exist we have*

$$E[(X \bullet A)_t | \mathcal{F}_t] = E\Big[\int_{[0,t]} X_s \, dA_s \Big| \mathcal{F}_t\Big] = \int_{[0,t]} E[X_s|\mathcal{F}_t] \, dA_s = (E[X_{(\cdot)}|\mathcal{F}_t] \bullet A)_t.$$

Here we implicitly take the measurable versions of all conditional expectations.

Proof. Again, we can use a monotone class argument. Let $\mathcal{H}$ denote those bounded functions for which the result holds. Then $\mathcal{H}$ contains the constants, and is closed under uniform convergence and for uniformly bounded monotone increasing sequences, by the dominated convergence theorem. As in the Proof of Lemma 8.1.3, let $\mathcal{C} = \{I_B I_{\{t>s\}} : s \in [0, \infty[\cup \{0-\}$ and $B \in \mathcal{F}\}$ so for any $X \in \mathcal{C}$ we have

$$\begin{aligned}E[(X \bullet A)_t | \mathcal{F}_t] &= E[I_B(A_t - A_{t\wedge s})|\mathcal{F}_t] = E[I_B|\mathcal{F}_t](A_t - A_{t\wedge s}) \\ &= E[I_B|\mathcal{F}_t] \int_{[0,t]} I_{]s,\infty]} dA_s = \int_{[0,t]} E[I_B I_{]s,\infty]}|\mathcal{F}_t] dA_s \\ &= (E[X_{(\cdot)}|\mathcal{F}_t] \bullet A)_t.\end{aligned}$$

Therefore, $\mathcal{H}$ contains $\mathcal{C}$ and so contains all bounded $\mathcal{B} \otimes \mathcal{F}$-measurable functions. By linearity, this extends to all $\mathcal{B} \otimes \mathcal{F}$-measurable functions whenever the integrals exist. □

Remark 8.1.8. After we establish Lemma 8.1.11, we can extend this result by linearity (Lemma 8.1.6) to $A \in \mathcal{V}$, provided all terms exist.

8.1.1 Total Variation Process

It is useful to have a notion of the 'total variation' of a process, in a similar way to when dealing with signed measures (as in Section 1.7).

Definition 8.1.9. *If $A \in \mathcal{W}$, then the process $D \in \mathcal{W}^+$ defined by*

$$D_t(\omega) = \int_{[0,t]} |dA_s(\omega)|$$

is called the total variation *of A. This process D is the unique process in $\mathcal{W}^+$ such that, for almost every ω, the measure $dD_t(\omega)$ on $[0, \infty[$ is the total variation of the signed measure $dA_t(\omega)$ (in the sense of Definition 1.7.7).*

Remark 8.1.10. As, for almost all ω,

$$D_t(\omega) = |A_0(\omega)| + \lim_{n\to\infty} \sum_{k=1}^{n} |A_{tk/n}(\omega) - A_{t(k-1)/n}(\omega)|,$$

it is clear that if A is Σ_x-measurable ($x = o, p$), then so is D.

The following lemma is a version of the Hahn Decomposition (Lemma 1.7.4) for processes. One could also obtain this result by applying Lemma 1.7.21 pathwise.

Lemma 8.1.11. *If $A \in \mathcal{W}$, there is a unique decomposition $A = B - C$, where $B, C \in \mathcal{W}^+$ and $B_t + C_t = D_t = \int_{[0,t]} |dA_s|$.*

If A is an optional (resp. predictable) process then B and C are optional (resp. predictable).

Proof. Write $B_t = (A_t + D_t)/2$ and $C_t = (D_t - A_t)/2$. Then B and C are increasing processes, and are the unique processes satisfying $A = B - C$ and $D = B + C$. The optional and predictable statements follow as soon as A, and hence D, B and C are in Σ_x. □

Remark 8.1.12. Because of this lemma, we can write $\mathcal{W} = \mathcal{W}^+ - \mathcal{W}^+$ and $\mathcal{V} = \mathcal{V}^+ - \mathcal{V}^+$. Frequently, this will allow us to prove a property holds for $\mathcal{W}^+$, and then claim that the property holds for all of $\mathcal{W}$ by linearity (as mentioned above, extending Lemma 8.1.7 to allow integrators $A \in \mathcal{V}$ can be done in this way).

We make the following definition.

Definition 8.1.13. *We denote by $\mathcal{A}^+$ the set of* adapted, integrable, increasing *processes, that is,*

$$\mathcal{A}^+ = \big\{A \in \mathcal{V}^+ : E[A_\infty] < \infty\big\}.$$

Similarly, we denote by $\mathcal{A}$ the set of adapted processes of integrable variation*, that is, the set of processes*

$$\mathcal{A} = \Big\{ A \in \mathcal{V} : E[D_\infty] = E\Big[\int_{[0,\infty[} |dA_s| \Big] < \infty \Big\}$$
$$= \Big\{ A \in \mathcal{V} : D = \int_{[0,\cdot]} |dA_s| \in \mathcal{A}^+ \Big\}.$$

As before, $\mathcal{A}_0^+$ and $\mathcal{A}_0$ denote processes, respectively in $\mathcal{A}^+$ and $\mathcal{A}$, for which $A_0 = 0$.

Example 8.1.14. For N a Poisson process, $N \notin \mathcal{A}^+$, as N_∞ is not integrable. For any fixed time $T < \infty$, we know that $N^T \in \mathcal{A}_0^+ \subset \mathcal{V}_0^+$, as it is an increasing *integrable* process. If $X_t = N_t - \lambda t$, then $X^T \in \mathcal{A}$. The process $\exp((N^T)^2) \in \mathcal{V}^+$ is not in $\mathcal{A}^+$ for any deterministic T (as it has infinite expectation).

We now consider the measure on $\mathcal{B} \otimes \mathcal{F}$ induced by a process $A \in \mathcal{A}^+$.

Definition 8.1.15. *If $A \in \mathcal{A}^+$ (or, more generally, $\mathcal{W}^+$), then a nonnegative measure μ_A can be defined on $([0,\infty[\times \Omega, \mathcal{B} \otimes \mathcal{F})$ by*

$$\mu_A(C) = E[(I_C \bullet A)_\infty] = E\Big[\int_{[0,\infty[} I_C \, dA_s \Big]$$

for each set $C \in \mathcal{B} \otimes \mathcal{F}$. We call μ_A the Doléans measure *associated with A.*

Lemma 8.1.16. *For any $A \in \mathcal{W}^+$ and $C \in \mathcal{B} \otimes \mathcal{F}$, we have*

$$\begin{aligned}\mu_A(C) &= E[(I_C \bullet A)_\infty] \\ &\leq (I_{[0,\infty[\times\pi(C)} \bullet A)_\infty = E[I_{\pi(C)} A_\infty] = \mu_A(I_{[0,\infty[\times\pi(C)}),\end{aligned}$$

where π is the canonical projection $[0,\infty[\times \Omega \to \Omega$, as in Definition 7.3.3.

Proof. As $C \subseteq [0,\infty[\times \pi(C)$ the result follows by nonnegativity of the measure μ_A. □

We have the following characterization for finite measures of this form.

Theorem 8.1.17. *Suppose μ is a measure on $([0,\infty[\times \Omega, \mathcal{B} \otimes \mathcal{F})$. For μ to be of the form μ_A, where $A \in \mathcal{A}^+$, it is necessary and sufficient that*

(i) μ has finite mass,
(ii) the evanescent sets have μ-measure zero, and
(iii) for every $t \in [0,\infty[$ and $H \in \mathcal{F}$,

$$\mu\big([0,t] \times H\big) = \int_{[0,\infty[\times\Omega} E[I_H|\mathcal{F}_t] I_{[0,t]} d\mu,$$

where we take the càdlàg version of the conditional expectation.

In this case, A is unique up to indistinguishability.

Proof. Necessity. Suppose $A \in \mathcal{A}^+$ and $\mu = \mu_A$. Then $\mu([0,\infty[\times\Omega) = E[A_\infty] < \infty$. If $B \subset [0,\infty[\times\Omega$ is evanescent, then by Lemma 8.1.16, $\mu(B) \leq E[I_{\pi(B)}A_\infty] = 0$. Finally, as A is adapted, by Lemma 8.1.7 and Remark 8.1.5 we have, for $t \in [0,\infty[$ and $H \in \mathcal{F}$,

$$\begin{aligned}\mu\big([0,t]\times H\big) &= E[I_H A_t] = E\big[E[I_H A_t|\mathcal{F}_t]\big] = E\big[E[I_H|\mathcal{F}_t]A_t\big] \\ &= \int_{[0,\infty[\times\Omega} E[I_H|\mathcal{F}_t]I_{[0,t]}d\mu_A.\end{aligned}$$

Sufficiency. Given μ satisfying (i) and (ii), for each $t \in [0,\infty[$ define a measure m_t on $(\Omega,\mathcal{F})$ by $m_t(H) = \mu([0,t]\times H)$. From property (ii), m_t is absolutely continuous with respect to the underlying measure P on $(\Omega,\mathcal{F})$.

Write $\hat{A}_t$ for (a version of) the Radon–Nikodym derivative dm_t/dP. As μ is nonnegative, $\hat{A}_t(\omega)$ is nondecreasing in t. We can therefore write $A_t = \inf_{r>t}\hat{A}_r = \hat{A}_{t+}$, to obtain a right-continuous nondecreasing process A_t. As μ has finite mass, m_t is uniformly bounded, and hence $E[A_t]$ is uniformly bounded in t, so $A \in \mathcal{A}^+$. For any sequence $t_n \downarrow t$, by the dominated convergence theorem,

$$\mu([0,t]\times H) = \lim_n \mu([0,t_n]\times H) = \lim_n m_{t_n}(H) = \lim_n E[I_H\hat{A}_{t_n}] = E[I_H A_t],$$

so we see that A_t is also a version of dm_t/dP. By construction, for $t \in [0,\infty[$ and $H \in \mathcal{F}$,

$$\mu_A([0,t]\times H) = E[(I_H \bullet A)_t] = E[I_H A_t] = \mu([0,t]\times H).$$

As sets of the form $[0,t]\times H$ generate the σ-algebra $\mathcal{B}\otimes\mathcal{F}$ and $\mu(I_{[0,\infty[\times\Omega}) = E[A_\infty] < \infty$, we have $\mu = \mu_A$ (by Lemma A.1.18).

We must show that if $\mu = \mu_A$ then A is adapted. For $t \in [0,\infty[$ and $H \in \mathcal{F}$,

$$\begin{aligned}E[I_H A_t] = \mu_A([0,t]\times H) &= \int_{[0,\infty[\times\Omega} E[I_H|\mathcal{F}_t]I_{[0,t]}d\mu_A \\ &= E\big[E[I_H|\mathcal{F}_t]A_t\big] = E\big[E[I_H|\mathcal{F}_t]E[A_t|\mathcal{F}_t]\big] \\ &= E\Big[E\big[E[A_t|\mathcal{F}_t]I_H\big|\mathcal{F}_t\big]\Big] = E\big[E[A_t|\mathcal{F}_t]I_H\big].\end{aligned}$$

That is, $E[I_H A_t] = E\big[I_H E[A_t|\mathcal{F}_t]\big]$ for all $H \in \mathcal{F}$, so $A_t = E[A_t|\mathcal{F}_t]$ almost surely, which implies A is adapted.

Uniqueness. If $B \in \mathcal{A}^+$ is another such process, then for each t, B_t is also a version of dm_t/dP, so $A_t = B_t$ a.s. However, A and B are both right-continuous, therefore they are indistinguishable by Lemma 3.2.10. $\square$

Corollary 8.1.18. *If $A, B \in \mathcal{A}^+$ are such that $E[(I_C \bullet A)_\infty] = E[(I_C \bullet B)_\infty]$ for all $C \in \mathcal{B}\otimes\mathcal{F}$, then $A = B$ up to indistinguishability.*

8.1.2 Stochastic Integrals and Stopping Times

We now focus on the interaction between the integral and stopping times, for processes $A \in \mathcal{V}$.

Definition 8.1.19. *Suppose $A \in \mathcal{V}^+$. For $t \in [0, \infty[$ and $\omega \in \Omega$ define*

$$C_t(\omega) = \inf\{s : A_s(\omega) > t\}.$$

By Remark 3.2.16, or as the debut of a progressive set, C_t is a stopping time. The process C is called the time change *associated with A.*

This name is motivated by the following version of Theorem 1.3.45.

Lemma 8.1.20. *For every process X with Borel-measurable paths (in particular, for any $\mathcal{B} \otimes \mathcal{F}$-measurable X),*

$$\int_{[0,\infty[} X_t \, dA(t) = \int_{[0,A_{\infty-}]} X_{C(t)} \, dt = \int_{[0,\infty]} X_{C(t)} I_{\{C_t < \infty\}} \, dt$$

whenever the integral is defined.

Theorem 8.1.21. *Suppose X and Y are two $\mathcal{B} \otimes \mathcal{F}$-measurable processes such that, for every stopping time T,*

$$E[X_T I_{\{T<\infty\}}] = E[Y_T I_{\{T<\infty\}}].$$

Then, for every process $A \in \mathcal{V}$ and every stopping time T,

$$E\Big[\int_{[0,T]} X_s \, dA_s\Big] = E\Big[\int_{[0,T]} Y_s \, dA_s\Big]$$

whenever the final terms are defined.

Proof. By linearity and Lemma 8.1.11, we can assume without loss of generality that $A \in \mathcal{V}^+$. Define $A_t^T := A_{t \wedge T}$, and write $\{C_t\}_{t \in [0,\infty[}$ for the time change associated with $\{A_t^T\}_{t\in[0,\infty[}$. By Lemma 8.1.20 and Fubini's theorem,

$$\begin{aligned} E\Big[\int_{[0,T]} X_s dA_s\Big] &= E\Big[\int_{[0,\infty[} X_s dA_s^T\Big] = E\Big[\int_{[0,\infty[} X_{C_s} I_{\{C_s<\infty\}} ds\Big] \\ &= \int_{[0,\infty[} E[X_{C_s} I_{\{C_s<\infty\}}] ds. \end{aligned}$$

Similarly,

$$E\Big[\int_{[0,T]} Y_s dA_s\Big] = \int_{[0,\infty[} E[Y_{C_s} I_{\{C_s<\infty\}}] ds.$$

As C_s is a stopping time for all s, the result is proven. □

Corollary 8.1.22. *Suppose $A \in \mathcal{V}$ and M is a uniformly integrable martingale. Then, for every stopping time T such that the terms are defined,*

$$E\Big[\int_{[0,T]} M_s\, dA_s\Big] = E[M_T A_T].$$

Proof. Write $X = MI_{[\![0,T]\!]}$ and $Y = M_T I_{[\![0,T]\!]}$. For any stopping time S,

$$E[X_S I_{\{S<\infty\}}] = E[M_S I_{[\![0,T]\!]}(S) I_{\{S<\infty\}}] = E[M_S I_{\{S\leq T\}}(S) I_{\{S<\infty\}}]$$

and by the optional stopping theorem (Theorem 5.3.1),

$$E[Y_S I_{\{S<\infty\}}] = E\big[E[M_T I_{\{S\leq T\}}(S) I_{\{S<\infty\}}|\mathcal{F}_S]\big] = E[M_S I_{\{S\leq T\}}(S) I_{\{S<\infty\}}].$$

Consequently, by Theorem 8.1.21 and Remark 8.1.5,

$$\begin{aligned} E\Big[\int_{[0,T]} M_s dA_s\Big] &= E\Big[\int_{[0,T]} X_s dA_s\Big] = E\Big[\int_{[0,T]} Y_s dA_s\Big] \\ &= E\Big[\int_{[0,T]} M_T dA_s\Big] = E[M_T A_T]. \end{aligned}$$

□

8.1.3 Decomposition of Finite Variation Processes

In Chapter 1 we have seen that, to every increasing function $\alpha(t)$ defined on $[0,\infty[$, there corresponds a measure $d\alpha(t)$. In the terminology of measure theory, α is continuous if and only if $d\alpha$ is a *diffuse* measure (that is, the support of $d\alpha$ has no atoms). If $d\alpha$ is purely atomic (that is, the support of $d\alpha$ consists of a countable number of atoms), then α is said to be *purely discontinuous.* Every σ-finite measure on $[0,\infty[$ can be written in a unique way as the sum of a diffuse and a purely atomic measure. This corresponds to the unique decomposition of an increasing function into the sum of a continuous and purely discontinuous function. For finite variation processes we have the following analogous result.

Theorem 8.1.23. *Suppose $A \in \mathcal{W}$, that is, A is a càdlàg process of finite variation. Then there is a unique decomposition $A = A^c + A^d$, where $A^c \in \mathcal{W}$ is a continuous process and $A^d \in \mathcal{W}$ is a purely discontinuous process (i.e. the paths of A^d are purely discontinuous). Furthermore,*

- *if $A \in \mathcal{W}^+$, that is, A is nondecreasing, then so are A^c and A^d,*
- *if A is adapted then A is optional, A^d is optional, and A^c is predictable,*
- *if A is predictable then both A^d and A^c are predictable.*

Proof. By Lemma 8.1.11, we can assume that $A \in \mathcal{W}^+$. As A is càdlàg, it has at most countably many jumps for almost all ω, so we write

$$A_t^d(\omega) = \sum_{r \le t} \Delta A_r(\omega) = \sum_{r \le t} \big(A_r(\omega) - A_{r-}(\omega)\big), \tag{8.1}$$

which is a pure jump process (i.e. it is either constant or discontinuous at every point), is right-continuous and is measurable. This then gives $A^c := A - A^d$, which is continuous. This decomposition is clearly unique.

If A is nondecreasing, its jumps are nonnegative, so A^d is nondecreasing. Furthermore, as $A_t - A_s \ge A_t^d - A_s^d$ for all $s < t$, we see that A^c is nondecreasing.

If A is adapted, it is optional by Theorem 7.2.7. As the left limit $\{A_{t-}\}_{t\ge 0}$ is then predictable (Theorem 7.2.4) and the sum in (8.1) is over a countable set of stopping times (Exercise 7.7.1), we see that A^d is optional, and is predictable if A is predictable. As A^c is then adapted and continuous, it follows that A^c is predictable. □

The following result is stated for increasing processes; however its extension to general finite variation processes is straightforward (by Lemma 8.1.11).

Theorem 8.1.24. *Suppose $A \in \mathcal{V}^+$ is a purely discontinuous Σ_x-measurable process ($x = o, p$). Then there is a sequence $\{\alpha_n\}_{n\in\mathbb{N}}$ of positive numbers and a sequence $\{T_n\}_{n\in\mathbb{N}}$ of stopping times $T_n \in \mathcal{T}_x$ such that $A_t = \sum_{n\in\mathbb{N}} \alpha_n I_{[\![T_n,\infty[\![}(t)$.*

Proof. Consider the predictable case; the optional case is similar. By Exercise 7.7.1, there is a sequence $\{S_n\}_{n\in\mathbb{N}}$ of predictable stopping times which exhausts the jumps of A. Therefore

$$A_t = \sum_{n\in\mathbb{N}} (A_{S_n} - A_{S_n-}) I_{[\![S_n,\infty[\![},$$

and it is enough to prove the result when the process is of the form $X = (A_S - A_{S-}) I_{[\![S,\infty[\![}$, where S is a predictable stopping time. By Exercise 7.7.5, the random variable $(A_S - A_{S-})$ is $\mathcal{F}_{S-}$-measurable, and so there is an increasing sequence $\{X_n\}_{n\in\mathbb{N}}$ of simple $\mathcal{F}_{S-}$-measurable functions such that $X = \lim_n X_n$ a.s. Consequently, there is a sequence $\{\alpha_n\}_{n\in\mathbb{N}}$ of positive numbers and a sequence $\{H_n\}_{n\in\mathbb{N}}$ of elements of $\mathcal{F}_{S-}$ such that

$$X = \sum_{n\in\mathbb{N}} \alpha_n I_{H_n}.$$

Restricting S to these sets H_n, we write $T_n = S_{H_n}$, so that T_n is a predictable stopping time. Hence, we see that $A_t = \sum_n \alpha_n I_{[\![T_n,\infty[\![}$. □

8.2 The Projection Π_x and Dual Projection Π_x^*

In Lemma 8.1.7 we have seen that, when $A \in \mathcal{V}$, the stochastic integral and the conditional expectation commute. In Chapter 7, we saw that the projection

Π_x behaves in a similar manner to the conditional expectation. This motivates the more general study of the interaction of the stochastic integral (for $A \in \mathcal{W}$) and the projection Π_x. Furthermore, as this is a linear operator, we shall see that the adjoint operator Π_x^* also has interesting properties.

8.2.1 Integrals and the Projection

Remark 8.2.1. It is a consequence of the following theorem that, if a process $A \in \mathcal{V}^+$ is Σ_x-measurable, then the corresponding signed measure μ_A is determined by its restriction to Σ_x. (Simply take X and Y to be indicator functions.)

Theorem 8.2.2. *Suppose $A \in \mathcal{V}^+$ is Σ_x-measurable ($x = o, p$). If X and Y are two $\mathcal{B} \otimes \mathcal{F}$-measurable processes whose projections $\Pi_x X$ and $\Pi_x Y$ are indistinguishable, then $E[(X \bullet A)_\infty] = E[(Y \bullet A)_\infty]$ (whenever these terms are defined).*

Proof. It is enough to consider the case where X and Y are bounded, the general result then follows by linearity and monotone convergence.

Optional case. By uniqueness, $\Pi_o X = \Pi_o Y$ if and only if $E[X_T I_{\{T<\infty\}}] = E[Y_T I_{\{T<\infty\}}]$ for every $T \in \mathcal{T}_o$. Therefore, because A is adapted, the result follows from Theorem 8.1.21.

Predictable case. As Π_p factors through Π_o, we can suppose X and Y are optional. By the decomposition of Theorem 8.1.23, A can be written

$$A_t = A_t^c + \sum_{n \in \mathbb{N}} \alpha_n I_{[\![T_n, \infty[\![}(t)$$

where $\{T_n\}_{n \in \mathbb{N}} \subset \mathcal{T}_p$ is a sequence of predictable stopping times which exhausts the jumps of A. It is, therefore, enough to establish the result for each term in this decomposition. The sets $\{X \neq \Pi_p X\}$ and $\{Y \neq \Pi_p Y\}$ are thin sets, by Theorem 7.6.9. Therefore, as $\Pi_p X = \Pi_p Y$ up to evanescence, the set $\{X \neq Y\}$ is also a thin set. As the continuous part A^c does not charge any thin set, we have

$$E[(X \bullet A^c)_\infty] = E[(Y \bullet A^c)_\infty].$$

For the discontinuous part, let $B^n := I_{[\![T_n, \infty[\![}$ for $T_n \in \mathcal{T}_p$. Then, as $\Pi_p X = \Pi_p Y$, by Theorem 7.6.2 we have

$$E[(X \bullet B^n)_\infty] = E[X_{T_n} I_{\{T_n<\infty\}}] = E[Y_{T_n} I_{\{T_n<\infty\}}] = E[(Y \bullet B^n)_\infty].$$

By linearity, the result follows. □

Corollary 8.2.3. *Suppose $A \in \mathcal{V}^+$ is Σ_x-measurable ($x = o, p$). If X and Y are two $\mathcal{B} \otimes \mathcal{F}$-measurable processes such that $\Pi_x X = \Pi_x Y$, then, for every pair of stopping times $S \leq T$, provided all terms are well defined, we have*

$$E\Big[\int_{]\!]S,T]\!]} X_t \, dA_t \Big| \mathcal{F}_S\Big] = E\Big[\int_{]\!]S,T]\!]} Y_t \, dA_t \Big| \mathcal{F}_S\Big].$$

Proof. For any $H \in \mathcal{F}_S$,

$$E\Big[I_H \int_{]\!]S,T]\!]} X_t \, dA_t\Big] = E\big[\big((I_{]\!]S_H,T_H]\!]}X)\bullet A\big)_\infty\big]$$

and similarly for Y. As $I_{]\!]S_H,T_H]\!]}$ is predictable, by Corollary 7.6.7,

$$\Pi_x(I_{]\!]S_H,T_H]\!]}X) = I_{]\!]S_H,T_H]\!]}\Pi_x X = I_{]\!]S_H,T_H]\!]}\Pi_x Y.$$

Therefore, the two expectations are equal, by Theorem 8.2.2. □

Corollary 8.2.4. *For any bounded or nonnegative $\mathcal{B} \otimes \mathcal{F}$-measurable process X, any Σ_x-measurable $A \in \mathcal{V}^+$ and any stopping times $S \leq T$, we have $E[(X\bullet A)_\infty] = E[((\Pi_x X)\bullet A)_\infty]$ and*

$$E\Big[\int_{]\!]S,T]\!]} X_t \, dA_t \Big| \mathcal{F}_S\Big] = E\Big[\int_{]\!]S,T]\!]} (\Pi_x X)_t \, dA_t \Big| \mathcal{F}_S\Big].$$

Theorem 8.2.5. *Suppose that μ_A is the measure generated by the increasing process $A \in \mathcal{W}^+$. Then A is Σ_x-measurable ($x = o, p$) if and only if the integral satisfies the following condition:*

> *"If X and Y are two bounded $\mathcal{B} \otimes \mathcal{F}$-measurable processes such that $\Pi_x X = \Pi_x Y$, then $E[(X\bullet A)_\infty)] = E[(Y\bullet A)_\infty]$."*

Proof. The necessity of the condition is established by Theorem 8.2.2. We need to show sufficiency.

Optional case. As A is càdlàg, we need only show it is adapted. As in Theorem 8.1.17, this is the case if for all $H \in \mathcal{F}$ and $t \in [0, \infty[$ we have

$$E[I_H A_t] = E\big[E[I_H|\mathcal{F}_t]A_t\big].$$

However, the processes $X_s = I_H(\omega)I_{[0,t]}(s)$ and $Y_s = E[I_H|\mathcal{F}_s]I_{[0,t]}(s)$ have the same Π_o projection, namely Y_t (where $E[I_H|\mathcal{F}_s]$ is taken to be càdlàg), and so the result follows from the assumed condition.

Predictable case. As A satisfies the condition when $x = p$, then it certainly satisfies the condition when $x = o$, and so A is optional by the above argument. For every $H \in \mathcal{F}$, and $T \in \mathcal{T}_p$, consider the processes $X_t = I_H I_{]\!]0,T]\!]}$ and $Y_t = E[I_H|\mathcal{F}_{T-}]I_{]\!]0,T]\!]}$. As T is predictable, if M is a càdlàg modification of $\{E[I_H|\mathcal{F}_t]\}_{t\geq 0}$ then by Theorem 7.6.2 and Corollary 7.6.7, up to indistinguishability

$$(\Pi_p X)_t = M_{t-}I_{]\!]0,T]\!]} = (\Pi_p Y)_t.$$

Using the assumed condition, this implies that

$$E[I_H A_T] = E[(I_H\bullet A)_T] = E\big[(E[I_H|\mathcal{F}_{T-}]\bullet A)_T\big] = E\big[E[A_T|\mathcal{F}_{T-}]I_H\big]$$

and hence $A_T = E[A_T|\mathcal{F}_{T-}]$. Therefore, A_T is $\mathcal{F}_{T-}$-measurable for every $T \in \mathcal{T}_p$.

As A is adapted and càdlàg, the set $\{A \neq \Pi_p A\}$ is a thin set (Exercise 7.7.1), so it has an exhausting sequence $\{T_i\}_{i\in\mathbb{N}}$ (Theorem 7.5.2).

As $\Pi_p(I_{[\![S]\!]}) = 0$ for any totally inaccessible stopping time S, we know by assumption that $E[A_S - A_{S-}] = E[I_{[\![S]\!]}\bullet A] = 0$. Hence, by Theorem 6.2.9, we can assume without loss of generality that each stopping time T_i is predictable. Therefore, A_{T_i} is $\mathcal{F}_{T_i-}$-measurable and the set

$$\{A > (\Pi_p A)\} \cap [\![T_i]\!] = [\![(T_i)_{\{A_{T_i} > (\Pi_p A)_{T_i}\}}]\!]$$

is a predictable set (Theorem 6.3.1).

For any $B \in \Sigma_p$, the assumed condition implies

$$0 = E\Big[\big((I_B(A - \Pi_p A))\bullet A\big)_\infty\Big] = E\Big[\sum_i I_B(A_{T_i} - (\Pi_p A)_{T_i})(A_{T_i} - A_{T_i-})\Big].$$

Considering the set $B = \{A > \Pi_p A\} \cap [\![T_i]\!]$, we see that, for each i and almost all ω, either A is continuous at T_i or $A_{T_i} \leq (\Pi_p A)_{T_i}$. Similarly, considering $B = \{A < \Pi_p A\} \cap [\![T_i]\!]$, we see that for all i and almost all ω, we also have $A_{T_i} \geq (\Pi_p A)_{T_i}$ or A is continuous at T_i. However, as A_{T_i} is $\mathcal{F}_{T_i-}$-measurable, those ω where A is continuous at T_i form a predictable set, on which, by the projection theorem, $A_{T_i} = (\Pi_p A)_{T_i}$. Hence we have $A_{T_i} = (\Pi_p A)_{T_i}$ for almost all ω.

Therefore, as the exhausting sequence is countable, we see that $A = \Pi_p A$ up to indistinguishability on the set $\{A \neq \Pi_p A\}$. By contradiction, this implies $A = \Pi_p A$ up to indistinguishability. Hence A is predictable. □

8.2.2 Integrals and the Dual Projection

Consider a $\mathcal{B} \otimes \mathcal{F}$-measurable process of integrable variation A. Then the associated map

$$X \mapsto E\Big[\int_{[0,\infty[} X_s(\omega)\, dA_s(\omega)\Big] = E[(X\bullet A)_\infty]$$

is a bounded linear functional of the bounded $\mathcal{B} \otimes \mathcal{F}$-measurable processes. To see this, note that linearity is clear from the definition of the integral and the expectation, and that

$$E[(X\bullet A)_\infty] \leq \big\|\sup_t X_t\big\|_\infty E\Big[\int_{[0,\infty[} |dA|_s\Big],$$

so the map is bounded (in the sense of Definition 1.5.7) under the $X \mapsto \|\sup_t X_t\|_\infty$ norm. The projection Π_x is also a bounded linear operator on the bounded measurable processes, and the results of Section 8.2.1 indicate that the process $(\Pi_x X)\bullet A$ may be of interest. Therefore, we are also interested in understanding Π_x^*, the adjoint operator (or 'dual projection') as defined by

Lemma 1.5.10, as this moves the projection operation from the integrand X to the integrator A.

In this setting, however, the result of Lemma 1.5.10 is unhelpfully abstract. In particular, we would like to see if Π_x^* can act on A as a *process*, rather than on the linear operator $X \mapsto E[(X \bullet A)_\infty]$.

Theorem 8.2.6. *Let A be a càdlàg $\mathcal{B} \otimes \mathcal{F}$-measurable process of integrable variation, that is, $A \in \mathcal{W}$ and $E\big[\int_{[0,\infty[} |dA|\big] < \infty$. Then, for $x = o, p$, there is a unique Σ_x-measurable process $\Pi_x^* A \in \mathcal{A}$ such that, for every bounded $\mathcal{B} \otimes \mathcal{F}$-measurable process X,*

$$E\big[\big((\Pi_x X)\bullet A\big)_\infty\big] = E\big[\big(X \bullet (\Pi_x^* A)\big)_\infty\big] = E\big[\big((\Pi_x X)\bullet (\Pi_x^* A)\big)_\infty\big].$$

The process $\Pi_o^* A$ is called the dual optional projection *of A, and $\Pi_p^* A$ is called the* dual predictable projection *of A.*

Proof. By Lemma 8.1.11, without loss of generality we can assume $A \in \mathcal{W}^+$ and $E[A_\infty] < \infty$. Given the increasing $\mathcal{B} \otimes \mathcal{F}$-measurable process A we can define a measure μ_x on $[0, \infty[\times \Omega$ by

$$\mu_x(B) = E\big[\big((\Pi_x I_B)\bullet A\big)_\infty\big].$$

As Π_x is monotone and preserves constants, it is easy to check that μ_x is a measure. For any $r \in [0, \infty[$ and $H \in \mathcal{F}$, the processes $X = I_H I_{[0,r]}$ and $Y = E[I_H|\mathcal{F}_r] I_{[0,r]}$ have the same Π_x projection. Therefore, μ_x satisfies

$$\mu_x([0, r] \times H) = E[(X \bullet A)_\infty] = E[(Y \bullet A)_\infty] = \int_{[0,\infty[\times\Omega} E[I_H|\mathcal{F}_r] I_{[0,r]} d\mu_x,$$

so by Theorem 8.1.17 we know μ is generated by a unique process, denoted $\Pi_x^* A \in \mathcal{A}^+$. From Theorem 8.2.5, we see that $\Pi_x^* A$ is Σ_x-measurable. □

Remark 8.2.7. $\Pi_x^* A$ should not be confused with $\Pi_x A$, as will be made clear in Example 8.2.17 below.

Corollary 8.2.8. *If $A \in \mathcal{A}^+$ is Σ_x-measurable, then $\Pi_x^* A = A$.*

Proof. By Theorem 8.2.5,

$$E\big[(X \bullet A)_\infty\big] = E\big[\big((\Pi_x X)\bullet A\big)_\infty\big] = E\big[(X \bullet (\Pi_x^* A))_\infty\big]$$

for all $\mathcal{B} \otimes \mathcal{F}$-measurable processes X. By Corollary 8.1.18, the processes A and $\Pi_x^* A$ are indistinguishable. □

Theorem 8.2.9. *Suppose $A \in \mathcal{A}$, X is Σ_x-measurable and A-integrable and $X \bullet A \in \mathcal{A}$, for $x = o, p$. Then*

$$\Pi_x^*(X \bullet A) = X \bullet (\Pi_x^* A).$$

Proof. First suppose X is bounded. By the Radon–Nikodym theorem, we know that $Y\bullet(X\bullet A) = XY\bullet A$, for any bounded process Y. For any set $H \in \Sigma_x$, we have

$$\begin{aligned}
E[(I_H\bullet\Pi_x^*(X\bullet A))_\infty] &= E[(\Pi_x I_H\bullet(X\bullet A))_\infty] = E[(I_H\bullet(X\bullet A))_\infty] \\
&= E[((I_H X)\bullet A)_\infty] = E[(\Pi_x(I_H X)\bullet A)_\infty] \\
&= E[(I_H X\bullet(\Pi_x^* A))_\infty] = E[(I_H\bullet(X\bullet\Pi_x^* A))_\infty].
\end{aligned}$$

Therefore, as measures, $\Pi_x^*(X\bullet A)$ and $X\bullet(\Pi_x^* A)$ must agree when restricted to the Σ_x-measurable sets. As they are Σ_x-measurable processes, the result follows (Remark 8.2.1). Finally, we can extend this to a general X using linearity and a standard monotone approximation argument. □

Theorem 8.2.10. *Suppose A is a continuous $\mathcal{B}\otimes\mathcal{F}$-measurable process of integrable variation. Then $\Pi_o^* A$ is a continuous process and $\Pi_o^* A = \Pi_p^* A$.*

Proof. Suppose $T \in \mathcal{T}_o$ and $X_t = I_{[\![T]\!]}(t)$. Then X is optional, so

$$E\big[(X\bullet A)_\infty\big] = E\big[\big((\Pi_o X)\bullet A\big)_\infty\big] = E\big[\big(X\bullet(\Pi_o^* A)\big)_\infty\big] = E[\Pi_o^* A_T - \Pi_o^* A_{T-}].$$

By continuity of A,

$$E\big[\big((\Pi_o X)\bullet A\big)_\infty\big] = E[A_T - A_{T-}] = 0,$$

so all the above terms are zero. As $\Pi_o^* A$ is càdlàg and optional, its jumps occur on a thin set (Exercise 7.7.1). However, we have seen that A is a.s. continuous at any stopping time, and hence on any thin set, and so must be continuous everywhere. Therefore, $\Pi_o^* A$ is predictable (Theorem 7.2.4) and $\Pi_o^* A = \Pi_p^* A$. □

The following fundamental result relates the dual projection with martingale properties of A.

Theorem 8.2.11. *Suppose $A \in \mathcal{A}$ and $A_0 = 0$ a.s. Then the following conditions are equivalent:*

(i) A is a martingale,
(ii) $\Pi_p^* A$ is indistinguishable from the zero process,
(iii) the restriction of μ_A to Σ_p is zero.

Proof. To show (i) and (iii) are equivalent, note that $\mu_A([\![0_F]\!]) = 0$ as $A_0 = 0$ a.s. and A is right-continuous (up to modification, which is enough here). Furthermore, for any constants $s < t$ and any $B \in \mathcal{F}_s$ we have $\mu_A(]\!]s_B, t_B]\!]) = E[I_B(A_t - A_s)]$. Hence, as these intervals generate Σ_p (Corollary 7.2.5), we see that $E[A_t|\mathcal{F}_s] = A_s$ if and only if the restriction of μ_A to Σ_p is zero. Therefore (i) is equivalent to (iii).

To show (ii) and (iii) are equivalent, consider a process X of the form $X = I_{\rrbracket S,T\rrbracket}$ for $S, T \in \mathcal{T}_o$, so that X is predictable. Then

$$\mu_A\big(\rrbracket S,T\rrbracket\big) = E[(X\bullet A)_\infty] = E\big[\big((\Pi_p X)\bullet A\big)_\infty\big] = E\big[\big(X\bullet(\Pi_p^* A)\big)_\infty\big].$$

Similarly, with $X = I_{\llbracket 0_F\rrbracket}$ for $F \in \mathcal{F}_0$, we have

$$\mu_A(\{0\}\times F) = E\big[\big(I_{\llbracket 0_F\rrbracket}\bullet(\Pi_p^* A)\big)_\infty\big].$$

The equivalence of (ii) and (iii) follows because Σ_p is also generated (Theorem 7.1.9) by stochastic intervals of the forms $\rrbracket S,T\rrbracket$ for $S, T \in \mathcal{T}_o$ and $\llbracket 0_F\rrbracket$ for $F \in \mathcal{F}_0$. $\square$

Corollary 8.2.12. *If $A \in \mathcal{A}$, then the process $B = A - \Pi_p^* A$ is a martingale.*

Proof. We know A is adapted and $\Pi_p^* B = 0$ by Corollary 8.2.8. The result follows. $\square$

Remark 8.2.13. Because of this, the dual predictable projection $\Pi_p^* A$ is sometimes referred to as the *compensator* of A (cf. Theorem 5.5.18).

Corollary 8.2.14. *If a predictable process $A \in \mathcal{A}$ is a martingale, then, for all t, $A_t = A_0$ a.s. In other words, all predictable finite variation martingales are constant.*

Proof. Let $B = A - A_0$, so $B \in \mathcal{A}$, $B_0 = 0$ and B and is a martingale. Therefore, from the theorem, $\Pi_p^* B = A - A_0$ is indistinguishable from the zero process. $\square$

We can obtain a converse to this theorem, in the following way.

Lemma 8.2.15. *Suppose A is an increasing $\mathcal{B}\otimes\mathcal{F}$-measurable process and $\Pi_x^* A$ is its dual Σ_x projection $(x = o, p)$. If X is a bounded $\mathcal{B}\otimes\mathcal{F}$-measurable process and S, T are stopping times, $S \le T$, then*

$$\begin{aligned} E\Big[\int_{\rrbracket S,T\rrbracket} (\Pi_x X_t)\, dA_t \Big| \mathcal{F}_S\Big] &= E\Big[\int_{\rrbracket S,T\rrbracket} X_t\, d(\Pi_x^* A)_t \Big| \mathcal{F}_S\Big] \\ &= E\Big[\int_{\rrbracket S,T\rrbracket} (\Pi_x X_t)\, d(\Pi_x^* A)_t \Big| \mathcal{F}_S\Big]. \end{aligned}$$

Proof. The result follows from the definition of $\Pi_x^* A$, by considering the process $I_{\rrbracket S_B,T_B\rrbracket} X$ for any $B \in \mathcal{F}_S$, as in Corollary 8.2.3. $\square$

Corollary 8.2.16. *If $A, B \in \mathcal{A}$ and $A_0 = B_0$, then the following are equivalent:*

(i) $\Pi_p^* A = \Pi_p^* B$ *up to indistinguishability,*
(ii) $A - B$ *is a martingale.*

Proof. That (ii) implies (i) follows from applying Theorem 8.2.11 to the martingale $A-B$. To show (i) implies (ii), take $X_t = 1$ for all t and $S = s \leq t = T$ in the above lemma. Then, as $\Pi_p^* A = \Pi_p^* B$,

$$E\Big[\int_{]s,t]} dA \Big| \mathcal{F}_s\Big] = E\Big[\int_{]s,t]} d(\Pi_p^* A) \Big| \mathcal{F}_s\Big] = E\Big[\int_{]s,t]} d(\Pi_p^* B) \Big| \mathcal{F}_s\Big]$$
$$= E\Big[\int_{]s,t]} dB \Big| \mathcal{F}_s\Big].$$

That is, $E[A_t - A_s|\mathcal{F}_s] = E[B_t - B_s|\mathcal{F}_s]$, so $E[A_t - B_t|\mathcal{F}_s] = A_s - B_s$. Integrability is guaranteed by the fact $A - B \in \mathcal{A}$. □

We can now give a simple example of these projections, based on the Poisson process.

Example 8.2.17. Let N^T be a Poisson process with parameter λ, stopped at some deterministic time T. Then N^T is optional (as it is right continuous), is increasing, and $E[N^T_\infty] = E[N_T] < \infty$, so $N^T \in \mathcal{A}^+$.

- By Theorem 7.6.5, as $P(N^T_t \neq N^T_{t-}) = 0$ for each t, we see that $(\Pi_p N^T)_t = N^T_{t-}$ up to indistinguishability.
- By Theorem 5.5.18 we have $N^T - \lambda(t \wedge T)$ is a martingale, and by Lemma 8.2.16, it follows that $(\Pi_p^*(N^T))_t = \lambda(t \wedge T)$.

Note, in particular, that $\Pi_p(N^T) \neq \Pi_p^*(N^T)$.

Given the results of Section 8.3, all these arguments can be extended to remove the necessity of stopping N at time T.

8.2.3 A Bound on $\Pi_x^* A$

We now seek to give a bound on $\Pi_x^* A$ in terms of A. To do this, we require the following general inequality, which is a simplified and weakened[2] version of a result in Neveu [138], but is sufficient for our purposes (compare with Lemmata 4.5.5 and 11.5.1).

These results allow us to obtain bounds on processes in the space $\mathcal{A}^p$, where $\mathcal{A}^p$ is given the natural definition $\mathcal{A}^p = \{A \in \mathcal{V} : \|A\|_{\mathcal{A}^p} < \infty\}$, and

$$\|A\|_{\mathcal{A}^p} = \Big\| \int_{[0,\infty]} |dA|_s \Big\|_{L^p}.$$

In particular, $\mathcal{A}^1 = \mathcal{A}$.

[2]Neveu proves the result with x^p replaced by a general convex function.

Lemma 8.2.18. *Suppose X and Y are nonnegative random variables such that, for any $\lambda \geq 0$, we have $E[(X-\lambda)^+] \leq E[YI_{\{X>\lambda\}}]$. Then for any $p \in [1,\infty[$,*

$$E[X^p] \leq p^p E[Y^p].$$

If Y is a constant, then for p an integer we have the tighter bound

$$E[X^p] \leq p! Y^p.$$

Proof. The case $p=1$ is trivial by assumption with $\lambda = 0$. For any $\lambda, \beta > 0$, we know that

$$\beta E[(X/\beta - \lambda)^+] \leq E[YI_{\{X>\beta\lambda\}}].$$

By simple calculations,

$$\int_{[0,\infty]} \beta(X/\beta - \lambda)^+ \lambda^{p-2} d\lambda = \frac{\beta}{p(p-1)}\Big(\frac{X}{\beta}\Big)^p$$

and

$$\int_{[0,\infty]} YI_{\{X>\beta\lambda\}}\lambda^{p-2}d\lambda = Y\int_{[0,X/\beta]}\lambda^{p-2}d\lambda = \frac{1}{p-1}Y\Big(\frac{X}{\beta}\Big)^{p-1},$$

so by Fubini's theorem and our initial inequality,

$$\beta E\Big[\Big(\frac{X}{\beta}\Big)^p\Big] \leq E\Big[pY\Big(\frac{X}{\beta}\Big)^{p-1}\Big]. \tag{8.2}$$

By Young's inequality (Lemma 1.5.26),

$$pY\Big(\frac{X}{\beta}\Big)^{p-1} \leq Y^p + (p-1)\Big(\frac{X}{\beta}\Big)^p,$$

so (8.2) becomes

$$\frac{\beta}{\beta^p}E[X^p] \leq E\Big[Y^p + \frac{(p-1)}{\beta^p}X^p\Big]$$

and we see, if $\beta = p$,

$$E[X^p] \leq p^p E[Y^p].$$

If Y is a constant, then (8.2) evaluated with $\beta = 1$ implies that, for every p, $E[X^p] \leq pYE[X^{p-1}]$. Iterating this inequality gives the result. □

Theorem 8.2.19. *For any $p \in [1,\infty[$, there exists a constant C_p such that, for any $A \in \mathcal{V}$, if $B = \Pi_x^* A$ for $x = o, p$ then*

$$E\Big[\Big(\int_{[0,\infty]} |dB|\Big)^p\Big] \leq C_p E\Big[\Big(\int_{[0,\infty]} |dA|\Big)^p\Big].$$

Proof. By Theorem 8.2.6, if $X = \text{sign}(dB)$ (which is a bounded Σ_x-measurable process, defined as in Remark 1.7.11), then we know that, for any Σ_x-measurable set $D \subseteq \Omega \times [0, \infty]$,

$$E\Big[\int_{[0,\infty]} I_D |dB|\Big] = E\Big[\int_{[0,\infty]} I_D X dB\Big]$$
$$= E\Big[\int_{[0,\infty]} I_D X dA\Big] \leq E\Big[\int_{[0,\infty]} I_D |dA|\Big],$$

so the compensator of the measure $|dA|$ does not grow slower than $|dB|$. Therefore, it is enough to establish the inequality when A is increasing.

Let $T = \inf\{t : B_t \geq \lambda\}$. By Theorem 8.2.6 applied with $X = I_D I_{[\![T,\infty]\!]}$ for $D \in \mathcal{F}_{T-}$ we know that

$$E[B_\infty - B_{T-}|\mathcal{F}_{T-}] = E[A_\infty - A_{T-}|\mathcal{F}_{T-}] \leq E[A_\infty|\mathcal{F}_{T-}],$$

and so, as $I_{\{B_\infty \geq \lambda\}} = I_{\{T<\infty\}}$ is $\mathcal{F}_{T-}$ measurable,

$$E[(B_\infty - \lambda)^+] \leq E[(B_\infty - B_{T-})I_{\{B_\infty > \lambda\}}] \leq E[A_\infty I_{\{B_\infty > \lambda\}}].$$

Applying Lemma 8.2.18, we have the result. □

We would very much like to have a converse bound. However, this is not, in general, possible. Instead, we obtain the following weaker inequality.

Theorem 8.2.20. *For every $p \geq 1$ there exists a constant C_p such that, for any $A \in \mathcal{V}^+$ (which implies A is optional), if $B = \Pi_p^* A$ then*

$$E[A_\infty^p] \leq C_p E\Big[B_\infty^p + \big(\sup_t \Delta A_t\big)^p\Big].$$

Proof. Let $T = \inf\{t : A_t \geq \lambda\}$. For any $D \in \mathcal{F}_{T-}$, we know that $I_D I_{]\!]T,\infty]\!]}$ is predictable. Applying Theorem 8.2.6,

$$E[A_\infty - A_T|\mathcal{F}_{T-}] = E[B_\infty - B_T|\mathcal{F}_{T-}] \leq E[B_\infty|\mathcal{F}_{T-}]$$

and so

$$E[(A_\infty - \lambda)^+] \leq E[(A_\infty - A_T + \Delta A_T)I_{\{A_\infty > \lambda\}}] \leq E\Big[\big(B_\infty + \sup_t \Delta A_t\big)I_{\{A_\infty > \lambda\}}\Big].$$

Applying Lemma 8.2.18 and simple calculations, we have the result. □

These results can be seen as a special case of a general inequality.

Theorem 8.2.21 (Garsia's Lemma). *Let A be an adapted nonnegative increasing process, and Y a nonnegative integrable random variable. Suppose one of the following conditions holds:*

(i) For any stopping time T,

$$E[A_\infty|\mathcal{F}_T] - A_{T-} \leq E[Y|\mathcal{F}_T],$$

(ii) A is predictable, $A_0 = 0$ and for any predictable time T,

$$E[A_\infty|\mathcal{F}_T] - A_T \leq E[Y|\mathcal{F}_T].$$

Then, for all $\lambda \geq 0$, we have $E[(A_\infty - \lambda)^+] \leq E[Y I_{\{A_\infty > \lambda\}}]$, so Lemma 8.2.18 can be applied.

Proof. Consider case (i). Let $T = \inf\{t : A_t \geq \lambda\}$. Then $A_{T-} \leq \lambda$. Since

$$\{A_\infty \geq \lambda\} = \{T < \infty\} \cup \{T = \infty, A_\infty \geq \lambda\} \in \mathcal{F}_T,$$

by assumption we have

$$E[(A_\infty - \lambda)^+] \leq E[(A_\infty - A_{T-}) I_{\{A_\infty > \lambda\}}] \leq E[Y I_{\{A_\infty > \lambda\}}].$$

Now consider case (ii). In this case, $T = \inf\{t : A_t \geq \lambda\}$ is a predictable time, and $T > 0$. Then let $\{T_n\}_{t \geq 0}$ be a sequence of predictable times announcing T. By assumption,

$$E[A_\infty|\mathcal{F}_{T_n}] - A_{T_n} \leq E[Y|\mathcal{F}_{T_n}].$$

Letting $n \to \infty$, by martingale and monotone convergence, we obtain

$$E[A_\infty|\mathcal{F}_{T-}] - A_{T-} \leq E[Y|\mathcal{F}_{T-}].$$

As

$$\{A_\infty \geq \lambda\} = \{T < \infty\} \cup \{T = \infty, A_\infty \geq \lambda\} \in \mathcal{F}_{T-}$$

and $A_{T-} \leq \lambda$, the result follows. □

8.3 Locally Finite Variation Processes

Many of the processes we are interested in are not of finite variation when we consider the whole time interval $[0, \infty]$. For this reason, it is of interest to consider the space of 'locally finite variation processes', where 'locally' is taken in the stochastic sense, that is, a process X is in $\mathcal{A}_{\text{loc}}$ if and only if there exists a sequence of stopping times $\{T_n\}_{n \in \mathbb{N}}$ with $T_n \to \infty$ and $X^{T_n} \in \mathcal{A}_{\text{loc}}$ for every n.

Remark 8.3.1. It is easy to verify that $\mathcal{V} = \mathcal{V}_{\text{loc}}$, but $\mathcal{A} \neq \mathcal{A}_{\text{loc}}$.

Most of the results proved above for processes in $\mathcal{A}$ or $\mathcal{A}^+$ have 'local' versions valid for processes in $\mathcal{A}_{\text{loc}}$ or $\mathcal{A}^+_{\text{loc}}$. In particular, using linearity, the dual predictable projections can be extended to processes which satisfy 'locally' the conditions of Theorem 8.2.6.

More formally, suppose A is an increasing $\mathcal{B} \otimes \mathcal{F}$-measurable process for which there is an increasing sequence $\{T_n\}_{n\in\mathbb{N}}$ of stopping times such that $\lim_n T_n = \infty$ a.s. and $E[A^{T_n}_\infty] = E[A_{T_n}] < \infty$ for each n. Then the dual Σ_x projection can be defined for each stopped process A^{T_n}, and the dual Σ_x projection of the process A can be defined by 'pasting' together the pieces, for example, by

$$\Pi^*_x A = \Pi^*_x(A^{T_1}) I_{[\![0,T_1]\!]} + \sum_{n=2}^{\infty} \Pi^*_x(A^{T_n}) I_{]\!]T_{n-1},T_n]\!]} \tag{8.3}$$

Then $\Pi^*_x A$ is a Σ_x-measurable process in $\mathcal{A}^+_{\text{loc}}$. Alternative equivalent definitions of the 'pasting' are given in Exercise 8.4.5, and the pasting can easily be shown to be independent of the choice of localizing sequence.

Here, for example, is the 'local' version of Theorem 8.1.17.

Theorem 8.3.2. *Suppose μ is a positive measure on $([0,\infty[\times\Omega, \mathcal{B}\otimes\mathcal{F})$. For μ to be of the form μ_A where $A \in \mathcal{A}^+_{\text{loc}}$, it is necessary and sufficient that*

(i) there is an increasing sequence of stopping times $\{T_n\}_{n\in\mathbb{N}}$ such that $\lim_n T_n = \infty$ and $\mu([\![0,T_n]\!]) < \infty$ for each n,
(ii) the evanescent sets have measure zero,
(iii) for every $t \in [0,\infty[$ and $H \in \mathcal{F}$,

$$\mu([0,t]\times H) = \int_{[0,\infty[\times\Omega} E[I_H|\mathcal{F}_t] I_{[\![0,t]\!]} d\mu.$$

Proof. For each n consider the measure μ_n on $([0,\infty[\times\Omega, \mathcal{B}\otimes\mathcal{F})$ defined by $\mu_n(B) = \mu(B \cap [\![0,T_n]\!])$. Then μ_n satisfies the conditions of Theorem 8.1.17, and so is of the form $\mu_n = \mu_{A^n}$, where $A^n \in \mathcal{A}^+$. However, by uniqueness $A^m_t(\omega) = A^n_t(\omega)$ if $t \le T_n(\omega)$, so a process $A \in \mathcal{A}^+_{\text{loc}}$ is obtained by 'pasting' the processes A^n together using (8.3). □

Lemma 8.3.3. *Suppose $A \in \mathcal{A}_{\text{loc}}$. Then there is a unique predictable process $\tilde{A} \in \mathcal{A}_{\text{loc}}$ such that $A - \tilde{A}$ is a local martingale starting at zero.*

Proof. Again we can suppose $A_0 = 0$. Because A is locally integrable, there is an increasing sequence $\{T_n\}_{n\in\mathbb{N}}$ of stopping times such that $\lim T_n = \infty$ and $E\big[\int_0^{T_n} |dA_s|\big] < \infty$ for all n. By Corollary 8.2.12 the dual predictable projection $\Pi^*_p A^{T_n}$ is such that $A^{T_n} - \Pi^*_p A^{T_n}$ is a martingale.

By uniqueness, if $T_m > T_n$ then $(\Pi^*_p A^{T_m})^{T_n} = \Pi^*_p A^{T_n}$, so we can define a predictable process $\tilde{A}$ by putting $\tilde{A}^{T_n} = \Pi^*_p A^{T_n}$. Clearly $A - \tilde{A}$ is then a local martingale. □

We therefore extend the definition of Theorem 8.2.6.

Definition 8.3.4. *The process $\tilde{A}$ constructed in Lemma 8.3.3 is called the* dual predictable projection *or* compensator *of $A \in \mathcal{A}_{\mathrm{loc}}$ and is written $\Pi_p^* A$.*

The following extension of Exercise 7.7.9 will be used in the coming chapter.

Lemma 8.3.5. *Suppose $A \in \mathcal{V}_0$ is either predictable or has bounded jumps. Then $A \in \mathcal{A}_{0,\mathrm{loc}}$.*

Proof. Write $D_t := \int_{[0,t]} |dA_s|$ and

$$S_k = \inf\{t : D_t \geq k\}.$$

Because $A_0 = 0$, we see that S_k is almost surely positive. If A has bounded jumps, one can immediately take the localizing sequence $T_n = S_n$.

From Lemma 8.1.11, if A is predictable, then the increasing, process D is predictable. Write $B_k = \{(t,\omega) : D \geq k\}$. Then S_k is the debut of $B_k \in \Sigma_p$, and because D is right continuous, $[\![S_k]\!] \subset B_k$, so S_k is predictable by Lemma 7.3.7.

Suppose $\{S_m^k\}_{m\in\mathbb{N}}$ is a sequence of stopping times which announce S_k. Then the variation of A on $[\![0, S_m^k]\!]$ is less than k, and so is integrable. Define

$$T_n = \bigvee_{k,m\in\{1,2,\ldots,n\}} S_m^k.$$

Then $\lim T_n = \infty$ and $E\big[\int_{[\![0,T_n]\!]} |dA_s|\big] \leq n$, so $A \in \mathcal{A}_{0,\mathrm{loc}}$. □

8.4 Exercises

Exercise 8.4.1. Let A be a Σ_x-measurable ($x = o, p$), càdlàg, locally finite variation process. Suppose that $\mu_A([\![T]\!]) = 0$ for all stopping times $T \in \mathcal{T}_x$. Show that A is continuous, up to indistinguishability.

Exercise 8.4.2. Let $M \in \mathcal{A}$ be an integrable variation càdlàg martingale. Show that, if X is a predictable $|dM|$-integrable process with $E\big[\int |X|\,|dM|\big] < \infty$, then $X \bullet M$ is a uniformly integrable martingale. (Hint: Use Theorems 8.2.6, 5.4.6 and 8.2.11.)

Exercise 8.4.3. Construct a filtered probability space with an example of a process $A \in \mathcal{A}_{\mathrm{loc}}$ and a bounded stopping time T such that $A^T \notin \mathcal{A}$.

Exercise 8.4.4. Let $A \in \mathcal{A}_{\mathrm{loc}}$ and $T \in \mathcal{T}_x$ for $x = o, p$. Show that $(\Pi_x A)^T = \Pi_x(A^T)$ and $(\Pi_x^* A)^T = \Pi_x^*(A^T)$, that is, the projection and dual projection commute with stopping the process.

Exercise 8.4.5. Show that, for $A \in \mathcal{A}_{\text{loc}}$, the 'pasting' given in (8.3) agrees with the alternatives

$$\begin{aligned}\Pi_x^* A &= \Pi_x^*(A^{T_1}) + \sum_{n=2}^{\infty} \Big(\Pi_x^*(A^{T_n}) - \Pi_x^*(A^{T_{n-1}})\Big) \\ &= \sup_n \{\Pi_x^*(A^{T_n}) - \infty I_{\{t>T_n\}}\} \\ &= \lim_n \{\Pi_x^*(A^{T_n})\},\end{aligned}$$

and that $\Pi_x^* A$ is independent of the choice of localizing sequence for A.

Exercise 8.4.6. Let N be a Poisson process with jumps at times $T_1, T_2, \ldots$. Show that $(X \bullet N)_t = \sum_{\{n:t\leq T_n\}} X_{T_n}$ for any measurable process X. Hence show that if $(X \bullet N)_t = 0$ almost surely and X is predictable, then $X = 0$ $dt \times dP$-almost everywhere. Give a counterexample when X is not predictable.

Exercise 8.4.7. Let N be a Poisson process with parameter λ. Show that

$$\int_{[0,t]} N_{s-} dN_s = \frac{(N_t - 1)N_t}{2},$$

and hence that

$$(\Pi_p^*(N^2))_t = \lambda \int_{[0,t]} \Big(N_{s-} + \frac{1}{2}\Big) ds.$$

Exercise 8.4.8. Let N be a Poisson process with parameter λ, and X be the associated martingale, i.e. $X_t = N_t - \lambda t$. Show that $\bar{X}_t := \sup_{s\leq t} X_t$ is in $(\mathcal{A}_0^+)_{\text{loc}}$ and that $(\Pi_p^*(\bar{X}))_t = \lambda \int_{[0,t]} I_{\{\bar{X}_s - X_s < 1\}} ds$.

Exercise 8.4.9. Let N be a Poisson process, and X be the associated martingale $X_t = N_t - \lambda t$. Let T_n be the time of the nth jump of N and $\{\mathcal{F}_t\}_{t\geq 0}$ the completed filtration generated by N. Show that a process Y_t is a local martingale if and only if $Y_t = \int_{[0,t]} Z_t dX_t$ for some predictable process Z_t. (Hint: Exercise 5.7.10 may prove helpful)

This very useful result is known as the *martingale representation theorem* for Poisson processes.

9

The Doob–Meyer Decomposition

In the previous chapter, we have seen that, for any process $X \in \mathcal{A}$, we can find a predictable process $Y = \Pi_p^* X$ such that $X - Y$ is a martingale. This is a fundamentally useful property, and in this chapter we show that a similar decomposition holds for all right-continuous local supermartingales (and hence local submartingales). To obtain this, we first consider the particularly 'nice' class of processes given by class (D). Recall that a right-continuous uniformly integrable supermartingale X is said to be of class (D) if the set of random variables $\{X_T\}_{T \in \mathcal{T}}$ is uniformly integrable (where $\mathcal{T}$ is the set of all stopping times). These were introduced in Section 5.6.

As before, we assume the usual conditions on the filtration throughout, and all martingales will be taken to be their càdlàg versions. For simplicity, we shall assume $\mathbb{T} = [0, \infty]$ or $[0, \infty[$ and $\mathcal{F}_{\infty-} = \mathcal{F}_\infty$, so that $M_t \to M_\infty$ a.s. for any uniformly integrable martingale M. If A is an increasing process, we know that $A_{\infty-} = \lim_n A_n$ exists a.s., and we shall write $A_\infty = A_{\infty-}$ for simplicity. Unless otherwise indicated, all (in)equalities should be read as 'up to indistinguishability'.

9.1 Decompositions of Potentials

In Chapter 5 we showed (Theorem 5.4.3) that a uniformly integrable supermartingale can be expressed as the sum of a martingale and a potential (i.e. a nonnegative càdlàg supermartingale with $\lim_{t\to\infty} E[X_t] = 0$, Definition 5.4.1); this was called its Riesz decomposition. Instead of a potential, we will now show that uniformly integrable supermartingales can be expressed as the sum of a martingale and a decreasing process. We begin by showing that this is true of the potentials, which are special cases of supermartingales.

S.N. Cohen, R.J. Elliott, *Stochastic Calculus and Applications*,
Probability and Its Applications, DOI 10.1007/978-1-4939-2867-5_9

Definition 9.1.1 (Doob–Meyer Decomposition). *Suppose X is a càdlàg process. Then X is said to have a* Doob–Meyer decomposition *if there is a right-continuous local martingale M and a predictable finite variation process $A \in \mathcal{V}_0$ such that $X = M - A$.*

Remark 9.1.2. For X a uniformly integrable càdlàg supermartingale, we know from Theorem 4.6.5 that X has a Riesz decomposition $X = Y + Z$ where Y is a uniformly integrable martingale and Z is a potential. Therefore, X has a Doob–Meyer decomposition if and only if its potential Z has a Doob–Meyer decomposition.

Lemma 9.1.3. *Let Z be a potential and suppose Z has a decomposition $Z = M - A$, where M is a martingale and $A \in \mathcal{V}_0$. Then*

$$Z = \{E[A_\infty|\mathcal{F}_t]\}_{t\geq 0} - A$$

In this situation we say that the potential Z is generated by the increasing process *A.*

Proof. Suppose Z has a Doob–Meyer decomposition $Z = M - A$. Then

$$E[Z_t] = E[M_t] - E[A_t]$$

and $\lim_{t\to\infty} E[Z_t] = 0$. As

$$\lim_{s,t\to\infty} E\big[|Z_t - Z_s|\big] \leq \lim_{s,t\to\infty} E[Z_t + Z_s] = 0,$$

we know that Z converges in L^1, and, as A is an increasing process, A converges a.s. By monotone convergence,

$$E[A_\infty] = \lim_{t\to\infty} E[A_t] = \lim_{t\to\infty} (E[M_t] - E[Z_t]) = E[M_0] < \infty,$$

so that $A \in \mathcal{A}_0^+$, and furthermore A converges in L^1. We can then define

$$M_\infty = \lim_{t\to\infty} M_t = \lim_{t\to\infty} (Z_t + A_t) = A_\infty,$$

the limits being taken both a.s. and in L^1. By Theorem 2.5.8, M is uniformly integrable and hence $M_t = E[A_\infty|\mathcal{F}_t]$. The result follows. □

Lemma 9.1.4. *Without loss of generality, in the decomposition of Lemma 9.1.3, the nondecreasing process A can be assumed to be predictable (so we have a Doob–Meyer decomposition, as in Definition 9.1.1).*

Proof. If A is not predictable, consider the dual predictable projection $B_t = (\Pi_p^* A)_t$. By linearity of Π_p^*, we know $\Pi_p^*(A - B)$ is indistinguishable from the zero process, and from Corollary 8.2.16, we see that $A - B$ is a martingale. As $A - B \in \mathcal{A}$, we know $A_t - B_t$ converges almost surely and in L^1 as $t \to \infty$, hence it is uniformly integrable and

$$E[A_\infty - B_\infty | \mathcal{F}_t] = A_t - B_t.$$

Therefore, we can write

$$Z_t = E[B_\infty | \mathcal{F}_t] - B_t.$$

That is, A and B generate the same potential. □

Theorem 9.1.5. *If a potential Z has a decomposition*

$$Z_t = M_t - A_t = E[A_\infty | \mathcal{F}_t] - A_t$$

with $A \in \mathcal{A}_0^+$ and M a càdlàg martingale, then Z is of class (D). Furthermore, the increasing process A can be taken to be predictable and, in this case, for every stopping time T,

$$Z_T = E[A_\infty | \mathcal{F}_T] - A_T \quad \text{a.s.}$$

Proof. In the above decomposition $M_T = E[A_\infty | \mathcal{F}_T]$ for any stopping time T, so the set of random variables $\{M_T\}_{T \in \mathcal{T}}$ is uniformly integrable. Because $A \in \mathcal{A}_0^+$ is an increasing process,

$$0 \geq -A_T \geq -A_\infty \quad \text{a.s.}$$

for any stopping time T, so the set of random variables $\{-A_T\}_{T \in \mathcal{T}}$ is uniformly integrable. Therefore, Z is of class (D).

Replacing A with $\Pi_p^* A$, we obtain a new decomposition

$$Z = M + (-A + \Pi_p^* A) - \Pi_p^* A = \tilde{M} - \tilde{A},$$

where $\tilde{M} = M + (-A + \Pi_p^* A)$ is a martingale and $\tilde{A} = \Pi_p^* A$ is a predictable process in $\mathcal{A}_0^+$.

We can therefore suppose that A is predictable. By optional stopping (Theorem 5.3.1), for any stopping time T, we see

$$Z_T = E[A_\infty | \mathcal{F}_T] - A_T \quad \text{a.s.}$$

□

The main result of this chapter is the following theorem, which is a converse of the above, and shows that if a potential is of class (D) then it has a Doob–Meyer decomposition.

Theorem 9.1.6 (Doob–Meyer Decomposition: Class (D) Potentials). *Suppose Z is a potential of class (D). Then there is a unique predictable integrable increasing càdlàg process $A \in \mathcal{A}_0^+$ such that Z is the potential generated by A. That is, up to indistinguishability,*

$$Z_t = E[A_\infty | \mathcal{F}_t] - A_t.$$

Proof. Uniqueness. If $A, B \in \mathcal{A}_0^+$ are two predictable processes such that $Z = M - A = N - B$ for some martingales M, N, then $A - B$ is a predictable process in $\mathcal{A}_0$ which is also a martingale. By Theorem 8.2.11, we know that $A - B$ is indistinguishable from the zero process. □

The *existence* of a Doob–Meyer decomposition for a potential of class (D) will be established following a sequence of lemmata. We take the assumptions of the theorem as given for the remainder of this section. The first step is to construct a measure on the predictable σ-algebra Σ_p, sometimes called the *Doléans measure* of the potential. From Theorem 7.1.9 we know that Σ_p is generated by stochastic intervals of the forms $[\![0_A]\!]$ and $]\!]S, T]\!]$, where $A \in \mathcal{F}_0$ and S and T are arbitrary stopping times. Write $\mathcal{C}_p$ for the algebra of finite unions of such intervals.

For any $C \in \mathcal{C}_p$, define S_1 as the debut of C, T_1 as the debut of $]\!]S_1, \infty]\!] \cap C^c$, S_2 as the debut of $]\!]T_1, \infty[\![\cap C$ and so on. Using these stopping times, C can be written in a unique way as

$$C = [\![0_F]\!] \cup]\!]S_1, T_1]\!] \cup \ldots \cup]\!]S_n, T_n]\!],$$

where $F = \{S_1 = 0\} \in \mathcal{F}_0$. Define

$$\bar{C} := [\![0_F]\!] \cup [\![S_1, T_1]\!] \cup \ldots \cup [\![S_n, T_n]\!].$$

Using this representation, for a given potential Z, a function $\mu : \mathcal{C}_p \to \mathbb{R}$ can be defined by

$$\mu(C) := E[Z_{S_1} - Z_{T_1}] + \ldots + E[Z_{S_n} - Z_{T_n}]. \tag{9.1}$$

Clearly, if $C_1, C_2 \in \mathcal{C}_p$ and $C_1 \cap C_2 = \emptyset$, then $\mu(C_1 \cup C_2) = \mu(C_1) + \mu(C_2)$. As Z is a supermartingale, μ is nonnegative, and as Z is a potential, μ is bounded by $E[Z_0]$. Hence μ is a finitely additive set function on the algebra $\mathcal{C}_p$. The next two lemmata show that μ is countably additive. By Carathéodory's extension theorem (Theorem 1.2.7), this will allow us to extend μ to the σ-algebra $\Sigma_p = \sigma(\mathcal{C}_p)$.

Lemma 9.1.7. *Let μ be defined as in (9.1). For any $C \in \mathcal{C}_p$, any $\epsilon > 0$, there is an element $D \in \mathcal{C}_p$ such that $\bar{D} \subseteq C$ and $\mu(C) \leq \mu(D) + \epsilon$.*

Proof. It is sufficient to consider the case when $C =]\!]S, T]\!]$, for stopping times $S < T$ on $S < \infty$. Write S_n for the restriction of $S + n^{-1}$ to the set $\{S + n^{-1} < T\}$ and T_n for the restriction of T to the same set (in the sense of Definition 6.2.7).

By construction, $S_n > S$ on $\{S < \infty\}$ and $S_n \geq S$ for all ω. Furthermore, $S = \lim_n S_n$. Similarly $T_n = T$ on $\{S_n < \infty\}$, $T_n \geq T$ and $T = \lim_n T_n$. Therefore, for all n, $[\![S_n, T_n]\!] \subset]\!]S, T[\![$. As Z is right continuous, $\lim_n Z_{S_n} = Z_S$ a.s. and $\lim_n Z_{T_n} = Z_T$ a.s. By assumption, Z is of class (D), so $\{Z_{S_n}\}_{n\in\mathbb{N}}$ and $\{Z_{T_n}\}_{n\in\mathbb{N}}$ are uniformly integrable, and these limits hold in L^1. Consequently, $\lim_n E[Z_{S_n} - Z_{T_n}] = E[Z_S - Z_T]$, and the result is proven by taking $D = [\![S_n, T_n]\!]$ for n sufficiently large. □

Lemma 9.1.8. *Suppose* $\{C_n\}_{n\in\mathbb{N}}$ *is a decreasing sequence of elements of* $\mathcal{C}_p$ *such that* $\cap_n C_n = \emptyset$*. Then* $\lim_n \mu(C_n) = 0$*.*

Proof. Fix $\epsilon > 0$. From Lemma 9.1.7, for each n there is a set $D_n \in \mathcal{C}_p$ such that $\bar{D}_n \subset C_n$ and $\mu(C_n) \leq \mu(D_n) + 2^{-n}\epsilon$. Write

$$\Delta_n = D_1 \cap D_2 \cap \ldots \cap D_n \subseteq D_n.$$

Then

$$\begin{aligned} C_n \setminus \Delta_n &= C_n \cap (D_1 \cap D_2 \cap \ldots \cap D_n)^c \\ &= C_n \cap (D_1^c \cup D_2^c \cup \ldots \cup D_n^c) = \bigcup_{k=1}^n (C_n \cap D_k^c) \subseteq \bigcup_{k=1}^n (C_k \cap D_k^c). \end{aligned}$$

Therefore,

$$\mu(C_n \setminus \Delta_n) = \mu(C_n) - \mu(\Delta_n) \leq \sum_{k=1}^n \big(\mu(C_k) - \mu(D_k)\big) \leq \epsilon \sum_{k=1}^n 2^{-k} \leq \epsilon.$$

The sequence $\{\bar{\Delta}_n\}_{n\in\mathbb{N}}$ is decreasing. Write S_n for the debut of $\bar{\Delta}_n$. Then, because the intervals defining $\bar{\Delta}_n$ are closed for each ω (this is why we need them closed), the S_n are increasing. As $\cap_n \bar{\Delta}_n \subseteq \cap_n C_n = \emptyset$ we see $\lim_n S_n = \infty$ a.s. Now

$$\mu(\Delta_n) \leq E[Z_{S_n} - Z_\infty] = E[Z_{S_n}].$$

As Z is a potential of class (D), we have $\lim_n E[Z_{S_n}] = 0$ so $\lim_n \mu(\Delta_n) = 0$. Consequently, $\lim_n \mu(C_n) \leq \epsilon$ for every $\epsilon > 0$, that is, $\lim_n \mu(C_n) = 0$. □

The preceding lemma, together with Lemma A.1.4, shows that μ is a countably additive set function on the algebra $\mathcal{C}_p$. As $\mu\big([\![0, \infty[\![\big) = E[Z_0] < \infty$, it is a finite measure on $\mathcal{C}_p$ and so, by Caratheodory's extension theorem (Theorem A.1.12), μ can be extended in a unique way to a measure on the σ-algebra $\Sigma_p = \sigma(\mathcal{C}_p)$.

Lemma 9.1.9. *For* $H = [\![0]\!]$*, and for every predictable evanescent set* $H \subset [\![0, \infty[\![$*, we have* $\mu(H) = 0$*.*

Proof. The case $H = [\![0]\!]$ is obvious from the definition of μ.

Otherwise, because H is evanescent, its debut D_H is almost surely infinite. As $A = \{D_H < \infty\}$ is of probability zero, it belongs to $\mathcal{F}_0$. Therefore 0_A is a stopping time, and we have

$$H \subseteq [\![D_H, \infty[\![\subseteq [\![0_A]\!] \cup]\!]0_A, \infty[\![.$$

It follows that

$$\mu(H) \leq \mu\big(]\!]0_A, \infty[\![\big) = E[I_A Z_0] = 0.$$

□

Proof of Theorem 9.1.6. Existence. For every $C \in \mathcal{B} \otimes \mathcal{F}$, the process $\Pi_p I_C$ is predictable and bounded, and if $C \in \Sigma_p$ then $\Pi_p I_C = I_C$. Hence, as in the proof of Theorem 8.2.6, the measure μ defined on Σ_p can be immediately extended to a measure $\bar{\mu}$ defined on $\mathcal{B} \otimes \mathcal{F}$, by setting

$$\bar{\mu}(C) = \int_{\Omega \times [0,\infty[} (\Pi_p I_C) d\mu.$$

Clearly $\bar{\mu}$ has finite mass (because μ is bounded), does not charge evanescent sets and, as the predictable projection commutes with the conditional expectation, for any $H \in \mathcal{F}$,

$$\bar{\mu}\big([0,t] \times H\big) = \int_{[0,\infty[\times \Omega} E[I_H | \mathcal{F}_t] I_{[0,t]} d\bar{\mu}.$$

From Theorem 8.1.17, it follows that $\bar{\mu}$ is generated by an integrable increasing process $A \in \mathcal{A}_0^+$. From Theorem 8.2.5, A is predictable.

For any two stopping times S and T with $S \leq T$ by definition

$$\mu\big(]\!]S,T]\!]\big) = E[Z_S - Z_T] = E[A_T - A_S].$$

Therefore, for any $t \in [0,\infty[$ and $H \in \mathcal{F}_t$, setting $S = t_H$ and $T = \infty$ we have

$$E[I_H(A_\infty - A_t)] = E[I_H Z_t],$$

that is,

$$Z_t = E[A_\infty - A_t | \mathcal{F}_t] \quad \text{a.s.}$$

for all t. Therefore, Z is the potential generated by the increasing predictable process $A \in \mathcal{A}_0^+$, and

$$Z_t = E[A_\infty | \mathcal{F}_t] - A_t$$

gives the Doob–Meyer decomposition of Z. □

9.2 Decompositions of Supermartingales

We now seek to extend the Doob–Meyer decomposition from potentials to more general supermartingales (Fig. 9.1).

For a general right-continuous uniformly integrable supermartingale X, the potential in its Riesz decomposition (Theorem 5.4.3) is of class (D) if and only if X is of class (D). (This follows immediately from uniform integrability of the martingale term M, as every uniformly integrable martingale is of class (D).) This implies the following result.

Theorem 9.2.1 (Doob–Meyer Decomposition: Class (D) Supermartingales). *Suppose X is a right-continuous supermartingale of class (D). Then there exists a unique increasing predictable process $A \in \mathcal{A}_0^+$ such that the process $M = X + A$ is a uniformly integrable martingale.*

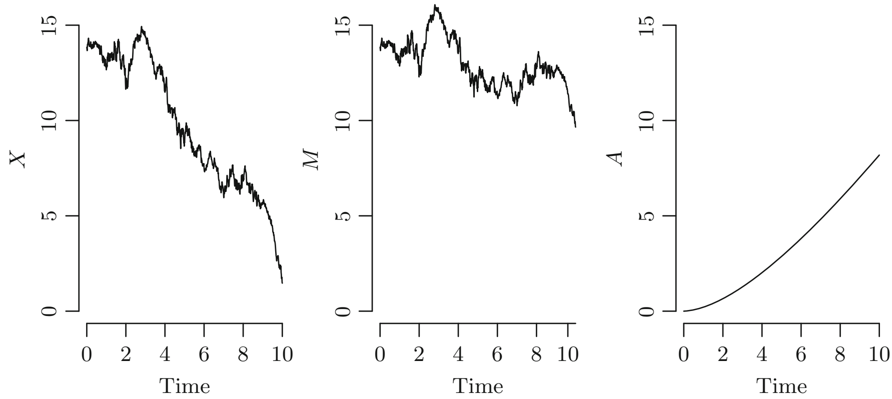

Fig. 9.1. A path of a supermartingale X, and the processes M and A, where $X = M - A$ is the Doob–Meyer decomposition.

Proof. Let $X = \tilde{M} + Z$ be the Riesz decomposition of X. Then, by the above remark, Z is of class (D) and so $Z_t = E[A_\infty|\mathcal{F}_t] - A_t$ for some $A \in \mathcal{A}_0^+$, by Theorem 9.1.6. If $M_t = \tilde{M}_t + E[A_\infty|\mathcal{F}_t]$, then we have $X = M - A$, and M is a uniformly integrable martingale. $\square$

Definition 9.2.2. *Suppose X is a right-continuous uniformly integrable supermartingale. Then X is said to be* regular *if, for every predictable stopping time T,*

$$E[X_{T-}] = E[X_T].$$

Lemma 9.2.3. *A right-continuous uniformly integrable supermartingale is regular if and only if*

$$X_{T-} = E[X_T|\mathcal{F}_{T-}] \quad a.s.$$

for every predictable stopping time T.

Proof. Let X have Riesz decomposition $X = M + Z$. If T is a predictable stopping time announced by the sequence $\{T_n\}_{n\in\mathbb{N}}$, then, as M is a right-continuous uniformly integrable martingale, Theorem 5.6.13 states that

$$M_{T-} = E[M_T|\mathcal{F}_{T-}].$$

By the optional stopping theorem, the sequence $\{Z_{T_n}\}_{n\in\mathbb{N}}$ is a supermartingale in the filtration $\{\mathcal{F}_{T_n}\}_{n\in\mathbb{N}}$. By Theorem 4.4.1 and nonnegativity of Z, the random variable $Z_{T-} = \lim_n Z_{T_n}$ is integrable. Consequently, X_{T-} is integrable. However, $X_{T_n} \geq E[X_T|\mathcal{F}_{T_n}]$, again by Doob's optional stopping theorem, so in the limit

$$X_{T-} \geq E[X_T|\mathcal{F}_{T-}],$$

because $E[X_T|\mathcal{F}_{T-}] = \lim_n E[X_T|\mathcal{F}_{T_n}]$ by the martingale convergence theorem and Lemma 6.2.14. Therefore, $E[X_{T-}] = E[X_T]$ if and only if $X_{T-} = E[X_T|\mathcal{F}_{T-}]$. $\square$

Remark 9.2.4. Clearly, a right-continuous uniformly integrable martingale is regular. Therefore, a right-continuous uniformly integrable supermartingale is regular if and only if the potential in its Riesz decomposition is regular.

If X is a supermartingale, then X is regular if and only if, up to indistinguishability,

$$(\Pi_p X)_t = X_{t-},$$

as then, for every $T \in \mathcal{T}_p$,

$$E[X_{T-}] = E[(\Pi_p X)_T] = E[X_T].$$

In particular, X is regular if X is continuous.

Lemma 9.2.5. *Suppose X is a right-continuous supermartingale of class (D), and A is the predictable increasing process in its Doob–Meyer decomposition. Then A is continuous if and only if X is regular.*

Proof. We have $X = M - A$, where $A \in \mathcal{A}_0^+$ and M is a uniformly integrable martingale. For every predictable stopping time T,

$$M_T = X_T + A_T \qquad \text{and} \qquad M_{T-} = X_{T-} + A_{T-}.$$

However, $E[M_T] = E[M_{T-}]$, so

$$E[A_T - A_{T-}] = E[X_T - X_{T-}].$$

From Theorem 7.5.4 and Exercise 7.7.1, as A is predictable, the set $\{A_t \neq A_{t-}\}$ is thin and predictable. Therefore, A is a.s. continuous if and only if it has no predictable jumps, that is, if and only if $E[X_T - X_{T-}] = 0$. □

Remark 9.2.6. In particular, if X is a continuous supermartingale of class (D), then it is certainly regular, and the processes in its Doob–Meyer decomposition are continuous.

By localization, we now obtain the Doob–Meyer decomposition of a general right-continuous local supermartingale.

Theorem 9.2.7 (Doob–Meyer Decomposition: Local Supermartingales). *Suppose X is a right-continuous local supermartingale. Then X has a unique Doob–Meyer decomposition, that is, a decomposition of the form*

$$X = M - A$$

where $A \in (\mathcal{A}_0^+)_{\mathrm{loc}}$ and is predictable, and $M \in \mathcal{M}_{\mathrm{loc}}$. In other words, A is a nondecreasing predictable process with $A_0 = 0$ a.s., and there is a nondecreasing sequence $\{T_n\}_{n\in\mathbb{N}}$ of stopping times such that $\lim_n T_n = \infty$ a.s., A_{T_n} is integrable and each stopped process M^{T_n} is a uniformly integrable martingale.

Proof. Uniqueness. This is similar to the proof of uniqueness of Theorem 9.1.6 above. Suppose

$$X = M - A = N - B$$

are two such decompositions. Then there is an increasing sequence $\{S_n\}_{n\in\mathbb{N}}$ of stopping times such that $S_n \to \infty$ and, for each n, we have $M^{S_n} - N^{S_n} = A^{S_n} - B^{S_n}$ is both a martingale and a predictable process of integrable variation. Therefore, by Corollary 8.2.11,

$$M^{S_n} - N^{S_n} = A^{S_n} - B^{S_n} = 0.$$

Letting $n \to \infty$, we see that $M = N$ and $A = B$.

Existence. First suppose X is a supermartingale stopped at time n. As X is a supermartingale, we can define the nonnegative process Z by

$$Z_t = X_t - E[X_n|\mathcal{F}_t] \geq 0,$$

so that Z is a potential. For each positive integer k, write

$$S_k = \inf\{t : Z_t \geq k\} \wedge k.$$

Note that $P(S_k = 0)$ can be positive, but $\lim_k S_k = \infty$ a.s. The process $I_{[\![0,S_k[\![}Z$ is a potential, and is uniformly bounded by k, so is of class (D). Applying the result of Theorem 5.3.1, $I_{[\![0,S_k[\![}Z$ has a Doob–Meyer decomposition, that is, there is a predictable increasing process $B^{(k)} \in \mathcal{A}_0^+$ such that

$$I_{[\![0,S_k[\![}Z_t = E[B_\infty^{(k)}|\mathcal{F}_t] - B_t^{(k)}.$$

This gives the equation, for $t \leq n$,

$$\begin{aligned} X_t^{(S_k)} &= E[X_n|\mathcal{F}_t]^{S_k} + Z_t^{S_k} \\ &= E[X_n|\mathcal{F}_t]^{S_k} + E[B_\infty^{(k)}|\mathcal{F}_t] - B_t^{(k)} + Z_{S_k} I_{[\![S_k,n[\![} \\ &= \tilde{M}_t^{(k)} - \tilde{A}_t^{(k)}, \end{aligned}$$

where $\tilde{M}_t^{(k)} = E[X_n|\mathcal{F}_t]^{S_k} + E[B_\infty^{(k)}|\mathcal{F}_t]$ is a martingale and $\tilde{A}_t^{(k)} = B_t^{(k)} - Z_{S_k} I_{[\![S_k,n[\![}$ is a process of integrable variation. From Corollary 8.2.12, we know that $\tilde{A}^{(k)} - \Pi_p^* \tilde{A}^{(k)}$ is a martingale. Writing $M^{(k)} = \tilde{M}^{(k)} - \tilde{A}^{(k)} + \Pi_p^* \tilde{A}^{(k)}$ and $A^{(k)} = \Pi_p^* \tilde{A}^{(k)}$, we conclude that $X^{S_k} = M^{(k)} - A^{(k)}$ is a Doob–Meyer decomposition for X^{S_k}. As X is a supermartingale, the optional stopping theorem can be used to check that $B^{(k)}$ is a nondecreasing process.

By uniqueness, if we can find a Doob–Meyer decomposition $X^{T_k} = M^{(k)} - A^{(k)}$ for a sequence of stopping times $T_k \to \infty$ a.s., then these decompositions must be consistent (i.e. $M^{(k)} - M^{(m)} = A^{(k)} - A^{(m)} = 0$ a.s. on $[\![0, T_m \wedge T_k]\!]$). Taking $M = \lim_k M^{(k)}$ and $A = \lim_k A^{(k)}$ we have the Doob–Meyer decomposition of X. Therefore, we know that any right continuous supermartingale stopped at n admits a Doob–Meyer decomposition.

Now suppose X is a general right-continuous local supermartingale. Then there exists a sequence of stopping times T_n such that $T_n \to \infty$ and X^{T_n} is a supermartingale. Replacing T_n by $T_n \wedge n$, we see that $X^{T_n \wedge n}$ is a supermartingale stopped at n, and so admits a Doob–Meyer decomposition. As the sequence $T_n \wedge n \to \infty$ a.s., as earlier, we can paste these decompositions together, which guarantees that X admits a Doob–Meyer decomposition. $\square$

Remark 9.2.8. While we have given our results for supermartingales, simply multiplying by -1 yields a corresponding decomposition $X = M + A$, where $M \in \mathcal{M}_{\mathrm{loc}}$, $A \in (\mathcal{A}_0^+)_{\mathrm{loc}}$, for X a right-continuous local submartingale.

Remark 9.2.9. For X a càdlàg supermartingale, we have $A = M - X$, and therefore A is also right-continuous. As A is predictable, it follows from Lemma 7.3.20 that A is a locally bounded process.

9.3 Local Time of Brownian Motion

In order to present a nontrivial example of the Doob–Meyer decomposition, we consider a particularly interesting quantity related to the zeros of Brownian motion. As is shown in Appendix A.5.1, this is a rather pathological set, as it is almost surely closed, uncountable, and of Lebesgue measure zero.

Definition 9.3.1. *Let X be a standard one-dimensional Brownian motion. Then, for any $a \in \mathbb{R}$, the process $\{|X_t - a|\}_{t\in[0,\infty[}$ is a submartingale, with Doob–Meyer decomposition*

$$|X_t - a| = M_t + L_t^a,$$

where $L^a \in (\mathcal{A}_0^+)_{\mathrm{loc}}$ is a nondecreasing predictable process, called the local time of X at a.

As $|X - a|$ is *not* a martingale (for example, with $a = 0$ we see $|X_0| = 0$ but $|X_t| > 0$ a.s. for all $t > 0$), the process L^a is not almost surely everywhere constant. A depiction of a path of L^0 is given in Fig. 9.2.

Theorem 9.3.2. *For any $a \in \mathbb{R}$, the process L^a is constant on the set $\{X_t \neq a\}$.*

Proof. As $X - a$ is also a Brownian motion (starting at $-a$), it is enough to consider the case $a = 0$. For any $\epsilon > 0$, we define a sequence of stopping times

$$\begin{aligned} S_n &= \inf\{t > T_{n-1} : |X_t| \geq \epsilon\} \wedge n, \\ T_n &= \inf\{t > S_n : |X_t| = 0\} \wedge n, \end{aligned}$$

with T_0 an arbitrary stopping time. Let $b_n = \mathrm{sign}(X_{S_n})$. By continuity of X and the optional stopping theorem, the process defined by

$$b_n I_{\{S_n<t\}} E\big[X_{t\wedge T_n} - X_{S_n}\big|\mathcal{F}_{t\wedge T_n}\big] = |X_{t\wedge T_n}| - |X_{t\wedge S_n}|$$

is a martingale. From the Doob–Meyer decomposition, we know that

$$|X_{t\wedge T_n}| - L_{t\wedge T_n} - |X_{t\wedge S_n}| + L_{t\wedge S_n}$$

defines a martingale, and taking a difference, $L_{t\wedge T_n} - L_{t\wedge S_n}$ is a predictable increasing process which is a martingale. Hence it is indistinguishable from

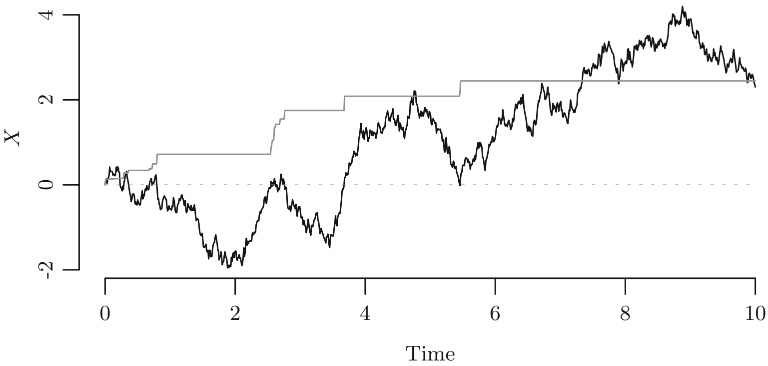

Fig. 9.2. A path of a Brownian motion, together with its (approximate) local time at zero in gray. Note that the local time only changes when the Brownian motion is zero (indicated by the dashed line).

the zero process (Corollary 8.2.14), and taking $t > n$ we see $L_{S_n} = L_{T_n}$ almost surely, so L is constant on the set $]\!]S_n, T_n]\!]$.

Therefore, for any $\epsilon > 0$, L is not increasing between the first time $|X|$ hits ϵ after T_0, and the next time $X_t = 0$. Taking $\epsilon \downarrow 0$, we conclude that L cannot be increasing on the set $\cup_{\epsilon>0}\{|X_t| > \epsilon\} = \{|X_t| > 0\} = \{X_t \neq 0\}$. □

Lemma 9.3.3. *For any $a \in \mathbb{R}$, the process L^a is almost surely continuous.*

Proof. As $\{|X_t - a|\}_{t\in[0,\infty[}$ is a continuous process, we see that the jumps of M and of L^a in its Doob–Meyer decomposition must cancel. As L^a is predictable, this implies that the jump times of M are predictable, and by 5.6.13, we have $M_T - M_{T-} = M_T - E[M_T|\mathcal{F}_{T-}]$ for all bounded predictable stopping times T. Taking an expectation, for T any bounded jump time of L^a, we have

$$E[L^a_T - L^a_{T-}] = E[M_T - M_{T-}] = E\big[M_T - E[M_T|\mathcal{F}_{T-}]\big] = 0$$

and as $L^a_T - L^a_{T-} \geq 0$, we see that L^a is continuous. □

Lemma 9.3.4. *For any $a \in \mathbb{R}$, the path $t \mapsto L^a_t(\omega)$ is almost surely not differentiable with respect to t.*

Proof. We know (Appendix A.5.1) that the zeros of X are a.s. a Lebesgue-null set, and L^a is constant except on the set $\{X_t = a\}$. As L^a is not a constant process, it follows that L^a is not absolutely continuous with respect to t (otherwise, by the Radon–Nikodym theorem, we could write $L_t^a = \int_{[0,t]} g_s ds$ for some g). This implies that L is a.s. not classically differentiable. □

We shall see that the local time makes another appearance when we come to the theory of stochastic integration (Section 14.3). We shall also prove further properties of the local time using this theory.

9.4 Exercises

Exercise 9.4.1. Give a careful statement of the Doob–Meyer decomposition for local submartingales.

Exercise 9.4.2. Let $X \in \mathcal{A}_0^+$. Then X is a submartingale, with Doob–Meyer decomposition $X = M + A$. Show that $M \in \mathcal{A}_0$ and $A \in \mathcal{A}_0^+$. Under what conditions does $X = A$ (up to indistinguishability)?

Exercise 9.4.3. Let $X \in \mathcal{A}_0^+$. Then X is a submartingale, with Doob–Meyer decomposition $X = M + A$. For any nonnegative bounded predictable process Y, show that $\{(Y \bullet X)_t\}$ is a submartingale with Doob–Meyer decomposition $(Y \bullet X) = (Y \bullet M) + (Y \bullet A)$. Show that this is not necessarily the case if Y is not predictable.

Exercise 9.4.4. Let $C(t)$ be an increasing continuous process such that $C(t)$ is a stopping time for every t, and X a Brownian motion. Let $Y_t = (X_{C(t)})^2$. Find the Doob–Meyer decomposition of Y in the filtration given by $\tilde{\mathcal{F}}_t = \mathcal{F}_{C(t)}$. (Hint: Compare with Exercise 5.7.11.)

Exercise 9.4.5. Let X be a local martingale satisfying $X_t > C$ a.s. for all t, for some constant $C \in \mathbb{R}$, and such that X_0 is integrable. Use Fatou's lemma to show that X is a supermartingale.

Exercise 9.4.6. Let X be a potential, and hence a local supermartingale, with Doob–Meyer decomposition $X = M - A$. For T a stopping time, show that, if $A = A^T$, then $X = X^T$. Give a counterexample if X is only a supermartingale (rather than a potential).

10

The Structure of Square Integrable Martingales

We assume, as in previous chapters, that we are working on a probability space $(\Omega, \mathcal{F}, P)$ which has a filtration $\{\mathcal{F}_t\}_{t\in[0,\infty]}$ satisfying the usual conditions and, for simplicity, $\mathcal{F}_\infty = \bigvee_{t<\infty} \mathcal{F}_t$. Furthermore, indistinguishable processes will be identified, so that when we speak of a process, we really mean an equivalence class of indistinguishable processes. When we speak of a martingale, we shall invariably mean its càdlàg version.

10.1 The $\mathcal{H}^p$ Space

Definition 10.1.1. *If X is any càdlàg process, the càdlàg process X^* is defined by*

$$X_t^* = \sup_{s\leq t} |X_t|.$$

If X is progressive but not càdlàg, then X^* is defined by

$$X_t^* = \lim_{s\downarrow t} \big\{ \sup_{u\leq s} |X_t| \big\},$$

and so is càdlàg and optional (given the right-continuity of $\{\mathcal{F}_t\}_{t\geq 0}$).

Recall that $\mathcal{M}$ is the space of uniformly integrable martingales, that $\mathcal{M}_{\text{loc}}$ is the space of local martingales, and that $\mathcal{M}_{0,\text{loc}}$ is the space of local martingales $\{X \in \mathcal{M}_{\text{loc}} : X_0 = 0\}$. We know that when considering random variables it is useful to work with the spaces L^p; we shall see that it is also convenient to have analogous spaces of martingales, which leads to the definition of the spaces $\mathcal{H}^p$.

S.N. Cohen, R.J. Elliott, *Stochastic Calculus and Applications*,
Probability and Its Applications, DOI 10.1007/978-1-4939-2867-5_10

Definition 10.1.2. *For M a martingale and $p \in [1,\infty[$, write*

$$\|M\|_{\mathcal{H}^p} := \|M^*_\infty\|_p = E[\sup_t |M_t|^p]^{1/p}.$$

Here $\|\cdot\|_p$ denotes the norm in L^p. Then $\mathcal{H}^p$ is the space of martingales such that

$$\|M\|_{\mathcal{H}^p} < \infty.$$

In a natural way, we have the spaces $\mathcal{H}^p_{\mathrm{loc}}$ for processes locally in $\mathcal{H}^p$, and $\mathcal{H}^p_0$ for processes X in $\mathcal{H}^p$ with $X_0 = 0$. Note that, for now, we exclude the case $p = \infty$, which will be defined later (see Remark 11.5.9).

Lemma 10.1.3. *$\mathcal{H}^p$ has the following properties.*

(i) If $p' \leq p$, then $\mathcal{H}^p \subset \mathcal{H}^{p'}$ and $\|M\|_{\mathcal{H}^{p'}} \leq \|M\|_{\mathcal{H}^p}$.
(ii) If $1 < p < \infty$ and $M \in \mathcal{M}$, then

$$\|M_\infty\|_p \leq \|M^*_\infty\|_p \leq q\|M_\infty\|_p\,, \quad \textit{where} \quad p^{-1} + q^{-1} = 1.$$

Proof. (i) If $p' \leq p$, then, using Jensen's inequality, $\|M^*_\infty\|_{p'} \leq \|M^*_\infty\|_p$. That $\mathcal{H}^p \subset \mathcal{H}^{p'}$ follows.
(ii) The first inequality is trivial and the second is the result of Doob's L^p inequality (Theorem 5.1.3) and Jensen's inequality.

□

Lemma 10.1.4. *For all $p \in [1,\infty[$, we have $\mathcal{H}^p \subset \mathcal{M}$ and, if $M \in \mathcal{H}^p$, then $|M_t|^p \to |M_\infty|^p$ in L^1.*

Proof. For $M \in \mathcal{H}^p$, we know M^*_∞ is integrable, and $|M_T| \leq |M^*_\infty|$ for all stopping times T. Therefore, the set $\{|M_T|\}_{T\in\mathcal{T}}$ is uniformly integrable, and so the martingale M is of class (D). That $M \in \mathcal{M}$ follows from Lemma 5.6.6.

As M is uniformly integrable, we know M_∞ exists and, as $|M_t|^p \leq (M^*_\infty)^p$ and $(M^*_\infty)^p$ is integrable, the set $\{|M_t|^p\}_{t\geq 0}$ is uniformly integrable. As $|M_t|^p$ converges almost surely to $|M_\infty|^p$, this implies $|M_t|^p \to |M_\infty|^p$ in L^1. □

Lemma 10.1.5. *Identifying indistinguishable martingales, $\mathcal{H}^p$ is a Banach space under the $\mathcal{H}^p$ norm, for all $p \in [1,\infty[$.*

Proof. The only nontrivial statement to verify is completeness. Suppose $\{X^n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathcal{H}^p$. By taking a subsequence, we can assume that $X^n(\omega)$ converges uniformly in t for almost all ω, and hence there exists an adapted càdlàg process X such that $E[\sup_t |X^n_t - X_t|^p] \to 0$. We need to verify that X is a martingale.

For every t, we also know that $X^n_t = E[X^n_\infty|\mathcal{F}_t]$, as $X^n \in \mathcal{M}$. From the above, $X^n_\infty \to X_\infty$ in L^p, and hence $X^n_t \to E[X_\infty|\mathcal{F}_t]$ in L^p for every t. It follows that X is a càdlàg modification of $\{E[X_\infty|\mathcal{F}_t]\}_{t\geq 0}$, so X is a martingale. Therefore, $X^n \to X$ in $\mathcal{H}^p$. □

Remark 10.1.6. Note that $\mathcal{H}^p$ requires in its definition that $E\big[\sup_t |M|_t^p\big] < \infty$, not just that $\sup_t E\big[|M|_t^p\big] < \infty$. In other works (see, for example, [131]), the spaces $\{\mathcal{H}^p\}_{1<p<\infty}$ are sometimes defined as those martingales for which

$$\sup_t E\big[|M_t|^p\big] < \infty.$$

By Jensen's inequality (Lemma 2.4.11), we know that, for all such martingales, $\{|M_t|^p\}_{t\in[0,\infty]}$ is a nonnegative submartingale. Therefore, Doob's L^p inequality, as expressed in Theorem 5.1.3, yields

$$\|M_\infty^*\|_p = E\big[\sup_t |M_t|^p\big]^{1/p} \leq q \sup_t E\big[|M_t|^p\big]^{1/p} < \infty,$$

where $p^{-1} + q^{-1} = 1$. Hence the definitions are equivalent for $p > 1$, on the space of martingales.

However, this is only true as the processes considered are martingales, and in general fails for local martingales. For example, we may have $\sup_t E\big[|M_t|^2\big] < \infty$ but $E\big[\sup_t |M_t|^2\big] = \infty$. See Exercise 14.7.11 for an example of this. For this reason, the definition $\|M\|_{\mathcal{H}^p} = E[\sup_t |M_t|^p]^{1/p}$ is preferable, as it implies (Exercise 10.4.6) that a local martingale M with $\|M\|_{\mathcal{H}^p} < \infty$ is necessarily a true martingale in $\mathcal{H}^p$.

The following result, from Dellacherie and Meyer [54], provides a simple way of approximating martingales in $\mathcal{H}^1$. Given all martingales are locally in $\mathcal{H}^1$ (Corollary 10.3.8), this gives a very general approximation result.

Theorem 10.1.7. *The space of bounded martingales (and hence $\mathcal{H}^p$ for any $p > 1$) is dense in $\mathcal{H}^1$.*

To prove this, we prove two useful lemmata.

Lemma 10.1.8. *Let $X \in \mathcal{H}^1$, and let $\{T_k\}_{k\in\mathbb{N}}$ be an increasing sequence of stopping times with $T_k \to \infty$ a.s. Then the stopped processes converge $X^{T_k} \to X$ in $\mathcal{H}^1$ as $k \to \infty$.*

Proof. We know $(X - X^{T_n})_\infty^* = \sup_{t\geq T_k} |X_t - X_{T_n}| \leq 2X_\infty^*$ and X_∞^* is integrable. As $X_{T_n} \to X_\infty$ almost surely as $n \to \infty$, by Doob's inequalities (Theorem 5.1.2) and dominated convergence, we know

$$\lambda P\big(\sup_t |X_t - X_t^{T_n}| > \lambda\big) \leq E\big[|X_\infty - X_\infty^{T_n}|\big] \to 0$$

for any $\lambda > 0$. Therefore, $(X - X^{T_n})_\infty^* \to 0$ in probability, and the result follows by dominated convergence. □

Lemma 10.1.9. *Let X^n and X be uniformly integrable martingales such that $E\big[|X_\infty^n|\big]$ and $E\big[|X_\infty|\big]$ are uniformly bounded, and such that $E\big[|X_\infty^n - X_\infty|\big] \leq 4^{-n}$. There exists a sequence of stopping times $\{T_k\}_{k\in\mathbb{N}}$ with $T_k \to \infty$ such that $(X^n)^{T_k} \in \mathcal{H}^1$ for all k, and $(X^n)^{T_k} \to X^{T_k}$ in $\mathcal{H}^1$. (To express this differently, $X^n \to X$ locally in $\mathcal{H}^1$.)*

Proof. Let $S_k := \inf\{t : |X_t| \geq k\}$, so $X^{S_k} \in \mathcal{H}^1$ and $S_k \uparrow \infty$. Define $Y^n := X^n - X$. Then, from Doob's maximal inequality (Theorem 5.1.2), we have, for any $\lambda > 0$,

$$\lambda P\big((Y^n)^*_\infty \geq \lambda\big) \leq 4^{-n}.$$

Therefore, $\sum_n P\big((Y^n)^*_\infty \geq 2^{-n}\big) < \infty$. By the Borel–Cantelli lemma, the increasing process $C_t = \sum_n (Y^n)^*_t$ is therefore finite valued, and the stopping times $R_k := \inf\{t : C_t \geq k\} \uparrow \infty$. Let $T_k := R_k \wedge S_k$. We know

$$(Y^n)^*_{T_k} \leq (Y^n)^*_{T_k-} \vee |(Y^n)_{T_k}| \leq k + |X^n_{T_k} - X_{T_k}|.$$

Therefore, $(Y^n)^{T_k} \in \mathcal{H}^1$, and we see $(X^n)^{T_k} \in \mathcal{H}^1$ for all k. From our assumptions, we see that $|X^n_{T_k} - X_{T_k}| \to 0$ in L^1, and it follows that $\{|X^n_{T_k} - X_{T_k}|\}_{k\in\mathbb{N}}$ is uniformly integrable. Therefore, $\{(Y^n)^*_{T_k}\}_{k\in\mathbb{N}}$ is uniformly integrable. From our above estimate of $P\big((Y^n)^*_\infty \geq \lambda\big)$, we know that $(Y^n)^* \to 0$ in probability, and therefore in L^1. □

Proof of Theorem 10.1.7.
Let $X \in \mathcal{H}^1$. Choose a sequence of bounded martingales $\{X^n\}_{n\in\mathbb{N}}$ such that $\|X^n_\infty - X_\infty\| \leq 4^{-n}$. (This exists because bounded functions are dense in L^1.) Using Lemma 10.1.9, we construct a sequence of stopping times $\{T_k\}_{k\in\mathbb{N}}$, and we see that $\{X^{n,k}\}_{n,k\in\mathbb{N}} = \{(X^n)^{T_k}\}_{n,k\in\mathbb{N}}$ is a family of bounded martingales with

$$\|X^{n,k} - X\|_{\mathcal{H}^1} \leq \|X^{n,k} - X^{T_k}\|_{\mathcal{H}^1} + \|X^{T_k} - X\|_{\mathcal{H}^1}.$$

Taking k and then n sufficiently large, and using Lemmata 10.1.8 and 10.1.9, we can select a Cauchy sequence of bounded martingales converging to X in $\mathcal{H}^1$. □

Theorem 10.1.10. *Suppose $1 < p < \infty$, and that $\{M^n\}_{n\in\mathbb{N}}$ is a sequence of martingales which converge in $\mathcal{H}^p$ to the martingale M. Then there is a subsequence $\{M^{n_k}\}_{k\in\mathbb{N}}$ such that, for almost every $\omega \in \Omega$, $M^{n_k}_t(\omega)$ converges uniformly to $M_t(\omega)$ on $[0, \infty]$.*

Proof. By definition,

$$\lim_n \|M^n - M\|_{\mathcal{H}^p} = \lim_n \|(M^n - M)^*_\infty\|_p = 0.$$

Therefore (cf. Lemma 1.3.38, Exercise 2.7.8), for almost any $\omega \in \Omega$, there is a subsequence $\{n_k\}_{k\in\mathbb{N}}$ such that

$$\sup_t |M^{n_k}_t(\omega) - M_t(\omega)| = 0 \text{ a.s.}$$

□

Remark 10.1.11. As in Lemma 5.5.6, the limit in $\mathcal{H}^p$ of a sequence of continuous martingales is, therefore, a continuous martingale, and the jumps of the limit are the limits of the jumps of an approximating sequence.

10.1.1 The Space $\mathcal{H}^2$.

Remark 10.1.12. From Lemma 10.1.3(ii) we see that if, $1 < p < \infty$, the norm $\|M_\infty\|_p$ is equivalent to the norm $\|M\|_{\mathcal{H}^p}$, and so $\mathcal{H}^p$ can be identified with the Banach space $L^p(\Omega, \mathcal{F}, P)$ by the map which associates $M \in \mathcal{H}^p$ with its terminal value $M_\infty \in L^p$. In particular, the map

$$M, N \mapsto E[M_\infty N_\infty]$$

forms an inner product on $\mathcal{H}^2$ inducing the desired topology. In general, we see that for $p \in]1, \infty[$, this association shows that the dual of $\mathcal{H}^p$ is given by $\mathcal{H}^q$, where $p^{-1} + q^{-1} = 1$, and any continuous linear functional ϕ on $\mathcal{H}^p$ can be written $\phi(M) = E[M_\infty N_\infty]$ for some $N \in \mathcal{H}^q$. The space $\mathcal{H}^1$ requires more careful analysis and is discussed in Appendix A.8.

Definition 10.1.13. *For convenience, processes in $\mathcal{H}^2$ will be called* square integrable martingales*.*

The following definition allows us to consider the 'geometry' of the space of local martingales.

Definition 10.1.14. *Two local martingales M, N in $\mathcal{M}_{\mathrm{loc}}$ are* orthogonal *if their product $MN = \{M_t N_t\}_{t \geq 0}$ is in $\mathcal{M}_{0,\mathrm{loc}}$. We shall then write $M \perp N$. Note that, in particular, orthogonality implies that*

$$M_0 N_0 = 0 \quad \text{a.s.}$$

Remark 10.1.15. Note that this definition implies that, if M and N are orthogonal, then there exists a 'localizing' sequence of stopping times T_n such that

$$E[M_{t \wedge T_n} N_{t \wedge T_n}] = 0$$

for all t and all n.

We now see that this is a 'stronger' geometry than that induced by associating $\mathcal{H}^2$ with L^2, as in Remark 10.1.12, as orthogonality of martingales M, N implies the random variables M_T, N_T are orthogonal in L^2 for *every* stopping time T.

Lemma 10.1.16. *Suppose $M, N \in \mathcal{H}^2$ are orthogonal. Then $MN \in \mathcal{H}_0^1$ (that is, MN is a uniformly integrable martingale with $E[|(MN)_\infty^*|] < \infty$ and $M_0 N_0 = 0$). In particular, for every stopping time $T \in \mathcal{T}$, the random variables M_T and N_T are orthogonal in L^2 (that is, $E[M_T N_T] = 0$).*

Conversely, if $M_0 N_0 = 0$ a.s. and the random variables M_T and N_T are orthogonal in L^2 for every $T \in \mathcal{T}$, then M and N are orthogonal.

Proof. Suppose $M, N \in \mathcal{H}^2$, so M_∞^* and N_∞^* are in L^2. Then their product $M_\infty^* N_\infty^*$ is in L^1, by the Cauchy–Schwarz inequality. Now

$$(MN)_\infty^* = \sup_t |M_t N_t| \leq M_\infty^* N_\infty^*,$$

so $MN \in \mathcal{H}_0^1$ as M and N are orthogonal. In particular, $MN \in \mathcal{M}$, the product is uniformly integrable and, for any $T \in \mathcal{T}$, we have $E[M_T N_T] = E[M_0 N_0] = 0$.

Conversely, suppose that for any $T \in \mathcal{T}$, we have $M_T N_T \in L^1$ so $E[|M_T N_T|] < \infty$, and $E[M_T N_T] = 0$. Therefore, by Theorem 5.4.6, MN is a uniformly integrable martingale. We know $M_0 N_0 = 0$ a.s., and hence $MN \in \mathcal{M}_{0,\text{loc}}$. □

Remark 10.1.17. For this reason, and to distinguish between different types of orthogonality, local martingales M and N with $MN \in \mathcal{M}_{0,\text{loc}}$ are sometimes called 'very strongly orthogonal', while if $E[M_\infty N_\infty] = 0$ we say they are 'weakly orthogonal'. We will use 'orthogonal' to mean 'very strongly orthogonal' for simplicity. Exercise 10.4.1 will show that these notions are indeed not equivalent.

10.1.2 Stable Subspaces

The following definition gives us a useful notion of 'closedness' of a set of processes, which is well behaved when we wish to work with stopped processes.

Definition 10.1.18. *A subspace $\mathcal{K} \subset \mathcal{H}^p$ is said to be* stable *if:*

(i) it is closed under the $\mathcal{H}^p$-norm topology,
(ii) it is closed under stopping, that is $T \in \mathcal{T}$ and $M \in \mathcal{K}$ imply $M^T \in \mathcal{K}$,
(iii) if $M \in \mathcal{K}$ and $A \in \mathcal{F}_0$ then $I_A M \in \mathcal{K}$.

Theorem 10.1.19. *Suppose $\mathcal{K}$ is a stable subspace of $\mathcal{H}^2$ and write $\mathcal{K}^\perp$ for the set of martingales $N \in \mathcal{H}^2$ such that $E[M_\infty N_\infty] = 0$ for all $M \in \mathcal{K}$. Then $\mathcal{K}^\perp$ is a stable subspace and, if $M \in \mathcal{K}$ and $N \in \mathcal{K}^\perp$, then M and N are orthogonal.*

Proof. Consider $M \in \mathcal{K}$, $N \in \mathcal{K}^\perp$ and $T \in \mathcal{T}$. Then $E[L_\infty N_\infty] = 0$ for all $L \in \mathcal{K}$ and $\mathcal{K}$ is closed under stopping. Taking $L = M^T \in \mathcal{K}$,

$$E[M_T N_\infty] = 0.$$

Thus,

$$\begin{aligned} E[M_T N_\infty] &= E\big[E[M_T N_\infty | \mathcal{F}_T]\big] \\ &= E\big[M_T E[N_\infty | \mathcal{F}_T]\big] = E[M_T N_T] = 0. \end{aligned}$$

Taking $T = 0$ and $A \in \mathcal{F}_0$,

$$E[I_A M_0 N_0] = 0 \quad \text{so} \quad M_0 N_0 = 0 \quad \text{a.s.},$$

and we see M and N are orthogonal. Furthermore,

$$E[I_A M_T N_T] = E[M_\infty (I_A N^T)_\infty] = 0,$$

which implies that $I_A N^T \in \mathcal{K}^\perp$ for any $N \in \mathcal{K}^\perp$, $T \in \mathcal{T}$ and $A \in \mathcal{F}_0$.

Finally, note that for any sequence $\{N^n\}_{n\in\mathbb{N}} \subset \mathcal{K}^\perp$ which converges in $\mathcal{H}^2$-norm to $N \in \mathcal{H}^2$, by Remark 10.1.12 we know that $N^n_\infty \to N_\infty$ in L^2. Hence for any $M \in \mathcal{K}$ we have

$$\begin{aligned}E[N_\infty M_\infty] &= E[N^n_\infty M_\infty] + E[(N_\infty - N^n_\infty)M_\infty] = 0 + E[(N_\infty - N^n_\infty)M_\infty]\\ &\leq \|N_\infty - N^n_\infty\|_2 \|M\|_2 \to 0.\end{aligned}$$

Therefore $N \in \mathcal{K}^\perp$ that is, $\mathcal{K}^\perp$ is closed in the $\mathcal{H}^2$ topology. Consequently, $\mathcal{K}^\perp$ is stable. □

Corollary 10.1.20. *Suppose $\mathcal{K} \subset \mathcal{H}^2$ is a stable subspace. Then every element $M \in \mathcal{H}^2$ has a unique decomposition*

$$M = N + N',$$

where $N \in \mathcal{K}$ and $N' \in \mathcal{K}^\perp$. Equivalently (as N is orthogonal to itself only if it is zero) we can write $\mathcal{H}^2 = \mathcal{K} \oplus \mathcal{K}^\perp$.

Proof. Suppose $\mathcal{K}_\infty$ is the closed subspace of L^2 generated by the random variables $\{M_\infty : M \in \mathcal{K}\}$ and, similarly, $\mathcal{K}^\perp_\infty$ is the closed subspace generated by $\{M_\infty : M \in \mathcal{K}^\perp\}$. By standard results on projections in Hilbert spaces (Lemma 1.5.21), for any $M \in \mathcal{H}^2$, M_∞ has a unique decomposition (up to equality a.s.)

$$M_\infty = N_\infty + N'_\infty,$$

where $N_\infty \in \mathcal{K}_\infty$ and $N'_\infty \in \mathcal{K}^\perp_\infty$. Then N (resp. N') is the càdlàg version of the martingale defined by

$$N_t = E[N_\infty|\mathcal{F}_t] \quad (\text{resp. } N'_t = E[N'_\infty|\mathcal{F}_t]).$$

N and N' are orthogonal by Theorem 10.1.19. □

10.2 The Space of Pure-Jump Martingales $\mathcal{H}^{2,d}$

Remark 10.2.1. If $\{X_t\}_{t\in[0,\infty]}$, is any process, we shall follow the notational convention introduced in Chapter 8 that $X_{0-} = 0$ a.s. unless indicated otherwise, so $X_0 = \Delta X_0$. However, when we say a process is continuous, we shall interpret that as continuity on the right at zero, so we do not require $X_0 = 0$ also.

Definition 10.2.2. *$\mathcal{H}^{2,c}_0 \subset \mathcal{H}^2$ will denote the space of continuous square integrable martingales with $M_0 = 0$.*

By Theorem 10.1.10, $\mathcal{H}_0^{2,c}$ is topologically closed, and clearly is closed under stopping and is a vector space, so $\mathcal{H}_0^{2,c}$ is stable.

Definition 10.2.3. *We define $\mathcal{H}^{2,d}$ to be the stable subspace orthogonal to $\mathcal{H}_0^{2,c}$, that is $\mathcal{H}^{2,d} = (\mathcal{H}_0^{2,c})^{\perp}$. Martingales in $\mathcal{H}^{2,d}$ are said to be* purely discontinuous, as they are orthogonal to every continuous local martingale.

We shall determine the structure of $\mathcal{H}^{2,d}$ by studying certain simple subspaces.

Definition 10.2.4. *Suppose $T \in \mathcal{T}$ is a stopping time. $\mathcal{H}_{(T)}^{2,d}$ will denote the space of martingales in $\mathcal{H}^{2,d}$ which are continuous outside the graph of T, and satisfy $M_0 = 0$ on the set $\{T > 0\}$. Note that $\mathcal{H}_{(T)}^{2,d}$ is a stable subspace.*

Lemma 10.2.5. *$\mathcal{H}_{(0)}^{2,d}$ is the space of constant processes.*

Proof. If $H \in \mathcal{H}_{(0)}^{2,d}$, then $M_t = H_t - H_0$ is a continuous martingale, but as $\mathcal{H}_{(0)}^{2,d} \subset \mathcal{H}^{2,d}$, it is also purely discontinuous. Therefore, $E[M_t^2] = E[M_0^2] = 0$, so $H_t = H_0$ a.s. for all t. □

10.2.1 Martingales of Integrable Variation

Theorem 10.2.6. *Suppose M is a martingale which is also a process of integrable variation. Then*

$$M = M_0 + A - \Pi_p^* A$$

where

$$A_t = \sum_{0<s\le t} \Delta M_s \in \mathcal{A},$$

and $\Pi_p^* A$ is continuous.

Proof. Consider the process

$$B = M - M_0 - A\,,$$

where A is given above. Then $B \in \mathcal{A}_0$ and B is continuous. By Corollary 8.2.8 we have $\Pi_p^* B = B$, so, from Theorem 8.2.11,

$$\Pi_p^* B = \Pi_p^*(M - M_0) - \Pi_p^* A = -\Pi_p^* A,$$

and it follows that

$$M = M_0 + A + B = M_0 + A - \Pi_p^* A.$$

Clearly, $\Pi_p^* A = -B$ is continuous. □

Lemma 10.2.7. *For M a martingale of integrable variation and any bounded càdlàg martingale N,*

$$E[M_\infty N_\infty] = E\Big[M_0 N_0 + \sum_{s>0} \varDelta M_s \varDelta N_s\Big] = E\Big[\sum_{s\geq 0} \varDelta M_s \varDelta N_s\Big].$$

Proof. The process $\{M_t - M_0\}_{t\geq 0}$ is in $\mathcal{A}_0$ and is a martingale, so, from Theorem 8.2.11(iii), the restriction to $\varSigma_p$ of the measure μ associated with $M - M_0$ is zero. The process $\{N_{t-}\}_{t\geq 0}$ is predictable, so

$$E\Big[\int_{[0,\infty[} N_{s-} dM_s\Big] = \int_{[0,\infty[\times \varOmega} N_{s-}(\omega) d\mu = 0.$$

However, by Theorem 7.6.5, the constant process N_∞ has optional projection $(\Pi_o N_\infty)_t = N_t$, so by Theorem 8.2.2,

$$E[M_\infty N_\infty] = E\Big[\int_{[0,\infty[} N_\infty dM_s\Big] = E\Big[\int_{[0,\infty[} N_s dM_s\Big],$$

recalling $M_{0-} = 0$. Subtracting, we find

$$E[M_\infty N_\infty] = E\Big[\int_{[0,\infty[} \varDelta N_s dM_s\Big] = E\Big[\sum_{s\geq 0} \varDelta M_s \varDelta N_s\Big].$$

Note that if $M_0 \neq 0$, $N_0 \neq 0$ this sum includes the term $\varDelta M_0 \varDelta N_0 = M_0 N_0$, using the convention $M_{0-} = N_{0-} = 0$. □

Corollary 10.2.8. *For any M, N as in Lemma 10.2.7 above, writing*

$$L_t := M_t N_t - M_0 N_0 - \sum_{0<s\leq t} \varDelta M_s \varDelta N_s$$

defines a process $L \in \mathcal{M}_0$.

Proof. Applying the lemma to the martingale N^T, stopped at an arbitrary stopping time $T \in \mathcal{T}$,

$$E[M_\infty N_T] = E\Big[\sum_{0\leq s\leq t} \varDelta M_s \varDelta N_s\Big].$$

However,

$$E[M_\infty N_T] = E\big[E[M_\infty | \mathcal{F}_T] N_T\big] = E[M_T N_T],$$

so $E[L_T] = 0$ for any $T \in \mathcal{T}$. It is also easy to verify that $E\big[|L_T|\big] < \infty$, from the assumptions on M and N. Therefore, by Theorem 5.4.6, L is a uniformly integrable martingale with $L_0 = 0$, that is, $L \in \mathcal{M}_0$. □

Corollary 10.2.9. *A martingale of integrable variation M is orthogonal to every continuous local martingale N with $N_0 = 0$.*

Proof. As N is continuous, it is locally bounded and $\Delta N_s \equiv 0$, so by Corollary 10.2.8, $MN \in \mathcal{M}_{0,\text{loc}}$. □

Theorem 10.2.10. *Suppose T is a totally inaccessible stopping time and $\Phi \in L^2(\mathcal{F}_T, P)$. Write, for $t < \infty$,*

$$A_t := \Phi I_{\{t \geq T\}},$$

so that $A \in \mathcal{A}_0$. Then $B := \Pi_p^* A$ is continuous, and $M = A - B$ is a square integrable martingale in $\mathcal{H}^{2,d}_{(T)} \subset \mathcal{H}^{2,d}$.

Proof. It is sufficient to consider the case when $\Phi \geq 0$ a.s., which implies that A is nonnegative and nondecreasing. Note that the fact T is totally inaccessible implies $T > 0$ a.s. From Corollary 8.2.12, the process $M = A - B$ is a martingale.

The measure on Σ_p associated with the dual predictable projection B coincides, by definition, with the measure μ_A associated with A. However, in Σ_o, the support of μ_A is the graph of the totally inaccessible stopping time T. Therefore, A, and hence B, does not charge any predictable stopping time. As B is predictable, Exercise 8.4.1 implies B is continuous. In particular, $B_0 = 0$.

By definition, $A_\infty = \lim_{t\to\infty} A_t = \Phi I_{\{T<\infty\}} \in L^2 \subset L^1$. We know that $E[B_\infty] = E[A_\infty]$, so $B_\infty = \lim_{t\to\infty} B_t \in L^1$ and $B \in \mathcal{A}_0^+$. To show M is square integrable, we must show that $B_\infty \in L^2$. From the proof of Lemma 9.1.4, A and B generate the same potential, that is

$$Z_t = E[A_\infty|\mathcal{F}_t] - A_t = E[B_\infty|\mathcal{F}_t] - B_t \quad \text{a.s.}$$

for all t.

Write N for the càdlàg martingale $\{E[B_\infty|\mathcal{F}_t]\}_{t\geq 0}$. Then the predictable projection of N is just the left continuous process $\{N_{t-}\}_{t\geq 0}$. Because B is continuous, it is predictable, so the predictable projection of the potential Z is given by $(\Pi_p Z)_t = Z_{t-}$.

From Corollary 8.1.22

$$E\Big[\int_{[0,\infty[} N_t dB_t\Big] = E\Big[\int_{[0,\infty[} E[B_\infty|\mathcal{F}_t] dB_t\Big] = E[B_\infty^2].$$

By Theorem 1.3.43, as $B \in \mathcal{A}_0^+$ is almost surely of finite variation and continuous, we know

$$2\int_{[0,t]} B_s dB_s = B_t^2,$$

and so

$$2E\Big[\int_{[0,\infty[} Z_t dB_t\Big] = 2E\Big[\int_{[0,\infty[} (N_t - B_t) dB_t\Big] = E[B_\infty^2].$$

By Corollary 8.2.8 and Theorem 8.2.6, as B is continuous (and hence predictable),

$$E[B_\infty^2] = 2E\Big[\int_{[0,\infty[} Z_{t-}dB_t\Big] = 2E\Big[\int_{[0,\infty[} Z_{t-}dA_t\Big].$$

However, as $Z_{t-} + A_{t-} = E[A_\infty|\mathcal{F}_{t-}]$ for all t, and A is nondecreasing and nonnegative,

$$E\Big[\int_{[0,\infty[} Z_{t-}dA_t\Big] \leq E\Big[\int_{[0,\infty[} (Z_{t-} + A_{t-})dA_t\Big] \leq E\Big[\big(\sup_t E[A_\infty|\mathcal{F}_{t-}]\big)A_\infty\Big]$$

and, by Theorem 5.1.3 and right-continuity,

$$\big\|\sup_t E[A_\infty|\mathcal{F}_{t-}]\big\|_2 = \big\|\sup_t E[A_\infty|\mathcal{F}_t]\,\big\|_2 \leq 2\|A_\infty\|_2.$$

Therefore,

$$\begin{aligned} E[B_\infty^2] &\leq 2E\big[\big(\sup_t E[A_\infty|\mathcal{F}_{t-}]\big)A_\infty\big] \\ &\leq 2E\big[\big(\sup_t E[A_\infty|\mathcal{F}_t]\big)^2\big] \\ &\leq 8E[A_\infty^2] < \infty, \end{aligned}$$

so $B_\infty \in L^2$ and $M \in \mathcal{H}^{2,d}_{(T)} \subset \mathcal{H}^{2,d}_0$. $\square$

Theorem 10.2.11. *Suppose $T > 0$ a.s. is a predictable stopping time and $\varPhi \in L^2(\mathcal{F}_T, P)$ is such that*

$$E[\varPhi|\mathcal{F}_{T-}] = 0 \quad \text{a.s.}$$

Then the process M defined by $M_t = A_t = \varPhi I_{\{t\geq T\}}$ is a square integrable martingale in $\mathcal{H}^{2,d}_{(T)} \subset \mathcal{H}^{2,d}_0$.

Proof. Because $\varPhi I_{\{T=\infty\}}$ is $\mathcal{F}_{T-}$-measurable, we can suppose that $\varPhi = 0$ a.s. on the set $\{T = \infty\}$. By the result of Exercise 7.7.5, if X is any predictable process, then $X_T I_{\{T<\infty\}}$ is $\mathcal{F}_{T-}$-measurable. Therefore,

$$E\Big[\int_{[0,\infty[} X_t dA_t\Big] = E[X_T(A_T - A_{T-})] = E\big[X_T E[\varPhi|\mathcal{F}_{T-}]\big] = 0,$$

so the restriction of the Doléans measure μ_A to Σ_p is zero. Therefore, by Theorem 8.2.11, we know $\Pi_p^* A$ is the zero process and $M = A$ is a martingale. Clearly

$$\|M_\infty^*\|_2 = \|A_\infty\|_2 = \|\varPhi\|_2 < \infty \quad \text{and} \quad M \in \mathcal{H}^{2,d}_{(T)}.$$ $\square$

Corollary 10.2.12. *Suppose, as above, that either*

(i) T *is a totally inaccessible stopping time and* $\Phi \in L^2(\mathcal{F}_T, P)$*, or*
(ii) $T > 0$ *a.s. is a predictable stopping time,* $\Phi \in L^2(\mathcal{F}_T, P)$ *and* $E[\Phi|\mathcal{F}_{T-}] = 0$ a.s.

Let $A_t := \Phi I_{\{t \geq T\}}$ *and write*

$$M = A - \Pi_p^* A\,.$$

Then, for every martingale $N \in \mathcal{H}^2$*, the process* L *defined by*

$$L_t := M_t N_t - M_0 N_0 - \Delta M_T \Delta N_T I_{\{t \geq T\}}$$

is in $\mathcal{M}_0$. *Therefore,* M *is orthogonal to every martingale in* $\mathcal{H}^2$ *which is continuous at* T*. In particular, we know*

$$\{M_t^2 - (\Delta M_T)^2 I_{\{t \geq T\}}\}_{t \geq 0} \in \mathcal{M}_0 \qquad \textit{and} \qquad E[M_\infty^2] = E[(\Delta M)_T^2].$$

Proof. In both cases M is a martingale of integrable variation with a single jump at T, so if N is a bounded càdlàg martingale, by Lemma 10.2.7,

$$E[M_\infty N_\infty] = E[\Delta M_T \Delta N_T] = E[\Phi \Delta N_T].$$

Theorem 10.1.10 and dominated convergence imply that this holds for general $N \in \mathcal{H}^2$ by approximation.

Suppose S is any stopping time. Then $N_\infty^S = N_S$, so applying the above equality to the martingale N^S we have

$$E[M_S N_S] = E[\Delta M_T \Delta N_T I_{\{T \leq S\}}],$$

that is,

$$E[L_S] = 0.$$

Therefore, by Lemma 5.4.6, L is a martingale, is uniformly integrable and $L_0 = 0$ a.s. □

Definition 10.2.13. *When* M *is as in Corollary 10.2.12, we say that it is a* compensated (single) jump martingale.

10.2.2 General Pure-Jump Martingales

In Theorem 10.2.6 it was shown that a martingale of integrable variation is the compensated sum of a series of jumps. This result is now extended to martingales in $\mathcal{H}^{2,d}$.

However, for martingales of integrable variation, as considered in Theorem 10.2.6, the process A obtained by summing all the jumps can be defined and the dual predictable projection $\Pi_p^* A$ obtained directly. In contrast, for general square-integrable martingales in $\mathcal{H}^{2,d}$, the compensated jump martingale for each jump must first be defined, and the convergence of these then discussed in $\mathcal{H}^2$.

This procedure is motivated by the work of Lévy, who considered processes with independent increments, and obtained the following result:

> Suppose X is a real càdlàg process with independent increments. In general $\sum_{0<s\leq t} \Delta X_s$ is not convergent, and the sample paths of X are not of bounded variation. However, for any $\epsilon > 0$ consider the sum of the jumps of size between ϵ and 1 :
>
> $$A^\epsilon_t = \sum_{0<s\leq t} \Delta X_s I_{\{\epsilon \leq |\Delta X_s| \leq 1\}}.$$
>
> Then it can be shown that A^ϵ is adapted and of integrable variation on any finite interval. However, A^ϵ is not predictable, because its jumps are all totally inaccessible. As the increments of X are independent, its dual predictable projection $(\Pi^*_p A^\epsilon)$ is of the form $c_\epsilon t$. As $\epsilon \to 0$, neither c_ϵ nor A^ϵ has a limit, but
>
> $$A^\epsilon_t - (\Pi^*_p A^\epsilon)_t$$
>
> does have a limit in L^2 for every t.

Theorem 10.2.14. *Suppose $M \in \mathcal{H}^{2,d}_0$. Then M is the sum (in $\mathcal{H}^2$) of a series of compensated jump martingales. Furthermore, M is orthogonal to every martingale $N \in \mathcal{H}^2$ which does not charge a common jump time with M.*

Proof. M is adapted and càdlàg, so, from Exercise 7.7.1, the set $\{(t,\omega) : \Delta M_t(\omega) \neq 0\}$ is thin, that is, it is contained in the union of graphs of a sequence of stopping times $\{S_n\}_{n\in\mathbb{N}}$

By restricting S_n to the set $([\![S_n]\!] \setminus \cup_{m<n} [\![S_m]\!])$, we may suppose that the $[\![S_n]\!]$ are disjoint. From Theorem 6.2.9, each $[\![S_n]\!]$ may be written as $[\![S^i_n]\!] \cup [\![S^a_n]\!]$, where S^i_n is totally inaccessible and S^a_n is accessible. Each $[\![S^a_n]\!]$ is contained in a countable union of graphs of predictable stopping times.

Therefore, by the above procedure, $\cup_n [\![S^a_n]\!]$ can be represented as the union of a sequence of disjoint graphs of predictable stopping times. Hence we can suppose that $\{(t,\omega) : \Delta M_t(\omega) \neq 0\}$ is contained in a set of the form $\cup_n [\![T_n]\!]$, where each T_n is either totally inaccessible or predictable, and their graphs are disjoint.

For each $n \in \mathbb{N}$ consider

$$A^n := \Delta M_{T_n} I_{\{t \geq T_n\}}, \qquad M^n = A^n - (\Pi^*_p A^n).$$

From the above results, M^n is a martingale in $\mathcal{H}^{2,d}$ which is continuous except at the stopping time T_n , and its jump at T_n is ΔM_{T_n} . Write

$$B^k = M^1 + \cdots + M^k.$$

Then $M - B^k$ is continuous at $T_1, \dots, T_k$ and so orthogonal to $M^1, \dots, M^k$ and to their sum B^k. Therefore,

$$\begin{aligned} E[(M_\infty)^2] &= E[(B^k_\infty)^2] + E[(M - B^k)^2_\infty] \\ &= \sum_{n=1}^{k} E[(M^n_\infty)^2] + E[(M - B^k)^2_\infty] \\ &= \sum_{n=1}^{k} E[\Delta M^2_{T_n}] + E[(M - B^k)^2_\infty]. \end{aligned}$$

As $M \in \mathcal{H}^{2,d}$, we know $\lim_k \big(\sum_{n=1}^k E\big[(\Delta M_{T_n})^2\big]\big)$ is finite, so the sequence of partial sums $B^k = M^1 + \cdots + M^k$ converges in $\mathcal{H}^2$ to a martingale B. Because $M - B^k$ is orthogonal to B^k for each k, for any stopping time S,

$$E\big[(M-B)_S B_S\big] = E\big[\lim_k (M - B^k)_S B^k_S\big] = \lim_k E\big[(M - B^k)_S B^k_S\big] = 0,$$

so $M - B$ and B are orthogonal.

From Theorem 10.1.10, there is a subsequence of $\{B^k\}_{k \in \mathbb{N}}$ whose sample paths converge uniformly, almost surely, to B. Therefore, $M - B$ is continuous. However $M \in \mathcal{H}^{2,d}_0$, so $M - B$ is orthogonal to M. Consequently, $M - B$ is orthogonal to M and B, hence to $M - B$, and so $M - B = 0$. That is, $M = B = \lim_k \big(\sum_{n=1}^k M^n \big)$. ☐

Corollary 10.2.15. *If $M \in \mathcal{H}^{2,d}$ is not zero at $t = 0$, the above result is still valid. This is because, following the convention that $M_{0-} = 0$ a.s., $\Delta M_0 = M_0$ and so we can write $T_0 = 0$ and*

$$\Delta M_{T_0} I_{\{t \geq T_0\}} = M^0_t\,.$$

Therefore, $(M - M_0) \in \mathcal{H}^{2,d}_0$. Using Theorem 10.2.14, we see that

$$M = \lim_k \Big(\sum_{n=0}^{k} M^n \Big).$$

Corollary 10.2.16. *The decomposition of the above theorem can be applied to any $M \in \mathcal{H}^2$, so that $M = B^k + (M - B^k)$. As in the theorem, $B = \lim_k B^k$ exists in $\mathcal{H}^2$, and $M - B$ is continuous.*

The martingale, $M - B$ is the projection of M onto $\mathcal{H}^{2,c}_0$, and B is the projection of M onto $\mathcal{H}^{2,d}$. That is,

$$M = M^c + M^d,$$

where $M^c = M - B \in \mathcal{H}^{2,c}_0$ and $M^d = B \in \mathcal{H}^{2,d}$. This decomposition is unique (up to indistinguishability).

Corollary 10.2.17. *For any $M \in \mathcal{H}^2$,*

$$E\Big[\sum_{s\in[0,\infty[} \Delta M_s^2\Big] \le E[M_\infty^2].$$

Equality holds here if, and only if, $M \in \mathcal{H}^{2,d}$. (Here again we use the convention that $\Delta M_0 = M_0$.)

Proof. Recall that M has at most countably many jumps, so the left-hand side of the inequality is well defined. Following the proof of the above theorem we have

$$E[M_\infty^2] = \sum_{n=1}^{k} E[\Delta M_{T_n}^2] + E[(M - B^k)_\infty^2].$$

For almost every ω, $\sum_{n\in\mathbb{N}} \Delta M_{T_n}^2(\omega) = \sum_{s\in[0,\infty[} \Delta M_s^2(\omega)$. So, in the limit,

$$E[M_\infty^2] = E\Big[\sum_{s\in[0,\infty[} \Delta M_s^2\Big] + E[(M-B)_\infty^2].$$

The result follows. □

Corollary 10.2.18. *For any $t \le \infty$ and $M \in \mathcal{H}^2$, we have $\sum_{s\le t} \Delta M_s^2 < \infty$ a.s.*

Corollary 10.2.19. *If M and N are in $\mathcal{H}^2$, we have the bounds*

(i)

$$\sum_s |\Delta M_s \Delta N_s| \le \Big(\sum_s \Delta M_s^2\Big)^{1/2} \Big(\sum_s \Delta N_s^2\Big)^{1/2},$$

and

(ii) $E\big[\sum_s |\Delta M_s \Delta N_s|\big] \le \|M_\infty\|_2 \|N_\infty\|_2$.

Proof. From Corollary 10.2.17 the right-hand side of (i) is in $L^1(\Omega)$ and is, therefore, almost surely finite. The result follows from the Cauchy–Schwarz inequality.

Taking expectations, part (ii) is also immediate from (i) and Corollary 10.2.17, using the Cauchy–Schwarz inequality (with respect to the expectation). □

Theorem 10.2.20. *Suppose M and N belong to $\mathcal{H}^2$ and one of them, say M, belongs to $\mathcal{H}^{2,d}$. Then, with the convention $M_0 = \Delta M_0$ and $N_0 = \Delta N_0$,*

(i) $E[M_\infty N_\infty] = E\big[\sum_{s\ge 0} \Delta M_s \Delta N_s\big]$, *and*

(ii) the process L defined by

$$L_t := M_t N_t - \sum_{0\le s\le t} \Delta M_s \Delta N_s$$

belongs to $\mathcal{H}_0^1$.

Proof. Suppose first that both M and N belong to $\mathcal{H}^{2,d}$. Then, from Corollary 10.2.17

$$E[M_\infty^2] = E\Big[M_0^2 + \sum_{s>0} \Delta M_s^2\Big], \qquad E[N_\infty^2] = E\Big[N_0^2 + \sum_{s>0} \Delta N_s^2\Big],$$

and

$$E[(M+N)_\infty^2] = E\Big[(M_0+N_0)^2 + \sum_{s>0} (\Delta M_s + \Delta N_s)^2\Big].$$

Part (i) follows by subtraction. Applying part (i) to the martingales M^T, N^T, stopped at $T \in \mathcal{T}$, we have $E[L_T] = 0$. Therefore, as in Theorem 5.4.6, L is a martingale. Furthermore, for any t,

$$|L_t| \leq M_t^* N_t^* + \sum_{s \leq t} |\Delta M_s \Delta N_s| \leq M_\infty^* N_\infty^* + \sum_{s \in [0,\infty[} |\Delta M_s \Delta N_s|,$$

which is integrable by Corollary 10.2.19, so $L \in \mathcal{H}_0^1$.

In general, suppose $N \in \mathcal{H}^2$ and $M \in \mathcal{H}^{2,d}$. Then, by Corollary 10.2.16, $N = N^c + N^d$, where $N^c \in \mathcal{H}_0^{2,c}$ and $N^d \in \mathcal{H}^{2,d}$. By orthogonality, $M_t N_t^c$ is a martingale which is zero at $t = 0$. Therefore, $E[M_\infty N_\infty^c] = 0$ and MN^c is in $\mathcal{H}_0^1$. As $MN = MN^c + MN^d$, the result follows. □

Theorem 10.2.21. *If $M \in \mathcal{H}^2 \cap \mathcal{V}$ (that is, it is a finite variation martingale with $E[\sup_t |M_t|^2] < \infty$), then $M \in \mathcal{H}^{2,d}$.*

Proof. If $M \in \mathcal{H}^2 \cap \mathcal{A}$ (that is, it has integrable variation), then, by Lemma 10.2.7,

$$E[M_\infty N_\infty] = E\Big[M_0 N_0 + \sum_{s>0} \Delta M_s \Delta N_s\Big],$$

for any bounded martingale N.

Both sides of the above identity are continuous in N under the $\mathcal{H}^2$ norm, and so the above identity is valid for $N \in \mathcal{H}^2$. Therefore, in particular,

$$E[M_\infty^2] = E\Big[M_0^2 + \sum_{s>0} \Delta M_s^2\Big],$$

for $M \in \mathcal{H}^2 \cap \mathcal{A}$ and, by continuity, for $M \in \mathcal{H}^2 \cap \mathcal{V}$. Therefore, by Corollary 10.2.17, $M \in \mathcal{H}^{2,d}$. □

The following theorem enforces the intuition that, for processes in $\mathcal{H}^{2,d}$, one only needs to pay attention to the jumps of the process.

Theorem 10.2.22. *Let $M, N \in \mathcal{H}^{2,d}$ be such that $\Delta M = \{M_t - M_{t-}\}_{t\geq 0}$ and $\Delta N = \{N_t - N_{t-}\}_{t \geq 0}$ are indistinguishable. Then M and N are indistinguishable.*

Proof. We have $L = M - N \in \mathcal{H}^{2,d}$, and ΔL is indistinguishable from the zero process. By Lemma 10.2.7, $E[L_\infty^2] = E\big[L_0^2 + \sum_s \Delta L_s^2\big] = 0$ and so, by Lemma 10.1.3, L is indistinguishable from the zero process. □

10.3 Localization

We now consider processes which are locally in $\mathcal{H}^2$. Recall that $\mathcal{H}^2_{\text{loc}} \subset \mathcal{M}_{\text{loc}}$ is the space of locally square integrable local martingales. We first show that all continuous local martingales are in $\mathcal{H}^2_{\text{loc}}$.

Lemma 10.3.1. *If M is a continuous local martingale, then $M \in \mathcal{H}^2_{\text{loc}}$.*

Proof. Write

$$T_n = \inf\{t : |M_t| \geq n\}.$$

As M is continuous, $M_s(\omega)$ is almost surely bounded on $[0,t]$ for all t, therefore $\lim_n T_n = \infty$ a.s. and $M^{T_n} \in \mathcal{H}^2$ because $|M_{t\wedge T_n}| \leq n$. □

Remark 10.3.2. If M is càdlàg but not continuous, M is bounded by n on the interval $[\![0, T_n[\![$, but generally one knows nothing about the jump at T_n. (Here T_n is as in the above lemma.)

On the other hand, if we know that

$$E[(\Delta M_T)^2] < \infty$$

for every bounded stopping time T, then we see that $M^d \in \mathcal{H}^2_{\text{loc}}$, and hence (by Lemma 5.6.11) that $M \in \mathcal{H}^2_{\text{loc}}$.

The following result, due to Doléans-Dade and Yen (see [134] and [61]) is sometimes called the 'fundamental theorem of local martingales', and will form the basis for much of our analysis of local martingales. It allows us to decompose a local martingale into the sum of a locally square-integrable local martingale and a local martingale of locally integrable variation.

Theorem 10.3.3. *Suppose $M \in \mathcal{M}_{\text{loc}}$, and let $a > 0$. Then M can be written as*

$$M = M_0 + U + V$$

where

- *U and V are local martingales and $U_0 = V_0 = 0$,*
- *$|\Delta U| \leq 2a$, so $U \in \mathcal{H}^p_{\text{loc}}$ for all p,*
- *V is locally of integrable variation and has finitely many jumps on any finite interval.*

Proof. It is sufficient to prove the result when $M \in \mathcal{M}_{0,\text{loc}}$. As M is càdlàg, $|\Delta M_s| > a$ for a.s. only finitely many s on any finite interval. Hence we can define

$$A_t = \sum_{s\leq t} \Delta M_s I_{\{|\Delta M_s|>a\}}$$

$$\text{and} \qquad A_t^{(+)} = \int_{[0,t]} |dA_s| = \sum_{s\leq t} |\Delta M_s| I_{\{|\Delta M_s|>a\}},$$

so $A \in \mathcal{V}$ and $A^{(+)} \in \mathcal{V}^+$. Put $T_n = \inf\left\{t : A_t^{(+)} \vee |M_t| > n\right\} \wedge S_n$, for S_n a sequence localizing M, so $M^{T_n} \in \mathcal{M}$ and, in particular, M_{T_n} is integrable. We know

$$A_{T_n}^{(+)} \leq n + |\Delta M_{T_n}| = n + |M_{T_n} - M_{T_n-}| \leq 2n + |M_{T_n}|,$$

and, as M_{T_n} is integrable, we see that $A_{T_n}^{(+)}$ is integrable, and hence $A \in \mathcal{A}_{\mathrm{loc}}$. Therefore, we can define $V = A - \Pi_p^* A$, which is locally a martingale of integrable variation, and $V_0 = 0$.

Clearly, $U := M - V$ is a local martingale, and we next show that its jumps are bounded. On the set $\{\Delta(\Pi_p^* A) = 0\}$, we have

$$|\Delta U| = |\Delta(M - V)| = |\Delta M| I_{\{|\Delta M| \leq a\}} \leq a. \tag{10.1}$$

As $\Pi_p^* A$ is a predictable càdlàg process, it is locally bounded and its jumps occur only on a predictable thin set (as in Exercise 7.7.1). That is,

$$\{\Delta(\Pi_p^* A) \neq 0\} \subseteq \bigcup_n [\![S_n]\!]$$

for $\{S_n\}_{n \in \mathbb{N}}$ a sequence of predictable stopping times. As $\Pi_p^* A$ is locally bounded, Theorem 7.6.5 implies that for any n, $\Delta(\Pi_p^* A)_{S_n}$ is $\mathcal{F}_{S_n-}$-measurable. Furthermore,

$$\Delta U = \Delta M - \Delta V = \Delta M I_{\{|\Delta M| \leq a\}} + \Delta(\Pi_p^* A),$$

so ΔU is also locally bounded. Hence, as U is a local martingale, Theorem 5.6.13 implies that $U_{S_n-} = E[U_{S_n} | \mathcal{F}_{S_n-}]$ and, therefore,

$$\begin{aligned}
\Delta U_{S_n} &= \Delta U_{S_n} - E[\Delta U_{S_n} | \mathcal{F}_{S_n-}] \\
&= \Delta(M - A)_{S_n} + \Delta(\Pi_p^* A)_{S_n} - E\big[\Delta(M - A)_{S_n} + \Delta(\Pi_p^* A)_{S_n} \big| \mathcal{F}_{S_n-}\big] \\
&= \Delta(M - A)_{S_n} - E\big[\Delta(M - A)_{S_n} \big| \mathcal{F}_{S_n-}\big].
\end{aligned}$$

So, as $\Delta(M - A) = \Delta M I_{\{|\Delta M| \leq a\}}$,

$$|\Delta U_{S_n}| \leq |(\Delta M I_{\{|\Delta M| \leq a\}})_{S_n}| + \Big| E\big[(\Delta M I_{\{|\Delta M| \leq a\}})_{S_n} \big| \mathcal{F}_{S_n-}\big] \Big| \leq 2a. \tag{10.2}$$

Combining (10.1) and (10.2), we see that $|\Delta U| \leq 2a$. As in Remark 10.3.2, it follows that $U \in \mathcal{H}^2_{\mathrm{loc}}$. □

Recall from Definition 10.1.14 that $M, N \in \mathcal{M}_{\mathrm{loc}}$ are *orthogonal* if $MN \in \mathcal{M}_{0,\mathrm{loc}}$, that is, their product is a local martingale starting at zero. The following result is analogous to Corollary 10.2.16.

Theorem 10.3.4. *Suppose $M \in \mathcal{M}_{\mathrm{loc}}$. Then $\mathcal{M}$ can be written in a unique way as $M = M^c + M^d$, where M^c and M^d are in $\mathcal{M}_{\mathrm{loc}}$, M^c is continuous (so $M_0^c = 0$), and M^d is orthogonal to every continuous local martingale.*

Proof. (a) *Uniqueness.* Suppose

$$M = M^{c,1} + M^{d,1} = M^{c,2} + M^{d,2},$$

where all the terms are local martingales, $M^{c,1}$ and $M^{c,2}$ are continuous, and $M^{d,1}$ and $M^{d,2}$ are orthogonal to every continuous local martingale. Then $M^{d,1} - M^{d,2} = M^{c,2} - M^{c,1}$ is a continuous local martingale which is orthogonal to itself. Therefore, $(M^{d,1} - M^{d,2})^2$ is a local martingale which is nonnegative and zero at $t = 0$. Consequently, $M^{d,1} = M^{d,2}$.

(b) *Existence.* Following the notation of Theorem 10.3.3, let

$$M = U + V$$

where $U \in \mathcal{H}^2_{\text{loc}}$ and V is a local martingale of locally integrable variation. By localizing U, from Corollary 10.2.16, we can define $U = U^c + U^d$, where $U^c \in \mathcal{H}^{2,c}_{0,\text{loc}}$ and $U^d \in \mathcal{H}^{2,d}_{\text{loc}}$ are the unique continuous and totally discontinuous local martingales in the decomposition of $U \in \mathcal{H}^2_{\text{loc}}$. Let $M^c = U^c$ and $M^d = U^d + V$.

We must show that M^d is orthogonal to every continuous local martingale N with $N_0 = 0$. Let $\{T_n\}_{n\in\mathbb{N}}$ be a sequence of stopping times such that $T_n \to \infty$, $(U^d)^{T_n} \in \mathcal{H}^{2,d}$, $V^{T_n} \in \mathcal{A}$ and N^{T_n} is bounded (this last property can be guaranteed by considering the stopping times $\inf\{t : |N_t| \geq n\}$.). For every n, we know $N^{T_n} \in \mathcal{H}^{2,c}_0$ so N^{T_n} is orthogonal to $(U^d)^{T_n} \in \mathcal{H}^{2,d}$, by definition, and to V^{T_n} by Corollary 10.2.9.

Consequently, $(U^d + V)N$ is a local martingale (with localizing sequence T_n) and $M^c = U^c$ and $M^d = U^d + V$ give the desired decomposition. □

The following lemma is particularly useful when applied to the process $X = \Delta M$, for M a local martingale.

Lemma 10.3.5. *Let X be an optional process which is zero except on a thin set. Define, for $a > 0$ and $t \geq 0$,*

$$A_t = \Big(\sum_{s\leq t} X_s^2\Big)^{1/2},$$

$$B_t^{(a)} = \Big(\sum_{s\leq t}(X_s^2 I_{\{|X_s|\leq a\}} + |X_s| I_{\{|X_s|\leq a\}})\Big).$$

Then $A \in \mathcal{A}_{\text{loc}}$ if and only if $B^{(a)} \in \mathcal{A}_{\text{loc}}$ for some $a > 0$, in which case $B^{(a)} \in \mathcal{A}_{\text{loc}}$ for all $a > 0$.

Proof. It is easy to see that the statements $A \in \mathcal{V}$, $A^2 \in \mathcal{V}$ and $B^{(a)} \in \mathcal{V}$ for any $a > 0$ are all equivalent, and that these are implied as soon as one of the processes is locally integrable.

To show that $A \in \mathcal{A}_{\mathrm{loc}}$ implies $B^{(a)} \in \mathcal{A}_{\mathrm{loc}}$ for all a, we first note that, as $B^{(a)} \in \mathcal{V}$, the stopping times $T_n := \inf\{t : B_t^{(a)} \geq n\} \to \infty$ almost surely. As $|b| \leq (b^2 + c^2)^{1/2}$, we know that $|X| \leq A$, and hence

$$\begin{aligned} B_{T_n}^{(a)} &\leq n + \big(X_{T_n}^2 I_{\{|X_{T_n}| \leq a\}} + |X_{T_n}| I_{\{|X_{T_n}| \leq a\}}\big) I_{\{T_n < \infty\}} \\ &\leq n + a^2 \vee A_{T_n}, \end{aligned}$$

so $B^{(a)} \in \mathcal{A}_{\mathrm{loc}}$ whenever $A \in \mathcal{A}_{\mathrm{loc}}$, for all values of $a > 0$.

Now suppose $B^{(a)} \in \mathcal{A}_{\mathrm{loc}}$ for some fixed $a > 0$. We know that $\mathcal{A} \in \mathcal{V}$ and so the stopping times $S_n := \inf\{t : A_t \geq n\} \to \infty$ almost surely. As $(b^2 + c^2)^{1/2} \leq |b| + |c|$, we have

$$\begin{aligned} A_{S_n} &\leq n + a \vee \big(X_{S_n}^2 I_{\{|X_{S_n}| \leq a\}} + |X_{S_n}| I_{\{|X_{S_n}| \leq a\}}\big) I_{\{S_n < \infty\}} \\ &\leq n + a \vee B_{S_n}^{(a)}, \end{aligned}$$

and we see that $A \in \mathcal{A}_{\mathrm{loc}}$ whenever $B^{(a)} \in \mathcal{A}_{\mathrm{loc}}$. □

A converse to the following lemma will be given in Theorem 11.5.11.

Lemma 10.3.6. *Suppose that $M \in \mathcal{M}_{0,\mathrm{loc}}$. For every finite t, the sum $\sum_{s \leq t} (\Delta M_s(\omega))^2$ is finite for almost every ω, and the process Y defined by*

$$Y_t = \Big(\sum_{s \leq t} (\Delta M_s)^2\Big)^{1/2}$$

is locally integrable (that is, $Y \in \mathcal{A}_{\mathrm{loc}}^+$).

Proof. We have $M = U^c + U^d + V$ in the notation of Theorem 10.3.4, taking the decomposition with $|\Delta U| \leq 1$. Therefore,

$$Y_t^2 = \sum_{s \leq t} (\Delta M_s)^2 \leq 2\Big(\sum_{s \leq t} (\Delta U_s^d)^2 + \sum_{s \leq t} (\Delta V_s)^2\Big).$$

Using the general inequality $(\sum |x_i|)^{1/2} \leq \sum |x_i|^{1/2}$, we also see that

$$Y_t \leq 2^{1/2}\Big(\Big(\sum_{s \leq t} (\Delta U_s^d)^2\Big)^{1/2} + \sum_{s \leq t} |\Delta V_s|\Big).$$

As $|\Delta U| \leq 1$ and V is of locally integrable variation, the process on the right is locally integrable, so $Y \in \mathcal{A}_{\mathrm{loc}}^+$ and is almost surely finite. □

Corollary 10.3.7. *If M is a local martingale, by stopping at the times $T_n = \inf\{t : |M_t| \geq n\} \wedge n$, we have*

$$M_t^{T_n} \leq n + |M_{T_n \wedge t}| \leq n + \Big(\sum_{s \leq t} (\Delta M_s)^2\Big)^{1/2} \in \mathcal{A}_{\mathrm{loc}}^+.$$

If S_n is a localizing sequence for the right-hand side, then $M^{T_n \wedge S_n} \in \mathcal{H}^1$, and hence $M \in \mathcal{H}_{\mathrm{loc}}^1$. Therefore, $\mathcal{H}_{\mathrm{loc}}^1 = \mathcal{M}_{\mathrm{loc}}$.

When we also know that the processes are of finite variation, we obtain further useful results.

Lemma 10.3.8. *Suppose $A \in \mathcal{V}$ (that is, it is adapted and almost surely of finite variation) and that A is a local martingale. Then we also have $A \in \mathcal{A}_{\mathrm{loc}}$, that is, A is locally of integrable variation.*

Proof. We can suppose $A_0 = 0$. In the notation of Theorem 10.3.3 A can be expressed as $U + V$, where $U \in \mathcal{H}^2_{\mathrm{loc}}$ and $V \in \mathcal{A}_{\mathrm{loc}}$. Let $\{T_n\}_{n\in\mathbb{N}}$ be an increasing sequence of stopping times such that $\lim_n T_n = \infty$ and each T_n reduces V. Write

$$S_n = \inf\Big\{t : \int_{[0,t]} |dA_s| \geq n\Big\}$$

and $R_n = T_n \wedge S_n$. We must show that

$$E\Big[\int_{]0,R_n]} |dA_s|\Big] < \infty.$$

However, from the definition of S_n, we know that

$$E\Big[\int_{[0,R_n[} |dA_s|\Big] \leq n.$$

As R_n reduces V, $\Delta V_{R_n} \in L^1(\Omega)$, and by construction ΔU is bounded. Therefore,

$$E[|\Delta A_{R_n}|] < \infty$$

and

$$E\Big[\int_{[0,R_n]} |dA_s|\Big] < \infty.$$

□

The following result is related to Corollary 8.2.14.

Lemma 10.3.9. *(i) Every predictable local martingale M is continuous.*
(ii) Every predictable local martingale M which is of finite variation is constant. In particular, if we also have $M_0 = 0$ a.s. then $M_s = 0$ a.s.

Proof. (i) By considering a stopping time which reduces M, we can suppose M is uniformly integrable. Because M is predictable, its discontinuities occur on a predictable thin set. Therefore, suppose T is a predictable stopping time. Then

$$\Delta M_T = M_T - M_{T-} = M_T - E[M_T|\mathcal{F}_{T-}]$$

by Theorem 6.2.18. However, M is predictable, so by Exercise 7.7.5, M_T is $\mathcal{F}_{T-}$ measurable and $\Delta M_T = 0$. It follows that M has no times of discontinuity.

(ii) If, furthermore, M is of finite variation, then the process $\{\int_{[0,t]} |dM_s|\}_{t\geq 0}$ is continuous, and so locally of integrable variation. By Corollary 8.2.14, $M_s = M_0$ for all s. □

10.4 Exercises

Exercise 10.4.1. Let W be a Brownian motion and ξ a nontrivial $\mathcal{F}_0$-measurable bounded random variable, in a filtered probability space satisfying the usual conditions. Suppose $E[\xi] = 0$, and let $X = \xi W$. Show that X_t and W_t are orthogonal random variables in L^2, for any t, but that X and W are *not* orthogonal martingales (in the sense of very strong orthogonality). Find a stopping time T such that $E[X_T W_T] \neq 0$.

Exercise 10.4.2. Let X be a local martingale in $\mathcal{H}^2_{\mathrm{loc}}$ and Y a nonnegative uniformly integrable martingale in $\mathcal{H}^2$ with $Y_0 = 1$. Define a probability measure $\mathbb{Q}$ by the Radon–Nikodym derivative $d\mathbb{Q}/d\mathbb{P} = Y_\infty$. Show that X is a local $\mathbb{Q}$-martingale if and only if X and Y are orthogonal.

Exercise 10.4.3. Let M be a local martingale of locally bounded variation. Show that for any bounded predictable process Z, the process $Z \bullet M$ is in $\mathcal{H}^{2,d}_{\mathrm{loc}}$.

Exercise 10.4.4. Let M be a local martingale and $M = U + V$ be the decomposition established in Theorem 10.3.3. Suppose the jumps of M are totally inaccessible. Show that U and V are orthogonal.

Exercise 10.4.5. Let A be a process of the form $A_t = \sum_{i=1}^\infty B_i I_{\{t \geq T_i\}}$, where $\{T_n\}_{n\in\mathbb{N}}$ is an increasing sequence of stopping times with $T_n \to \infty$ and $\{B_n\}_{n=1}^\infty$ is a sequence of random variables with $B_n \in \mathcal{F}_{T_n}$, for $\{\mathcal{F}_t\}_{t\geq 0}$ a filtration satisfying the usual conditions.

(i) Show that $A - \Pi^*_p A \in \mathcal{H}^q_{\mathrm{loc}}$ if $B_n \in L^q$ for every n.
(ii) Give an example of B_n such that $B_n \notin L^q$ for any q, but $A - \Pi^*_p A \in \mathcal{H}^q$ for every q.

Hint: Using the results of Section 8.2.3, we have the inequality $E[(\Pi^*_p A)^q_t]^{1/q} \leq C_q E[|A_t|^q]^{1/q}$ for any $q \geq 1$, for some constant C_q.

Exercise 10.4.6. Let $p \in [1, \infty[$ and let M be a local martingale such that $E[(M^*_\infty)^p] < \infty$. Show that M is a true martingale and hence $M \in \mathcal{H}^p$.

11

Quadratic Variation and Semimartingales

We now come to one of the key objects in stochastic analysis, and what fundamentally distinguishes the theory from classical calculus. This is the notion of the *quadratic variation* of a process.

As in the previous chapter, we assume that we have a filtration $\{\mathcal{F}_t\}_{t\in[0,\infty]}$ satisfying the usual conditions, and for simplicity, $\mathcal{F}_\infty = \mathcal{F}_{\infty-} = \bigvee_{t<\infty} \mathcal{F}_t$. When we say martingale, we invariably mean its càdlàg version, and all statements should be read as holding up to indistinguishability, unless otherwise indicated.

11.1 Quadratic Variation

Suppose $M \in \mathcal{H}_0^2$. Then $M_\infty^* = \sup_t |M_t| \in L^2(\Omega)$ and $\{M_t^2\}_{t\geq 0}$ is a submartingale such that, for all $t \geq 0$, we have $M_t^2 \leq (M_\infty^*)^2 \in L^1(\Omega)$. The right continuous version of the process $X = \{E[M_\infty^2|\mathcal{F}_t] - M_t^2\}_{t\geq 0}$ is, therefore, a càdlàg
supermartingale of class (D). Clearly $X_t \geq 0$ a.s. and $X_\infty = 0$, so X is a potential of class (D).

Definition 11.1.1. *For $M \in \mathcal{H}^2$, we denote by $\langle M\rangle$ the unique* predictable *increasing process in $\mathcal{A}_0^+$ given by the Doob–Meyer decomposition of the class (D) potential X defined by*

$$X_t = E[M_\infty^2|\mathcal{F}_t] - M_t^2.$$

(See Theorem 9.1.6.) The process $\langle M\rangle$ is called the predictable quadratic variation of M.

S.N. Cohen, R.J. Elliott, *Stochastic Calculus and Applications*,
Probability and Its Applications, DOI 10.1007/978-1-4939-2867-5_11

Remark 11.1.2. From this definition, we see that $\langle M\rangle$ is the unique increasing predictable process (indeed, the unique finite variation predictable process) such that

$$M^2 - \langle M\rangle$$

is a martingale and $\langle M\rangle_0 = 0$. Clearly, if $M_0 = 0$, then $M^2 - \langle M\rangle \in \mathcal{H}_0^1$.

Remark 11.1.3. At an intuitive level, the change in the predictable quadratic variation describes the 'variance' of our martingale, locally in time. For continuous martingales, the quantity $d\langle M\rangle/dt$, when defined, is often called the 'volatility' of the process (Fig. 11.1).

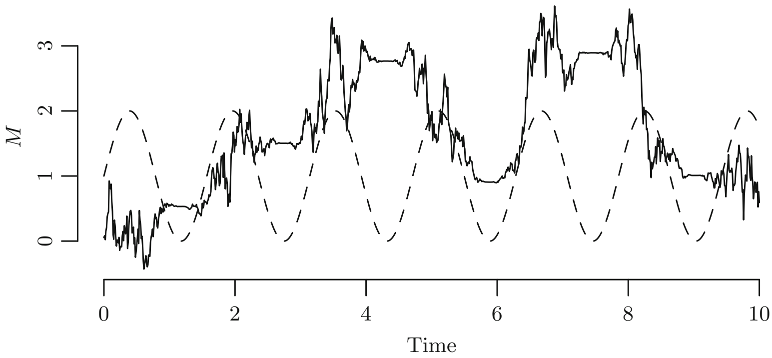

Fig. 11.1. A path of a continuous martingale with time-varying quadratic variation. The derivative of the quadratic variation $d\langle M\rangle/dt$ is the sinusoid given by the dashed line.

To deal with jumps effectively, we also require the following 'optional' quadratic variation.

Definition 11.1.4. *Suppose $M \in \mathcal{H}^2$ and $M = M^c + M^d$ is its decomposition as in Corollary 10.2.16 into a continuous martingale and a sum of compensated jump martingales. We define $[M]$ to be the optional increasing process*

$$[M]_t = \langle M^c\rangle_t + \sum_{0<s\leq t} \Delta M_s^2\,.$$

The process $[M]$ is called the optional quadratic variation *of M. From Corollary 10.2.17, $[M] \in \mathcal{A}_0^+$ and, from Corollary 10.2.18, $[M]_t < \infty$* a.s. *for all $t < \infty$.*

Remark 11.1.5. For $M = M^c + M^d \in \mathcal{H}^2$,

$$M_t^2 = (M_t^c)^2 + 2M_t^c M_t^d + (M_t^d)^2.$$

Because M^c and M^d are orthogonal and $M_0^c = 0$, we know that $M^c M^d$ is a martingale and is in $\mathcal{H}_0^1$, by the Cauchy–Schwarz inequality. From Theorem 10.2.20, $(M_t^d)^2 - \sum_{s \leq t} \Delta M_s^2$ is a martingale in $\mathcal{H}_0^1$ and, by Definition 11.1.1, $(M_t^c)^2 - \langle M^c \rangle_t$ is in $\mathcal{H}^1$. Therefore, $M_t^2 - [M]_t \in \mathcal{H}^1$.

The following are classic examples of these quantities. (Formally, in each case $X \notin \mathcal{H}^2$ unless we stop at some time. However, we shall see that this does not cause difficulties, as the quadratic variations are well defined in $\mathcal{H}^2_{\text{loc}}$.)

Example 11.1.6. For X a Brownian motion, from Exercise 5.7.6 we see that $X_t^2 - t$ is a martingale and hence (by uniqueness), $\langle X \rangle_t = t$. As a Brownian motion is continuous, $[X]_t = \langle X \rangle_t = t$.

Example 11.1.7. For N a Poisson process, we know that $X_t := N_t - \lambda t$ is a martingale, and (from Exercise 5.7.9), $E[(X_t - X_s)^2 | \mathcal{F}_s] = \lambda(t - s)$ for all $s \leq t$. By simple calculations, one can show that $\langle X \rangle_t = \lambda t$. On the other hand, X is a pure jump martingale, and its jumps are of size 1, so

$$[X]_t = \sum_{0<s\leq t} (\Delta N_s)^2 = \sum_{0<s\leq t} \Delta N_s = N_t \neq \lambda t = \langle X \rangle_t.$$

Remark 11.1.8. We shall see, in Corollary 14.1.2, that the optional quadratic variation also arises in the following way. For a martingale $M \in \mathcal{H}^2$ and each $N > 0$, let $\pi_N = \{0 = t_0^N < t_1^N < t_2^N < \ldots < t_N^N = t\}$ be a partition of the interval $[0, \infty[$ by *stopping times* $\{t_i^N\}_{i,N\in\mathbb{N}}$. Define $|\pi_N| = \max_i |t_i^N - t_{i-1}^N|$, and suppose $|\pi_N| \to 0$ as $N \to \infty$. Then

$$\sum_{i=1}^{N} (M_{t_{i+1}^N \wedge t} - M_{t_i^N \wedge t})^2 \to [M]_t \quad \text{as } N \to \infty,$$

the convergence taking place in probability for each t. It is this property which justifies the name 'quadratic variation'.

Note that, in general, this convergence does not hold pathwise, that is, if the $\{t_i^N\}_{i,N\in\mathbb{N}}$ are not required to be stopping times, then convergence is not guaranteed.

Lemma 11.1.9. *For $M \in \mathcal{H}^2$, $\langle M \rangle$ is the dual predictable projection of $[M]$, that is,*

$$\langle M \rangle = \Pi_p^*[M].$$

Proof. As in Remark 11.1.5, $[M] - \langle M \rangle$ is a martingale, so the result follows from Theorem 8.2.6 and Corollary 8.2.16. □

Lemma 11.1.10. *$\langle M^c \rangle$ is continuous.*

Proof. This follows from Remark 9.2.6. □

11.2 Quadratic Covariation

From the definition of the quadratic variation comes the quadratic covariation.

Definition 11.2.1. *Suppose $M, N \in \mathcal{H}^2$. Then we define the* predictable quadratic covariation

$$\langle M, N\rangle = \frac{1}{2}\Big(\langle M+N\rangle - \langle M\rangle - \langle N\rangle\Big).$$

One can check that $\langle M, N\rangle$ is the unique predictable process of integrable variation such that

$$MN - \langle M, N\rangle \in \mathcal{H}^1 \quad \textit{and} \quad \langle M, N\rangle_0 = 0\,.$$

Definition 11.2.2. *Suppose $M, N \in \mathcal{H}^2$. We define the* optional quadratic covariation

$$[M, N] = \frac{1}{2}\Big([M+N] - [M] - [N]\Big).$$

Then

$$[M, N] \in \mathcal{A}, \quad MN - [M, N] \in \mathcal{H}^1 \quad \textit{and} \quad [M, N]_0 = 0\,.$$

Remark 11.2.3. It is immediate from these definitions that $\langle M\rangle = \langle M, M\rangle$ and $[M] = [M, M]$.

Remark 11.2.4. Intuitively, $\langle M, N\rangle$ and $[M, N]$ both behave like inner products on a Hilbert space, acting on martingales in $\mathcal{H}^2$, locally in both time and space (see Exercise 11.7.1). In fact, it is clear from the definition that

$$(M, N)_{\mathcal{H}^2} := E[M_\infty N_\infty] = E[\langle M, N\rangle_\infty + M_0 N_0]$$

where $(M, N)_{\mathcal{H}^2}$ denotes the inner product in $\mathcal{H}^2$ induced by associating it with $L^2(\Omega, \mathcal{F}, P)$, as in Remark 10.1.12.

Furthermore, if we fix a martingale $M \in \mathcal{H}^2$, as $\langle M\rangle$ is an increasing process, we also know that

$$H, K \mapsto E\Big[\int_{[0,\infty]} H_t K_t \, d\langle M\rangle\Big]$$

is an inner product, for H, K in the space of predictable processes (up to an equivalence relation). This property is what we shall use to define the stochastic integral.

Remark 11.2.5. From the definitions,

$$[M, N]_t = \langle M^c, N^c\rangle_t + \sum_{0<s\le t} \Delta M_s \Delta N_s\,.$$

Furthermore, M and N are orthogonal if and only if $\langle M, N\rangle$ is the zero process and $M_0 N_0 = 0$, in which case $\langle M+N\rangle = \langle M\rangle + \langle N\rangle$. If M and N are orthogonal and do not jump simultaneously, then $[M+N] = [M] + [N]$ also.

Remark 11.2.6. From the definition, $M^2 - \langle M \rangle$ is a martingale in $\mathcal{H}^1$, so by Doob's optional stopping theorem (Theorem 5.3.1), for any $T \in \mathcal{T}$,

$$E[\,\langle M\rangle_\infty - \langle M\rangle_T|\mathcal{F}_T] = E[M_\infty^2|\mathcal{F}_T] - M_T^2\,.$$

Lemma 11.2.7. *Suppose $M, N \in \mathcal{H}^2$ and $T \in \mathcal{T}$. Then $\langle M, N\rangle^T = \langle M, N^T\rangle$.*

Proof. By the optional stopping theorem (Theorem 5.3.1),

$$(MN)^T - \langle M, N\rangle^T \in \mathcal{H}^1$$

and

$$M_t N_t^T - M_t^T N_t^T = E[\,(M_\infty - M_T)N_T|\mathcal{F}_t]\,.$$

Therefore,

$$MN^T - \langle M, N\rangle^T \in \mathcal{H}^1\,.$$

By definition

$$MN^T - \langle M, N^T\rangle \in \mathcal{H}^1 \quad \text{so} \quad \langle M, N^T\rangle - \langle M, N\rangle^T \in \mathcal{H}_0^1\,.$$

By Corollary 8.2.14, as a predictable finite variation martingale is indistinguishable from a constant, we conclude

$$\langle M, N^T\rangle = \langle M, N\rangle^T.$$

□

Corollary 11.2.8. *For $M, N \in \mathcal{H}^2$ and $T \in \mathcal{T}$,*

$$[M, N]^T = [M, N^T].$$

Proof. This follows from the definition and Lemma 11.2.7 because, for any $t \geq 0$,

$$[M, N]_t^T = \langle M^c, N^c\rangle_t^T + \sum_{0<s\leq T\wedge t} \Delta M_s \Delta N_s = [M, N^T]_t\,.$$

□

A similar argument to that of Lemma 11.2.7 yields the following.

Lemma 11.2.9. *The brackets $\langle\cdot,\cdot\rangle$ and $[\cdot,\cdot]$ are a.s. symmetric and bilinear, that is, for any $L, M, N \in \mathcal{H}^2$, any $\alpha \in \mathbb{R}$, $\langle M, N\rangle = \langle N, M\rangle$ and*

$$\langle \alpha L + M, N\rangle = \alpha\langle L, N\rangle + \langle M, N\rangle,$$

and similarly for $[\cdot,\cdot]$. (However, the null set where these properties fail may depend on L, M, N and α.)

Remark 11.2.10. In multiple dimensions, that is when M is $\mathbb{R}^m$ valued and N is $\mathbb{R}^n$ valued, both with components in $\mathcal{H}^2$, it is natural to define the covariation *matrix* to be the matrix process $[M, N]$ with components $[M, N]^{ij} = [M_i, N_j]$. Similarly for $\langle M, N\rangle$, which is the compensator of the process $MN^\top$ (in the sense that $MN^\top - \langle M, N\rangle$ is a matrix process with martingale components, with $N^\top$ denoting the transpose of N). One can also show (see Lemma 12.5.3) that a version of $[M, N]$ and $\langle M, N\rangle$ exists which is symmetric and positive semidefinite up to indistinguishability.

11.3 Localization

We now consider how these results work for more general (local) martingales. In $\mathcal{H}^2_{\text{loc}}$, the theory is reasonably straightforward, and we consider this case first.

Lemma 11.3.1. *Let $M \in \mathcal{H}^2_{\text{loc}}$. Then there exists a unique predictable process $\langle M \rangle \in \mathcal{A}^+_{\text{loc}}$, called the predictable quadratic variation, such that $M^2 - \langle M \rangle \in \mathcal{H}^1_{\text{loc}} = \mathcal{M}_{\text{loc}}$ and $\langle M \rangle_0 = 0$.*

Proof. Let $\{T_n\}_{n\in\mathbb{N}}$ be a sequence which localizes M, that is, an increasing sequence of stopping times $\{T_n\}_{n\in\mathbb{N}}$, such that $\lim_n T_n = \infty$ a.s. and $M^{T_n} \in \mathcal{H}^2$, for all n.

We know that the decomposition of Corollary 10.2.16,

$$M^{T_n} = (M^{T_n})^c + (M^{T_n})^d,$$

is unique, so, if $T_n \le T_m$, then $(M^{T_n})^c_t = (M^{T_m})^c_t$ and $(M^{T_n})^d_t = (M^{T_m})^d_t$, for $(t,\omega) \in [\![0, T_n]\!]$. Furthermore, the predictable quadratic variation process is unique, so

$$\langle M^{T_n} \rangle_t = \langle M^{T_m} \rangle_t = \langle M^{T_n}, M^{T_m} \rangle_t \quad \text{for} \quad (t,\omega) \in [\![0, T_n]\!].$$

We can, therefore, define the predictable quadratic variation process of $M \in \mathcal{H}^2_{\text{loc}}$ as the unique process $\langle M \rangle \in \mathcal{A}^+_{\text{loc}}$ such that

$$\langle M \rangle^{T_n}_t = \langle M^{T_n} \rangle_t \qquad \text{for all } n \in \mathbb{N}. \qquad \square$$

Given a continuous process is always in $\mathcal{H}^2_{\text{loc}}$ (Lemma 10.3.1), and the jumps of a local martingale are square summable (Lemma 10.3.6), we can define the optional quadratic variation for any process in $\mathcal{M}_{\text{loc}}$, not only for those in $\mathcal{H}^2_{\text{loc}}$.

Definition 11.3.2. *Suppose $M \in \mathcal{M}_{\text{loc}}$ and let $M = M^c + M^d$ be its unique decomposition into a continuous local martingale and a purely discontinuous local martingale (Theorem 10.3.4). Then the* optional quadratic variation $[M]$ *of M is the increasing process defined by*

$$[M]_t = \langle M^c \rangle_t + \sum_{0<s\le t} \Delta M_s^2 \,.$$

If $M, N \in \mathcal{M}_{\text{loc}}$, we define

$$\begin{aligned}[M,N]_t &= \frac{1}{2}\left([M+N]_t - [M]_t - [N]_t \right) \\ &= \langle M^c, N^c \rangle_t + \sum_{0<s\le t} \Delta M_s \Delta N_s \,.\end{aligned}$$

Remark 11.3.3. We need to be careful here, as $[M]$ exists for all $M \in \mathcal{M}_{\mathrm{loc}}$, but $\langle M \rangle$ exists only if $M \in \mathcal{H}^2_{\mathrm{loc}}$. One might hope to extend the definition of $\langle M \rangle$ by *defining* $\langle M \rangle = \Pi^*_p\big([M]\big)$, however $\langle M \rangle$ is then infinite unless $[M] \in \mathcal{A}^+_{\mathrm{loc}}$, which is equivalent to assuming $M \in \mathcal{H}^2_{\mathrm{loc}}$. (See also Definition 11.4.4.)

Lemma 11.3.4. *For $M \in \mathcal{H}^2_{\mathrm{loc}}$, we have $M^2 - [M] \in \mathcal{M}_{\mathrm{loc}}$, $[M] - \langle M \rangle \in \mathcal{M}_{\mathrm{loc}}$, and*

$$E\big[[M]_\infty\big] = E\big[\langle M \rangle_\infty\big] \leq \infty.$$

Proof. From Lemma 11.3.1 we know $M^2 - \langle M \rangle$ is a local martingale and applying Remark 11.1.5 locally shows $M^2 - [M]$ is a local martingale. Taking a difference, we have $[M] - \langle M \rangle \in \mathcal{M}_{0,\mathrm{loc}}$.

Taking an appropriate sequence of stopping times $T_n \uparrow \infty$, we have $E\big[[M]_{T_n} - \langle M \rangle_{T_n}\big] = 0$, and as $[M]$ and $\langle M \rangle$ are increasing processes, by monotone convergence

$$E\big[[M]_\infty\big] = \lim_n E\big[[M]_{T_n}\big] = \lim_n E\big[\langle M \rangle_{T_n}\big] = E\big[\langle M \rangle_\infty\big].$$

□

Remark 11.3.5. By rearrangement, we see that, for $M, N \in \mathcal{H}^2_{\mathrm{loc}}$,

$$\begin{aligned} MN - [M,N] &= \frac{1}{2}\big((M+N)^2 - M^2 - N^2\big) - \frac{1}{2}\big([M+N] - [M] - [N]\big) \\ &= \frac{1}{2}\Big(\big((M+N)^2 - [M+N]\big) - \big((M)^2 - [M]\big) - \big((N)^2 - [N]\big)\Big) \\ &\in \mathcal{M}_{0,\mathrm{loc}}, \end{aligned}$$

where the last line follows from Lemma 11.3.4. Similarly $[M,N] - \langle M,N \rangle$ and $MN - \langle M,N \rangle$ are in $\mathcal{M}_{0,\mathrm{loc}}$.

We can now (partially) extend Lemma 11.3.4 to the case where M is a general local martingale (rather than requiring $M \in \mathcal{H}^2_{\mathrm{loc}}$).

Lemma 11.3.6. *For $M, N \in \mathcal{M}_{\mathrm{loc}}$, we have $MN - [M,N] \in \mathcal{M}_{\mathrm{loc}}$.*

Proof. Using Theorem 10.3.3, localization and the rearrangement used in Remark 11.3.5, we can reduce our problem to three cases,

(i) $M, N \in \mathcal{H}^2_{\mathrm{loc}}$,
(ii) M, N both martingales of integrable variation,
(iii) M a martingale with bounded jumps, N a martingale of integrable variation.

Case (i) is the result of Lemma 11.3.4, and Case (ii) follows from Exercise 8.4.2 and the product rule of Theorem 1.3.43.

To consider Case (iii), observe that, by localizing, we can assume that M is a bounded martingale and $N \in \mathcal{A}$. Hence $[M,N]_t = \sum_{s\leq t} \Delta M_s \Delta N_s$, and we can write

$$M_tN_t - [M,N]_t = M_tN_t - \int_{[0,t]} M_s dN_s + \int_{[0,t]} M_{s-} dN_s.$$

As in Case (ii), we see from Exercise 8.4.2 that $\{\int_{[0,t]} M_{s-}dN_s\}_{t\geq 0}$ is a local martingale. Finally, Corollary 8.1.22 states that, after localization, for any stopping time T,

$$E\Big[M_TN_T - \int_{[0,T]} M_t dN_t\Big] = 0,$$

and Theorem 5.4.6 then implies that $MN - \int_{[0,\cdot]} M_s dN_s$ is a local martingale. □

Remark 11.3.7. For M a martingale in $\mathcal{H}^2_{\text{loc}}$, we can show (Exercise 11.7.6) that the maps

$$\mu_p : A \mapsto E\Big[\int I_A \, d\langle M\rangle\Big]$$
$$\mu_o : A \mapsto E\Big[\int I_A \, d[M]\Big]$$

are σ-finite measures on the progressive σ-algebra Σ_π, and that their restrictions to the predictable σ-algebra Σ_p agree. For simplicity, we sometimes write these measures as $d\mu_p = d\langle M\rangle \times dP$ and $d\mu_o = d[M] \times dP$.

11.4 The Kunita–Watanabe Inequality

Considering the quadratic variation as a type of inner product, the following result provides a useful variant of the Cauchy–Schwarz inequality. In the following, we shall write $\int(\cdot)d\langle M\rangle = \int_{[0,\infty[}(\cdot)d\langle M\rangle$ for the integral over the set $[0,\infty[$, for notational simplicity.

Theorem 11.4.1 (Kunita–Watanabe Inequality). *Suppose $M, N \in \mathcal{M}_{\text{loc}}$ and H and K are $\mathcal{B}\otimes\mathcal{F}$-measurable processes. Then, almost surely,*

$$\int |H_s|\,|K_s|\,|d[M,N]_s| \leq \Big(\int H_s^2 d[M]_s\Big)^{1/2}\Big(\int K_s^2 d[N]_s\Big)^{1/2}$$

and, if $\langle M\rangle$, $\langle N\rangle$ and $\langle M,N\rangle$ all exist,

$$\int |H_s|\,|K_s|\,|d\langle M,N\rangle_s| \leq \Big(\int H_s^2 d\langle M\rangle_s\Big)^{1/2}\Big(\int K_s^2 d\langle N\rangle_s\Big)^{1/2}.$$

Proof. We prove the first case only, the predictable variations case is identical. Take $s,t \in [0,\infty]$, with $s < t$. Then for any $\lambda \in \mathbb{R}$, we know $[M+\lambda N]$ is increasing, that is,

$$[M+\lambda N]_t - [M+\lambda N]_s \geq 0 \quad \text{a.s.}$$

Writing $\delta_s^t X = X_t - X_s$ for any process X, this implies that

$$\delta_s^t[M] + 2\lambda\delta_s^t[M,N] + \lambda^2\delta_s^t[N] \geq 0 \quad \text{a.s.}$$

and, taking

$$\lambda = -\left(\frac{\delta_s^t[M]}{\delta_s^t[N]}\right)^{1/2}$$

and rearranging, we have

$$\left|\delta_s^t[M,N]\right| \leq \left(\delta_s^t[M]\right)^{1/2}\left(\delta_s^t[N]\right)^{1/2} \qquad \text{a.s.}$$

for each value of t and s. The processes $[M]$, etc., are right continuous (by definition), so we can conclude that this inequality holds except on an evanescent set (by Lemma 3.2.10).

Consider a finite subdivision of $[0,\infty]$, say $0 = t_0 < t_1 < \cdots < t_n = \infty$, and a finite number of random variables $\{H_i, K_i\}_{i=0}^n$, all of which are bounded in absolute value by 1, almost surely. Put

$$H_t = \sum_{i=1}^n H_{t_i} I_{]t_{i-1},t_i]}(t) \quad \text{and} \quad K_t = \sum_{i=1}^n K_{t_i} I_{]t_{i-1},t_i]}(t).$$

Write the above inequality for $s = t_{i-1}$, $t = t_i$, multiply by $|H_{t_i}K_{t_i}|$, sum and apply the Cauchy–Schwarz inequality. Then, almost surely,

$$\begin{aligned}
&\int |H_s K_s|\,|d[M,N]_s| \\
&= \sum_{i=0}^n |H_{t_i}K_{t_i}|\,|\delta_{t_i}^{t_{i+1}}[M,N]| \\
&\leq \sum_{i=0}^n (H_t^2\delta_{t_i}^{t_{i+1}}[M])^{1/2}(K_{t_i}^2\delta_{t_i}^{t_{i+1}}[N])^{1/2} \\
&\leq \Big(\sum_{i=0}^n H_{t_i}^2\delta_{t_i}^{t_{i+1}}[M]\Big)^{1/2} \cdot \Big(\sum_{i=0}^n K_{t_i}^2\delta_{t_i}^{t_{i+1}}[N]\Big)^{1/2} \\
&= \Big(\int H_s^2 d[M]_s\Big)^{1/2}\Big(\int K_s^2 d[N]_s\Big)^{1/2}.
\end{aligned}$$

The left continuous step processes of the same form as H and K are an algebra, and they generate the product σ-algebra $\mathcal{B}\otimes\mathcal{F}$ on $[0,\infty]\times\Omega$. A monotone class argument (Theorem 7.4.1) then establishes the validity of the above inequality for all bounded processes H and K. That is, for $H, K \in B(\mathcal{B}\otimes\mathcal{F})$,

$$\int |H_s K_s|\,|d[M,N]_s| \leq \Big(\int H_s^2 d[M]_s\Big)^{1/2}\Big(\int K_s^2 d[N]_s\Big)^{1/2} \qquad \text{a.s.}$$

For general $\mathcal{B}\otimes\mathcal{F}$-measurable processes H and K, monotone convergence and approximation by bounded processes gives the result. $\square$

By taking an expectation and applying Hölder's inequality, we obtain the following corollary.

Corollary 11.4.2. *For H, K, M and N as in Theorem 11.4.1, if $1 < p < \infty$ and $p^{-1} + q^{-1} = 1$ then*

$$E\Big[\int |H_s|\,|K_s|\,|d[M,N]_s|\Big] \\ \leq \Big\|\Big(\int H_s^2 d[M]_s\Big)^{1/2}\Big\|_p \cdot \Big\|\Big(\int K_s^2 d[N]_s\Big)^{1/2}\Big\|_q$$

and, if the predictable quadratic variations exist,

$$E\Big[\int |H_s|\,|K_s|\,|d\langle M,N\rangle_s|\Big] \\ \leq \Big\|\Big(\int H_s^2 d\langle M\rangle_s\Big)^{1/2}\Big\|_p \cdot \Big\|\Big(\int K_s^2 d\langle N\rangle\Big)^{1/2}\Big\|_q.$$

Theorem 11.4.3. *(i) Suppose $M \in \mathcal{M}_{\mathrm{loc}}$ and N is any locally bounded martingale. Then the process $[M,N]$ is locally integrable.*

(ii) Suppose $M \in \mathcal{M}_{0,\mathrm{loc}}$ and there is an increasing sequence of stopping times $\{T_n\}_{n\in\mathbb{N}}$ with $\lim_n T_n = \infty$, such that $E\big[[M,N]_{T_n}\big] = 0$ for every bounded martingale N. Then $M = 0$.

Proof. We can suppose $M_0 = 0$ a.s., otherwise consider the local martingale $M - M_0$. Write $M = U + V$ as in Theorem 10.3.3, where $U \in \mathcal{H}^2_{0,\mathrm{loc}}$ and V is a martingale in $\mathcal{A}_{0,\mathrm{loc}}$. Let T be a stopping time which reduces U, V and N, so $U^T \in \mathcal{H}^2_0$, $V^T \in \mathcal{A}_0$ and N^T is bounded. Then, by the second Kunita–Wantabe inequality (Corollary 11.4.2),

$$E\Big[\int_{[0,\infty]} |d[U^T, N^T]_s|\Big] < \infty,$$

and, from the definition of $[U^T, N^T]$,

$$E\Big[\int_{[0,\infty]} d[U^T, N^T]_s\Big] = E[U_T N_T].$$

From Corollary 10.2.9, $V^c \equiv 0$ so

$$[V^T, N^T]_t = \sum_{0<s\leq(t\wedge T)} \Delta V_s \Delta N_s\,.$$

Because V^T is of integrable variation and N^T is bounded, $\int_{[0,\infty]} |d[V^T, N^T]_s|$ is integrable. Therefore,

$$\int_{[0,T]} |d[M,N]_s| = \int_{[0,\infty]} |d[M^T, N^T]_s| \leq \int |d[U^T, N^T]_s| + \int_{[0,\infty]} |d[V^T, N^T]_s|,$$

so $\int_{[0,T]} |d[M,N]_s|$ is integrable, proving (i) above. Also, $E\big[\int_{[0,\infty]} d[V^T, N^T]_s\big] = E[V_T N_T]$ by Lemma 10.2.7.

Now $E\big[[M,N]_T\big] = 0$ and $M_0 = 0$ implies that $E[M_T N_T] = 0$. However, N_T is any bounded $\mathcal{F}_T$-measurable random variable, so $M_T = 0$ a.s. Because M^T is a uniformly integrable martingale, $M_0^T = 0$ a.s., thus proving (ii). □

Definition 11.4.4. *If $M, N \in \mathcal{M}_{\text{loc}}$ are such that $[M,N] \in \mathcal{A}_{\text{loc}}$, then $\langle M, N\rangle$ is defined to be the dual predictable projection of $[M,N]$ (Definition 8.3.4).*

Note that, from Theorem 11.4.3(i), $\langle M, N\rangle$ is defined whenever $M \in \mathcal{M}_{\text{loc}}$ and N is a local martingale which is locally bounded. It is easy to check that this agrees with the definition given in Lemma 11.3.1 whenever $M, N \in \mathcal{H}^2_{\text{loc}}$.

The following lemma provides a useful method for determining whether a local martingale is a square integrable martingale.

Lemma 11.4.5. *Let M be a local martingale with $M_0 = 0$ and suppose there is a stopping time $T \leq \infty$ such that $E\big[[M]_T\big] < \infty$. Then M^T is a square integrable martingale. In particular, if $T = \infty$, then $M \in \mathcal{H}^2$, and if $E\big[[M]_T\big] = 0$, then $M^T = 0$ a.s.*

Proof. By considering a stopped process, we can assume that $T = \infty$. As $[M]_\infty$ is integrable, we know $\sum_t (\Delta M)_t^2$ is integrable, and so $M \in \mathcal{H}^2_{\text{loc}}$. From Lemma 11.3.4,

$$N_t = M_t^2 - [M]_t$$

is a local martingale, so there is an increasing sequence $\{T_n\}_{n\in\mathbb{N}}$ of stopping times such that $\lim_n T_n = \infty$ a.s. and N^{T_n} and M^{T_n} are uniformly integrable martingales. As $[M]$ is a nondecreasing process, we know

$$N_{T_n} + [M]_\infty = M_{T_n}^2 + ([M]_\infty - [M]_{T_n})$$

is the sum of two nonnegative random variables and, as N^{T_n} is a martingale,

$$E\big[M_{T_n}^2 + ([M]_\infty - [M]_{T_n})\big] = E\big[N_{T_n} + [M]_\infty\big] = E\big[[M]_\infty\big] < \infty$$

by assumption. Therefore, both $M_{T_n}^2$ and $[M]_{T_n}$ are integrable. By Doob's L^p inequality (Theorem 5.1.3) applied to the nonnegative submartingale $|M^{T_n}|$,

$$E\Big[\big(\sup_t\{|M_t^{T_n}|\}\big)^2\Big] \leq 4E[M_{T_n}^2] \leq 4E\big[[M]_\infty\big] < \infty.$$

Applying the monotone convergence theorem, we see that

$$\|M\|_{\mathcal{H}^2}^2 = E\Big[\Big(\sup_t\{|M_t|\}\Big)^2\Big] < \infty,$$

so M is bounded in $\mathcal{H}^2$ norm. As we know $\sup_{T\in\mathcal{T}_o} E[M_T^2] \leq \|M\|_{\mathcal{H}^2}^2$, we have that M is in class (D), and by Lemma 5.6.6, we see that M is a uniformly integrable martingale, and so is in $\mathcal{H}^2$. That $E\big[[M]_T\big] = 0$ implies $M = 0$ follows from Theorem 10.2.22. □

Lemma 11.4.6. *Suppose $M \in \mathcal{M}_{0,\mathrm{loc}}$. Then $[M]^{1/2}$ is a locally integrable process, that is, $[M]^{1/2} \in \mathcal{A}_{0,\mathrm{loc}}$.*

Proof. We know that $[M]_t = \langle M^c\rangle_t + \sum_{s\leq t}(\Delta M_s)^2$ so, by sublinearity of $x \mapsto x^{1/2}$,

$$[M]_t^{1/2} \leq \langle M^c\rangle_t^{1/2} + \Big(\sum_{s\leq t}(\Delta M_s)^2\Big)^{1/2}.$$

From Lemma 10.3.6 we know that $\Big(\sum_{s\leq t}(\Delta M_s)^2\Big)^{1/2}$ is locally integrable, and $\langle M^c\rangle$ is continuous (and hence locally integrable). The result follows. □

The following result ties together the quadratic variation and the convergence of a local martingale with a bound on the jumps (Fig. 11.2). This result is not true for general local martingales, as shown by Exercise 13.7.7.

Theorem 11.4.7. *Let M be a local martingale such that, for some $k > 0$, we have $E[(\Delta M_T)^2] < k$ for all stopping times T (for example, this holds if M is continuous or has uniformly bounded jumps). Let $A = \{\omega : [M]_\infty < \infty\}$. Then $I_A M$ converges to a finite limit a.s. (in other words, M converges on the set A) and $I_{A^c}M$ does not converge.*

Proof. Taking the stopping time $T_n = \inf\{t : [M]_t \geq n\}$, we observe that

$$E\big[(M_t^{T_n})^2\big] \leq E\big[[M]_\infty^{T_n}\big] \leq n + k < \infty$$

so M^{T_n} is a square integrable martingale, and so converges a.s. to a finite limit $M_\infty^{T_n}$. Observe that $\bigcup_n\{T_n = \infty\} = \big\{[M]_\infty < \infty\big\} = A$, and so $I_A M_\infty^{T_n}$ converges a.s. as $n \to \infty$, and equals the desired limit of $I_A M$.

Conversely, to consider the set A^c, take $S_n = \inf\{t : |M_t| > n\}$. Then M^{S_n} is a local martingale, and

$$E\big[[M]_{S_n}\big] = E[M_{S_n}^2] \leq 2(n^2 + k) < \infty.$$

Therefore, $[M]_{S_n}$ is almost surely finite, and $[M]_\infty$ is a.s. finite on the set

$$\big\{\sup_t |M_t| < n\big\} \subseteq \{S_n = \infty\}.$$

It follows that $[M]_\infty$ must be finite a.s. on the set $\{M_\infty \text{ exists }\} \subseteq \bigcup_n\{\sup_t |M_t| < n\}$. Therefore, M does not converge on the set $\{[M]_\infty = \infty\} = A^c$. □

11.5 The Burkholder–Davis–Gundy Inequality

We now derive a final inequality for local martingales, which can be seen as a complement to Doob's L^p inequality (Theorem 4.5.6) and relates the

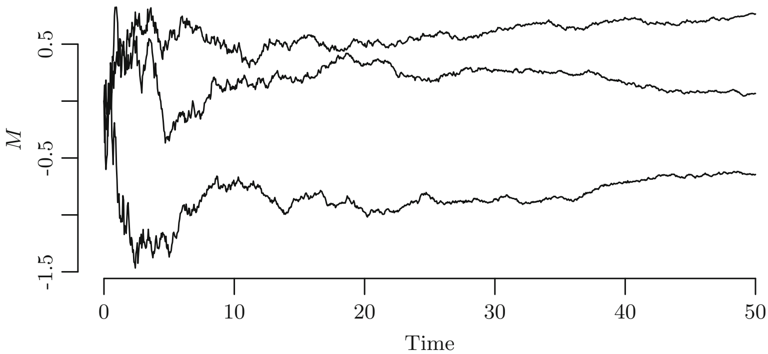

Fig. 11.2. Three paths of a continuous martingale with quadratic variation $\langle M\rangle_t = \int_{[0,t]}(1+s^2)^{-1}ds = \arctan(t)$.

maximum of a martingale with its quadratic variation. The key result of this section is Theorem 11.5.5. Our approach is to give a direct proof, which is conceptually simple but not particularly elegant. A beautiful proof is due to Garsia (see [85], or Dellacherie and Meyer [54]), using properties of BMO martingales. A recent proof due to Beiglböck and Siorpaes [6] obtains an inequality in discrete time; however does so 'pathwise', that is, for each ω, rather than in expectation.

The associated proofs may be omitted on a first reading.

We begin with the 'good-λ inequality', which is related to Lemma 8.2.18.

Lemma 11.5.1 (Good-λ Inequality). *Let $\beta > 0$ be a constant, and $\psi : \mathbb{R}^+ \to \mathbb{R}$ be a function satisfying $\psi(\delta) \to 0$ as $\delta \to 0$. For any $p \in\,]0,\infty[$, there exists a positive constant C_p such that, for any random variables X and Y satisfying*

$$P\big(X > \beta\lambda, Y < \delta\lambda\big) \le \psi(\delta)P(X \ge \lambda) \quad \textit{for all } \delta,\lambda \ge 0$$

we know

$$E[X^p] \le C_p E[Y^p].$$

If there exist bounds of this type for some $\beta > 0$ and some appropriate function ψ, then we say that the pair (X,Y) satisfies a 'good-λ inequality'.

Proof. From Theorem 1.3.45 and integration by parts, for any nonnegative random variable Z we know that

$$\int_{[0,\infty]} \lambda^{p-1}P(Z \ge \lambda)d\lambda = E\Big[\int_{[0,\infty]} I_{\{Z\ge\lambda\}}\lambda^{p-1}d\lambda\Big] = p^{-1}E[Z^p].$$

By assumption, we know

$$\begin{aligned} P(X > \beta\lambda) &\le P\big(X > \beta\lambda, Y < \delta\lambda\big) + P(Y \ge \delta\lambda) \\ &\le \psi(\delta)P(X \ge \lambda) + P(Y \ge \delta\lambda), \end{aligned}$$

and multiplying by λ^{p-1} and integrating gives

$$E\big[(X/\beta)^p\big] \le \psi(\delta)E[X^p] + E[(Y/\delta)^p].$$

For δ sufficiently small that $\psi(\delta) < \beta^{-p}$, by rearrangement,

$$E[X^p] \le \big(\delta^p(\beta^{-p} - \psi(\delta)\big)^{-1}E[Y^p]$$

as desired. □

Lemma 11.5.2. *Let X and Y be nonnegative processes such that $X-Y$ is a local martingale, and suppose there exists an adapted, left-continuous process Z such that $X-Y+Z \ge 0$. Suppose T is a stopping time with $X = Y = Z = 0$ on $[\![0,T]\!]$. Then, for all $\beta, \delta > 0$,*

$$P\big(X^*_\infty > \beta, Z^*_\infty < \delta\big) \le \frac{\delta}{\beta}P(T < \infty).$$

Proof. Let $T_1 = \inf\{t : X_t - Y_t \ge \beta - \delta\}$, $T_2 = \inf\{t : Z_t \ge \delta\}$ and $R = T_1 \wedge T_2$. On the set $\{X^*_\infty > \beta, Z^*_\infty < \delta\}$, we have $R = T_1$ and $X_R - Y_R \ge \beta - \delta$. We know $X^R - Y^R$ is a local martingale, and is bounded below by $-Z^R \ge -\delta$. Therefore, by Exercise 9.4.5, it is a supermartingale. By optional stopping,

$$\beta P\big(X^*_\infty > \beta, Z^*_\infty < \delta\big|\mathcal{F}_T\big) \le E\big[X_R - Y_R + \delta\big|\mathcal{F}_T\big] \le \delta$$

and so

$$\begin{aligned} P\big(X^*_\infty > \beta, Z^*_\infty < \delta\big) &= E\big[P\big(X^*_\infty > \beta, Z^*_\infty < \delta\big|\mathcal{F}_T\big)\big]P(T < \infty) \\ &\le \frac{\delta}{\beta}P(T < \infty). \end{aligned}$$

□

Lemma 11.5.3. *Let M be a local martingale such that $|\Delta M| \le L$ for some adapted, left-continuous process L. Then the pairs $(M^*_\infty, [M]^{1/2}_\infty + L^*_\infty)$ and $([M]^{1/2}_\infty, M^*_\infty + L_\infty)$ both satisfy good-λ inequalities.*

Proof. For any λ, let $T = \inf\{t : |M_t| \ge \lambda\}$, so that $P(T < \infty) \le P(M^*_\infty \ge \lambda)$. We know that $[M]$ is càdlàg, so

$$[M]_s \le L_s^2 + [M]_{s-}$$

and the right-hand side is a predictable process. Applying Lemma 11.5.2 with

$$X = (M - M^T)^2, \qquad Y = [M] - [M]^T, \qquad Z_s = (L^2 + [M]_{s-}) - (L^2 + [M]_{s-})^T,$$

we see $X - Y + Z \geq 0$ and so

$$P\big((M - M^T)^*_\infty > \beta\lambda, Z^*_\infty < \delta^2\lambda^2\big) = P\big(X^*_\infty > \beta^2\lambda^2, Z^*_\infty < \delta^2\lambda^2\big) \\ \leq \Big(\frac{\delta}{\beta}\Big)^2 P(M^*_\infty \geq \lambda).$$

Now, on the event $A = \big\{M^*_\infty > \beta\lambda, [M]^{1/2}_\infty + L^*_\infty < \delta\lambda\big\}$, we have

$$(Z^*)^{1/2} \leq [M]^{1/2}_\infty + L^*_\infty < \delta\lambda,$$
$$(\Delta M)^*_\infty \leq [M]^{1/2}_\infty \leq (Z^*_\infty)^{1/2} \leq [M]^{1/2}_\infty + L^*_\infty < \delta\lambda,$$
$$(M - M^T)^*_\infty \geq M^*_\infty - M^*_T \geq M^*_\infty - \lambda - (\Delta M)^*_\infty > \beta\lambda - \lambda - \delta\lambda.$$

In the third inequality, we have used the fact $M^*_T \leq M^*_{T-} + |\Delta M|_T \leq \lambda + (\Delta M)^*_\infty$. Therefore,

$$P(A) \leq P\big((M - M^T)^*_\infty > (\beta - 1 - \delta)\lambda, (Z^*_\infty)^{1/2} < \delta\lambda\big) \\ \leq \Big(\frac{\delta}{\beta - 1 - \delta}\Big)^2 P(M^*_\infty \geq \lambda),$$

and so $(M^*_\infty, [M]^{1/2}_\infty + L^*_\infty)$ satisfies a good-λ inequality with $\beta > 1$ and $\psi(\delta) = \delta^2(\beta - 1 - \delta)^{-2}$.

Similarly, to prove that $([M]^{1/2}_\infty, M^*_\infty + L^*_\infty)$ satisfies a good-λ inequality, let $T = \inf\{t : [M]_t \geq \lambda^2\}$. Applying Lemma 11.5.2 with

$$X = [M] - [M]^T, \qquad Y = (M - M^T)^2,$$
$$Z_s = 4\Big((M^*_{s-} + L_s)^2 - \big((M^*_{s-} + L_s)^2\big)^T\Big) \geq Y_s,$$

we obtain

$$P\big([M]_\infty - [M]_T > \beta\lambda, Z^*_\infty < \delta^2\lambda^2\big) = P\big(X^*_\infty > \beta^2\lambda^2, Z^*_\infty < \delta^2\lambda^2\big) \\ \leq \Big(\frac{\delta}{\beta}\Big)^2 P\big([M]^{1/2}_\infty \geq \lambda\big).$$

On the event $B = \big\{[M]^{1/2}_\infty > \beta\lambda, M^*_\infty + L^*_\infty < \delta\lambda\big\}$ we see

$$Z^*_\infty \leq 4(M^*_\infty + L^*_\infty)^2 < 4\delta^2\lambda^2,$$
$$\big((\Delta M)^*_\infty\big)^2 \leq (L^*_\infty)^2 \leq (M^*_\infty + L^*_\infty)^2 < \delta^2\lambda^2,$$
$$[M]_\infty - [M]_T \geq [M]_\infty - \lambda^2 - \big((\Delta M)^*_\infty\big)^2 > \beta^2\lambda^2 - \lambda^2 - \delta^2\lambda^2,$$

and so

$$P(B) \leq P\big([M]_\infty - [M]_T > (\beta^2 - 1 - \delta^2)\lambda^2, Z^*_\infty < 4\delta^2\lambda^2\big) \\ \leq \Big(\frac{4\delta}{\beta^2 - 1 - \delta^2}\Big)^2 P\big([M]^{1/2}_\infty \geq \lambda\big).$$

Hence $([M]^{1/2}_\infty, M^*_\infty + L^*_\infty)$ satisfies a good-λ inequality with $\beta > 1$ and $\psi(\delta) = 4\delta^2(\beta^2 - 1 - \delta^2)^{-1}$. $\square$

We now give a variant of Theorem 10.3.3, which will allow us to use Lemma 11.5.2, by decomposing a local martingale into a local martingale with jumps bounded below by a left-continuous process and a process with reasonable integrability properties.

Lemma 11.5.4. *For any $p \in [1,\infty[$ there exists a constant K_p such that, for any $\epsilon > 0$, a local martingale M has a decomposition $M = U + V$ where V is a pure-jump local martingale with*

$$E\Big[\Big(\int_{[0,\infty]} |dV|\Big)^p\Big]^{1/p} \leq K_p E\big[\big((\Delta M)^*_\infty\big)^p\big]^{1/p}$$

and U is a local martingale with

$$|\Delta U_s| \leq 4\big((\Delta M)^*_{s-} \vee \epsilon\big).$$

Proof. For notational simplicity, let $X = (\Delta M)^* \vee \epsilon$. Define

$$\tilde{V}_t = \sum_{s \leq t} \Delta M_s I_{\{|\Delta M_s| > 2X_{s-}\}}.$$

Let $T_1, T_2, \ldots$ be the increasing sequence of stopping times corresponding to the jumps of $\tilde{V}$. These are well ordered, as $X \geq \epsilon$. By construction, for any n,

$$|\Delta \tilde{V}_{T_n}| = X_{T_n} \leq X_{T_{(n+1)-}} \leq \frac{1}{2}|\Delta \tilde{V}_{T_{n+1}}|.$$

As $|\Delta \tilde{V}_{T_n}| \leq (\Delta M)^*_\infty$, it follows that

$$\sum_{n\in\mathbb{N}} |\Delta \tilde{V}_{T_n}| \leq \sum_{n\in\mathbb{N}} 2^{-n} (\Delta M)^*_\infty = 2(\Delta M)^*_\infty.$$

Therefore, $\tilde{V}$ is of finite variation.

As M is càdlàg and a local martingale, $\tilde{V}$ is locally integrable, so we can write $V = \tilde{V} - \Pi^*_p \tilde{V}$. By Theorem 8.2.19, for any $p \in [1,\infty[$ there exists a constant C_p such that

$$\begin{aligned} E\Big[\Big(\int_{[0,\infty]} |dV|\Big)^p\Big]^{1/p} &\leq E\Big[\Big(\int_{[0,\infty]} |d\tilde{V}|\Big)^p\Big]^{1/p} + E\Big[\Big(\int_{[0,\infty]} |d(\Pi^*_p \tilde{V})|\Big)^p\Big]^{1/p} \\ &\leq 2(1 + C_p) E\big[\big((\Delta M)^*_\infty\big)^p\big]^{1/p}. \end{aligned}$$

Writing $U = M - V$, by the same argument as in Theorem 10.3.3, on the set $\{\Delta \Pi^*_p \tilde{V} = 0\}$ we know that $|\Delta U_s| \leq 2X_{s-}$, and on the predictable thin set $\{\Delta \Pi^*_p \tilde{V} \neq 0\}$ we know that $\Delta U_s \leq 4X_{s-}$. The result follows, with $K_p = 2(1 + C_p)$. $\square$

Theorem 11.5.5 (Burkholder–Davis–Gundy (BDG) Inequality). *For any $1 \leq p < \infty$, there exist constants $c_p, C_p > 0$ such that, for any local martingale M with $M_0 = 0$,*

$$c_p E\big[[M]_\infty^{p/2}\big]^{1/p} \leq E\big[(M_\infty^*)^p\big]^{1/p} \leq C_p E\big[[M]_\infty^{p/2}\big]^{1/p}.$$

Proof. First take $M = U + V$ using the decomposition in Lemma 11.5.4, for an arbitrary $\epsilon > 0$. By Lemma 11.5.3, we know that $\big(U_\infty^*, [U]_\infty^{1/2} + L_\infty\big)$ and $\big([U]_\infty^{1/2}, U_\infty^* + L_\infty\big)$ both satisfy good-λ inequalities, where $L_t = 4\big((\Delta M)_{t-}^* \vee \epsilon\big)$. Therefore, there exist constants $C_{1,p}, C_{2,p}$ (independent of M) such that

$$\begin{aligned} E\big[(U_\infty^*)^p\big]^{1/p} &\leq C_{1,p}\|[U]_\infty^{1/2} + L_\infty\|_p, \\ E\big[[U]_\infty^{p/2}\big]^{1/p} &\leq C_{2,p}\|U_\infty^* + L_\infty\|_p. \end{aligned}$$

We also know

$$\|[V]_\infty^{1/2}\|_p = \Big\|\Big(\sum_s |\Delta V_s|^2\Big)^{1/2}\Big\|_p \leq \Big\|\sum_s |\Delta V_s|\Big\|_p \leq K_p\big\|(\Delta M)_\infty^*\big\|_p.$$

Using the fact $M = U + V$, we can combine these results to see that

$$\begin{aligned} \|M_\infty^*\|_p &\leq \|U_\infty^*\|_p + \|V_\infty^*\|_p \\ &\leq C_{1,p}\big\|[U]_\infty^{1/2} + 4\big((\Delta M)_\infty^* \vee \epsilon\big)\big\|_p + K_p\big\|(\Delta M)_\infty^*\big\|_p \\ &\leq C_{1,p}\big\|[U]_\infty^{1/2}\big\|_p + (K_p + 4C_{1,p})\big\|(\Delta M)_\infty^*\big\|_p + 4C_{1,p}\epsilon. \end{aligned}$$

Similarly, using the fact $U = M - V$, we know

$$\begin{aligned} [U]_t &= \langle M^c\rangle_t + \sum_{s\leq t}(\Delta M_s - \Delta V_s)^2 \leq \langle M^c\rangle_t + 2\sum_{s\leq t}\big((\Delta M_s)^2 + (\Delta V_s)^2\big) \\ &\leq 2[M]_t + 2[V]_t. \end{aligned}$$

Hence we have

$$\frac{1}{2}\|[U]_\infty^{1/2}\|_p \leq \|[M]_\infty^{1/2}\|_p + \|[V]_\infty^{1/2}\|_p \leq \|[M]_\infty^{1/2}\|_p + K_p\big\|(\Delta M)_\infty^*\big\|_p$$

and so

$$\|M_\infty^*\|_p \leq 2C_{1,p}\|[M]_\infty^{1/2}\|_p + (K_p + 4C_{1,p} + 2K_pC_{1,p})\big\|(\Delta M)_\infty^*\big\|_p + 4C_{1,p}\epsilon.$$

Finally, note that $(\Delta M)_\infty^* \leq [M]_\infty^{1/2}$, so letting $\epsilon \downarrow 0$ and $C_p = C_{1,p} + K_p + 4C_{1,p} + 2K_pC_{1,p}$ we have the second desired inequality,

$$\|M_\infty^*\|_p \leq C_p\|[M]_\infty^{1/2}\|_p.$$

To prove the first inequality, we know $[M] \leq 2[U] + 2[V]$, so we write

$$\begin{aligned}\frac{1}{2}\|[M]_\infty^{1/2}\|_p &\leq \|[U]_\infty^{1/2}\|_p + \|[V]_\infty^{1/2}\|_p \\ &\leq C_{2,p}\big\|U_\infty^* + 4\big((\Delta M)_\infty^* \vee \epsilon\big)\big\|_p + K_p\big\|(\Delta M)_\infty^*\big\|_p \\ &\leq C_{2,p}\|U_\infty^*\|_p + (K_p + 4C_{2,p})\big\|(\Delta M)_\infty^*\big\|_p + 4C_{2,p}\epsilon\end{aligned}$$

and, as $U^* \leq M^* + V^*$,

$$\|U_\infty^*\|_p \leq \|M_\infty^*\|_p + K_p\big\|(\Delta M)_\infty^*\big\|_p.$$

Finally, notice that $(\Delta M)^* \leq 2M^*$, so

$$\frac{1}{2}\|[M]_\infty^{1/2}\|_p \leq C_{2,p}\|M_\infty^*\|_p + 2K_pC_{2,p}\big\|M_\infty^*\big\|_p + 2(K_p + 4C_{2,p})\big\|M_\infty^*\big\|_p + 4C_{2,p}\epsilon.$$

Taking $\epsilon \downarrow 0$ and defining $c_p^{-1} = 2C_{2,p} + 4K_pC_{2,p} + 4(K_p + 4C_{2,p})$ gives the result. □

Remark 11.5.6. By considering stopped processes, we see that the Burkholder–Davis–Gundy inequality also holds for

$$c_pE\big[[M]_T^{p/2}\big]^{1/p} \leq E\big[(M_T^*)^p\big]^{1/p} \leq C_pE\big[[M]_T^{p/2}\big]^{1/p}$$

when T is any stopping time. By directly applying Lemma 11.5.3 and the good-λ inequality one can also show that, for *continuous* local martingales, this statement holds for $0 < p < \infty$.

Remark 11.5.7. By applying Doob's L^p inequality (Theorem 4.5.6, with localization and the monotone convergence theorem in the case of a local martingale), as $E[M_\infty^2] = E\big[[M]_\infty\big]$ for all $M \in \mathcal{H}_0^2$, we see that $C_2 = 4$ and $c_2 = 1$ satisfy the BDG inequality. Dellacherie and Meyer [54], following Garsia [85], show that in general one can take $c_p = 1/(4p)$ and $C_p = 6p$; however these values are not optimal. Osekowski [144] considers optimal values for these constants given some restrictions on the jumps.

Remark 11.5.8. When $M \in \mathcal{H}_{0,\mathrm{loc}}^2$, from Lemma 11.3.4 we see that, in the case $p = 2$, we can interchange the optional and predictable quadratic variations in the BDG inequality. By applying Theorems 8.2.19 and 8.2.20 to $\langle M\rangle = \Pi_p^*[M]$, we see that, for $p > 2$, we can find c_p, C_p such that

$$c_pE[\langle M\rangle_\infty^{p/2}]^{1/p} \leq E[(M_\infty^*)^p] \leq C_pE\big[\langle M\rangle_\infty^{p/2} + \sup_t(\Delta M_t)^p\big]^{1/p}.$$

Of course, in a continuous setting, we have $\langle M\rangle = [M]$, so the distinction between the inequality with the optional or predictable quadratic variation is irrelevant.

Remark 11.5.9. An immediate consequence of the BDG inequality is that, for any $p \in [1, \infty[$, the map $M \mapsto E\big[|M_0|^p + [M]_\infty^{p/2}\big]^{1/p}$ defines a seminorm on $\mathcal{M}$ equivalent to the $\mathcal{H}^p$ norm in Definition 10.1.2. This gives an alternative, and in some ways more natural, definition of the norm of $\mathcal{H}^p$. In particular, the convention for $p = \infty$ (where the norms are not equivalent, and we earlier left $\mathcal{H}^\infty$ undefined) is that

$$\mathcal{H}^\infty = \{M \in \mathcal{M}_{\mathrm{loc}} : \||M_0| + [M]_\infty\|_\infty < \infty\}$$

where $\|\cdot\|_\infty = \operatorname{ess\,sup}|\cdot|$ is the $L^\infty(\Omega)$ norm. We note immediately that any $\mathcal{H}^\infty$ local martingale is a uniformly integrable martingale, and is in $\mathcal{H}^p$ for every $p < \infty$.

Remark 11.5.10. It is easy to see from the definition that all bounded martingales are in $\mathcal{H}^p$ for every $p < \infty$. However, as we shall see in Exercise 11.7.10 (or simply consider a Brownian motion W stopped at $T = \inf\{t : |W_t| = 1\}$), it is *not* the case that all bounded martingales are in $\mathcal{H}^\infty$. Neither are all martingales in $\mathcal{H}^\infty$ bounded, simply consider a Brownian motion stopped at $T = 1$.

Using these estimates, we can obtain the following characterization of the purely discontinuous martingales. This will most often be used to specify a local martingale M by defining its jumps $X = \Delta M$.

Theorem 11.5.11. *Let X be an optional process with $X_0 = 0$. The following are equivalent.*

(i) There exists a local martingale M such that $\Delta M = X$.
(ii) The process defined by $Y_t := \big(\sum_{s\leq t} X_s^2\big)^{1/2}$ is in $\mathcal{A}^+_{\mathrm{loc}}$ and $\Pi_p X = 0$. (Here $\Pi_p X$ is defined as in Remark 7.6.4.)

In this case, there exists a unique purely discontinuous local martingale M such that $\Delta M = X$.

Proof. That (i) implies (ii) is the result of Lemma 10.3.6 and Corollary 7.6.6.

Conversely, suppose X is as stated, so our task is to construct M. By localization, we can suppose that $Y \in \mathcal{A}_0^+$, which implies that for every stopping time T, $X_T I_{\{T<\infty\}} \in L^1$ and $\{X \neq 0\}$ is a thin set. Using Theorem 6.2.9 and Theorem 7.5.2, we can find two sequences $\{T_n\}_{n\in\mathbb{N}}$ and $\{S_n\}_{n\in\mathbb{N}}$ such that each T_n is totally inaccessible, each S_n is predictable and $\{X \neq 0\} \subseteq \bigcup_{n\in\mathbb{N}}([\![T_n]\!] \cup [\![S_n]\!])$.

Write $A^{(n)} = X_{T_n} I_{[\![T_n,\infty[\![}$, so that $A^{(n)}$ is an optional integrable process which only charges the totally inaccessible stopping time T_n. Through a similar argument to that in Theorem 10.2.10, we see that $\Pi_p^* A^{(n)}$ is continuous and that $A^{(n)} - \Pi_p^* A^{(n)}$ is a martingale. Similarly, write $B^{(n)} = X_{S_n} I_{[\![S_n,\infty[\![}$. As S_n is predictable we know that

$$(\Pi_p^* B^{(n)})_{S_n} = E[X_{S_n}|\mathcal{F}_{S_n-}] = (\Pi_p^* X)_{S_n} = 0,$$

and hence $B^{(n)}$ is a martingale, through a similar argument to that in Theorem 10.2.11.

Define

$$M^{(n)} = \sum_{m \leq n} (A^{(m)} - \Pi_p^* A^{(m)} + B^{(m)})$$

so that $\Delta M^{(n)} = X$ on $[\![0, T_n \wedge S_n]\!]$, and $M^{(n)}$ is a purely discontinuous martingale. For any $m \leq n$,

$$[M^{(n)} - M^{(m)}]_\infty^{1/2} = \left(\sum_{0 \leq t < \infty} \left(X_t^2 \sum_{m < k \leq n} (I_{[\![T_n]\!]} + I_{[\![S_n]\!]}) \right) \right)^{1/2},$$

and by assumption, the term on the right tends to zero in L^1 as $m \wedge n \to \infty$. In particular, given the Burkholder–Davis–Gundy inequalities, we see that $M^{(n)}$ is a Cauchy sequence in $\mathcal{H}^1$, and therefore has a limit $M \in \mathcal{H}^1$. As convergence in $\mathcal{H}^1$ implies convergence uniformly in time (in L^1 with respect to ω, hence in probability, hence almost surely for a subsequence), we see that $\Delta M = \lim_{n \to \infty} \Delta M^{(n)} = X$.

Finally, the uniqueness of M among the purely discontinuous local martingales follows from Exercise 11.7.12. □

11.6 Semimartingales

We have now seen the importance of both processes of locally finite variation and local martingales. Unifying these is the class of processes called 'semimartingales', to which we now turn our attention.

Definition 11.6.1. *A process $X = \{X_t\}_{t \geq 0}$ is a* semimartingale *if it has a decomposition of the form*

$$X = X_0 + M + A\,,$$

where $M \in \mathcal{M}_{0,\mathrm{loc}}$ and $A \in \mathcal{V}_0$, that is, M is a local martingale and A is a càdlàg adapted process of almost surely finite variation, and $M_0 = A_0 = 0$. We write $\mathcal{S}$ for the space of semimartingales. Clearly, semimartingales are càdlàg and adapted.

Example 11.6.2. Clearly, local martingales and processes of finite variation are semimartingales. Furthermore, every right-continuous supermartingale (or submartingale) is a semimartingale, because a right-continuous supermartingale X has a Doob–Meyer decomposition $X = M - A$ with $M \in \mathcal{M}$ and $A \in \mathcal{A}$ (Theorem 9.2.7).

Lemma 11.6.3. *Every local semimartingale is a semimartingale.*

Proof. Suppose $\{T_n\}_{n\in\mathbb{N}}$ is a localizing sequence for a local semimartingale X. Then, for each n, we have a decomposition $X^{T_n} = X_0 + M^n + A^n$, where M^n is a local martingale and A^n a process of finite variation. Define the pasted processes M and A by

$$M_t := \sum_{n\in\mathbb{N}} I_{\{t\geq T_n\}}(M^n_{t\wedge T_{n+1}} - M^n_{T_n}),$$
$$A_t := \sum_{n\in\mathbb{N}} I_{\{t\geq T_n\}}(A^n_{t\wedge T_{n+1}} - A^n_{T_n}).$$

Because $T_n \to \infty$ a.s., we know that each of these sums is almost surely finite for each (t,ω), so these processes are well defined. It is easy to check that M is a local martingale (with localizing sequence $\{T_n\}_{n\in\mathbb{N}}$), A is a process of finite variation and $X = X_0 + M + A$. Therefore, X has the desired representation, and so is a semimartingale. □

Remark 11.6.4. Because there are local martingales which are processes of locally integrable (indeed, finite) variation, the decomposition of a semimartingale is not unique. However, consider two decompositions for a semimartingale X:

$$X_t = X_0 + M_t + A_t = X_0 + \overline{M}_t + \overline{A}_t\,.$$

Then $N = M - \overline{M} = A - \overline{A}$ is in $\mathcal{M}_{0,\mathrm{loc}}$, and by Lemma 10.3.8 is also a process of locally integrable variation. Applying Theorem 10.3.3, we can write $N = U + V$, where U, V are local martingales, $U \in \mathcal{H}^2_{\mathrm{loc}}$ and V is of locally integrable variation. This implies that $U = N - V$ is also of locally integrable variation, and so, by Theorem 10.2.21, $U \in \mathcal{H}^{2,d}_{\mathrm{loc}}$, that is, U has no continuous part. As V is also a pure jump martingale, this implies that N has no continuous martingale part (using Theorem 10.3.4). Therefore, $M^c = \overline{M}^c$, and the continuous part of the martingale in the decomposition of X is independent of the decomposition.

Definition 11.6.5. *Write*

$$X^c = M^c.$$

Then X^c is called the continuous martingale part *of X.*

Definition 11.6.6. *Suppose X is a semimartingale. Then the* optional quadratic variation *of X is the process*

$$[X]_t = \langle X^c\rangle_t + \sum_{0<s\leq t} (\Delta X_s)^2\,.$$

As before, if Y is a second semimartingale, define

$$[X,Y] = \frac{1}{2}\left([X+Y, X+Y] - [X,X] - [Y,Y]\right).$$

If $[X,Y] \in \mathcal{A}_{\mathrm{loc}}$, then the predictable quadratic variation *of X and Y is the process*

$$\langle X,Y\rangle = \Pi^*_p[X,Y].$$

By definition, we know $A \in \mathcal{V}_0$ and $\sum_{s\le t} \Delta A_s^2 \le \big(\sum_{s\le t} |\Delta A_s|\big)^2 < \infty$. From Lemma 10.3.6, we have $\sum_{s\le t} \Delta M_s^2 < \infty$. Therefore, $[X]_t$ is almost surely finite, as $(\Delta X_s)^2 \le 2\big((\Delta M_s)^2 + (\Delta A_s)^2\big)$.

Remark 11.6.7. It is easy to see that, given the definitions of $[X]$ and $\langle X\rangle$ for X a semimartingale, the Kunita–Watanabe inequality (Theorem 11.4.1) also holds for semimartingales.

Theorem 11.6.8. *Suppose $M \in \mathcal{M}_{0,\mathrm{loc}}$ and $A \in \mathcal{V}$. If A is predictable, then $[M, A] \in \mathcal{M}_{0,\mathrm{loc}}$.*

Proof. By definition
$$[M, A]_t = \sum_{0<s\le t} \Delta M_s \Delta A_s\,.$$
However, the local martingale M has a unique decomposition as the sum of a continuous local martingale M^c and a discontinuous martingale M^d. Furthermore, the discontinuous local martingale can be decomposed into the sum of a discontinuous local martingale M^{dp} which has only accessible jump times, and a discontinuous local martingale M^{dq} which has only totally inaccessible jump times.

Because A is predictable, we know $A = \Pi_p^* A$, the processes $\{A_t\}_{t\ge 0}$ and $\{A_{t-}\}_{t\ge 0}$ are locally bounded (Lemma 7.3.20) and A is a.s. continuous at every totally inaccessible stopping time. Therefore, with probability one,
$$[M, A]_t = \sum_{0<s\le t} \Delta M_s \Delta A_s = \sum_{0<s\le t} \Delta A_s \Delta M_s^{dp} = \int_{]0,t]} \Delta M_s dA_s.$$
After localizing to ensure integrability, for any stopping time T, we calculate
$$\begin{aligned} E\big[[M, A]_T\big] &= E\Big[\int_{]0,T]} \Delta M_s dA_s\Big] = E\Big[\int_{]0,T]} \Delta M_s d(\Pi_p^* A)_s\Big] \\ &= E\Big[\int_{]0,T]} \Pi_p(\Delta M)_s dA_s\Big] = 0, \end{aligned}$$
where $\Pi_p(\Delta M) = 0$ follows from Corollary 7.6.6. Applying Theorem 5.4.6 we conclude that $[M, A]$ is a local martingale. □

11.6.1 Special Semimartingales

For local supermartingales, the Doob–Meyer decomposition (Theorem 9.2.7) allowed us to write $X = M - A$, where A is a *predictable, locally integrable* process and M a local martingale. Generalizing this requirement leads to the class of ‘special semimartingales’.

Definition 11.6.9. *A semimartingale X with $X_0 \in L^1$ is called a* special semimartingale *if there is a decomposition*

$$X = X_0 + M + A,$$

in which A is locally of integrable variation, that is, $M \in \mathcal{M}_{0,\text{loc}}$, $A \in \mathcal{A}_{0,\text{loc}}$. We write $\mathcal{S}_{\text{Sp}}$ for the space of special semimartingales.

Theorem 11.6.10. *For a semimartingale X with $X_0 \in L^1$, the following are equivalent:*

(i) X is a special semimartingale,
(ii) the increasing process $(\sum_{0<s\leq t} \Delta X_s^2)^{1/2}$ is locally integrable,
(iii) for every semimartingale decomposition of X, the process A is locally of integrable variation,
(iv) there is a semimartingale decomposition of X in which A is predictable (and this decomposition is unique),
(v) the increasing process defined by $X_t^* = \sup_{s\leq t} |X_s|$ is locally integrable,
(vi) the increasing process defined by $D_t = \sup_{s\leq t} |\Delta X_s|$ is locally integrable.

Definition 11.6.11. *The decomposition of a special semimartingale as in Theorem 11.6.10(iv) is called the* canonical decomposition*.*

Proof. We shall show that (iv) $\Rightarrow$ (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) $\Rightarrow$ (iv) and (i) $\Rightarrow$ (v) $\Rightarrow$ (vi) $\Rightarrow$ (i).

(iv) $\Rightarrow$ (i). Here we are assuming that $X = X_0 + M + A$, where $M_0 = A_0 = 0$ a.s. and A is predictable. Therefore, the process $\{\int_{[0,t]} |dA_s|\}_{t\geq 0}$ is predictable (see Lemma 8.1.11), and so the stopping time

$$S_k = \inf\left\{t : \int_{[0,t]} |dA_s| \geq k\right\} > 0 \tag{11.1}$$

is predictable. Suppose $\{S_m^k\}_{m\in\mathbb{N}}$ is a sequence of stopping times which announce S_k. Then the variation of A on $[\![0, S_m^k]\!]$ is less than k and so integrable. Writing $T_n = \sup_{k\leq n, m\leq n} S_m^k$ we see that $\lim T_n = \infty$ and $A^{T_n} \in \mathcal{A}$, and so A is locally integrable.

(i) $\Rightarrow$ (ii). Write

$$X = X_0 + M + A\,.$$

Then, by sublinearity of $x \mapsto x^{1/2}$,

$$\begin{aligned}\Big(\sum_{s\leq t}(\Delta X_s)^2\Big)^{1/2} &\leq \Big(\sum_{s\leq t}(\Delta M_s)^2\Big)^{1/2} + \Big(\sum_{s\leq t}(\Delta A_s)^2\Big)^{1/2}\\ &\leq [M]_t^{1/2} + \sum_{s\leq t}|\Delta A_s|\\ &\leq [M]_t^{1/2} + \int_{[0,t]}|dA_s|,\end{aligned}$$

and this is locally integrable from Lemma 11.4.6, because A is assumed locally integrable.

(ii) ⇒ (iii). Clearly, for any decomposition $X = X_0 + A + M$,

$$\Big(\sum_{s\le t}(\Delta A_s)^2\Big)^{1/2} \le \Big(\sum_{s\le t}(\Delta X_s)^2\Big)^{1/2} + [M]_t^{1/2},$$

so, if (ii) holds, by Lemma 11.4.6 the left-hand side is an increasing locally integrable process. Suppose $\{R_n\}_{n\in\mathbb{N}}$ is an increasing sequence of stopping times, $\lim_n R_n = \infty$, such that

$$E\Big[\Big(\sum_{u\le R_n}(\Delta A_u)^2\Big)^{1/2}\Big] < \infty,$$

and let S_k be as in (11.1). If $T_n = R_n \wedge S_n$ then certainly

$$E\big[|\Delta A_{T_n}|\big] < \infty,$$

so

$$E\Big[\int_{[0,T_n]} |dA_s|\Big] \le n + E\big[|\Delta A_{T_n}|\big] < \infty.$$

(iii) ⇒ (iv). Suppose $X = X_0 + M + A$ where A is locally integrable. By Lemma 8.3.3, there is a unique predictable process $\widetilde{A} = \Pi_p^* A$ such that $A - \widetilde{A}$ is a local martingale. Writing

$$X = X_0 + (M + A - \widetilde{A}) + \widetilde{A}\,,$$

the term in parentheses is a local martingale and $\widetilde{A} \in \mathcal{A}_{\mathrm{loc}}$ is predictable.

To show uniqueness, suppose

$$X = X_0 + M + A = X_0 + \overline{M} + \overline{A}\,,$$

where M and $\overline{M}$ are in $\mathcal{M}_{0,\mathrm{loc}}$ and A and $\overline{A}$ are both predictable. Then $\overline{M} - M = A - \overline{A}$ is a predictable local martingale which is locally of integrable variation. Therefore, after localizing, we can apply Corollary 8.2.14 to see it is the zero process. Therefore, the canonical decomposition of a special semimartingale is unique.

(i) ⇒ (v). If X is a special semimartingale, there is a decomposition $X = X_0 + M + A$ in which $A \in \mathcal{A}_{\mathrm{loc}}$. As M is a local martingale, from Theorem 10.3.3, we have $M = U + V$, where $U \in \mathcal{H}^2_{\mathrm{loc}}$ and V is locally of integrable variation. Using this decomposition, we have

$$X_t^* \le |X_0| + U_t^* + \int_{[0,t]} \big(|dV|_s + |dA|_s\big).$$

Localizing and applying Doob's maximal inequality (Theorem 5.1.3) to U, we see that each of the terms on the right-hand side is locally integrable.

(v) $\Rightarrow$ *(vi)*. This follows because $|\Delta X_t| = |X_t - X_{t-}| \le 2X_t^*$.
(vi) $\Rightarrow$ *(i)*. Suppose $X = X_0 + M + A$ is a decomposition of X. Now

$$|\Delta A_t| \le |\Delta M_t| + |\Delta X_t| \le 2M_t^* + D_t\,,$$

and the right-hand side is increasing and locally integrable. Consider an increasing sequence of stopping times $\{R_n\}_{n\in\mathbb{N}}$, such that $\lim_n R_n = \infty$ and

$$E\big[\sup_{s\le R_n} |\Delta A_s|\big] \le E[2M_{R_n}^* + D_{R_n}] < \infty$$

for all n. Write

$$S_n = \inf\Big\{t : \int_{[0,t]} |dA_s| \ge n\Big\} \quad \text{and} \quad T_n = S_n \wedge R_n\,.$$

Then

$$\begin{aligned} E\Big[\int_{[0,T_n]} |dA_s|\Big] &= E\Big[\int_{[0,T_n[} |dA_s| + |\Delta A_{T_n}|\Big] \\ &\le n + E\big[\sup_{s\le R_n} |\Delta A_s|\big] < \infty. \end{aligned}$$

Therefore, $A \in \mathcal{A}_{\mathrm{loc}}$ and the result is proven. □

Corollary 11.6.12. *Every local special semimartingale is a special semimartingale.*

Proof. If X is a local special semimartingale then, from Lemma 11.6.3, X is a semimartingale. Using property (v) of Theorem 11.6.10, as X is locally a special semimartingale, we know that there is a localizing sequence $\{T_n\}_{n\in\mathbb{N}}$ such that $(X^{T_n})_t^*$ is locally integrable for each n, that is, there exists a doubly indexed sequence of stopping times $\{S_{n,m}\}_{n,m\in\mathbb{N}}$ such that $(X^{S_{n,m}})_t^*$ is integrable for each n, m, and $\lim_{m\to\infty} S_{n,m} = T_n$ for each n. Taking $R_n = \max_{m\le n} S_{n,m}$, we see that $R_n \to \infty$ and $(X^{R_n})_t^*$ is integrable for each n, and so X is a special semimartingale. □

11.7 Exercises

Exercise 11.7.1. Show that, for all t, the map $M, N \mapsto [M,N]_t(\omega)$ on $\mathcal{H}^2$ behaves like an inner product, that is, it is symmetric, bilinear and nonnegative definite, except on some null set in ω which may depend on M and N. What does it mean if $[M,M]_t(\omega) = 0$ a.s. for some t, for some non-null set of ω?

Exercise 11.7.2. Show that, for X a local martingale, $[X]$ is the unique nondecreasing optional process such that $[X]_0 = 0$, $X^2 - [X]$ is a local martingale and $\Delta[X] = (\Delta X)^2$.

Exercise 11.7.3. Give an example of a martingale which is in $\mathcal{H}^1_{\mathrm{loc}}$ but not $\mathcal{H}^2_{\mathrm{loc}}$.

Exercise 11.7.4. For M a martingale of finite variation, show that the set

$$\Big\{Z \bullet M : Z \text{ predictable}, E\Big[\int Z_s^2\, d\langle M\rangle_s\Big] < \infty\Big\}$$

is stable, in the sense of Definition 10.1.18.

Exercise 11.7.5. Consider the setting of Exercise 5.7.11. Find the quadratic variations (predictable and optional) of the time changed martingale in terms of the original martingale, assuming $C(t)$ is a predictable stopping time for each t. Use this to construct a continuous martingale with quadratic variation $\langle M\rangle_t = t^2$, and a pure jump martingale with quadratic variation $\langle M\rangle_t = \sqrt{t}$.

Exercise 11.7.6. For M a $\mathcal{H}^2_{\mathrm{loc}}$ martingale, show that the maps

$$\begin{aligned}\mu_p &: A \mapsto E\Big[\int I_A\, d\langle M\rangle\Big],\\ \mu_o &: A \mapsto E\Big[\int I_A\, d[M]\Big]\end{aligned}$$

are measures on the progressive σ-algebra Σ_π, and that their restrictions to the predictable σ-algebra Σ_p agree.

Exercise 11.7.7. Give an example of a set $A \in \Sigma_\pi$ and a $\mathcal{H}^2_{\mathrm{loc}}$ martingale M such that there is no measure ν on $\mathcal{B}$ with

$$E\Big[\int I_A d[M]\Big] = \int E[I_A] d\nu.$$

Exercise 11.7.8. Let M be a finite variation martingale with totally inaccessible jump times. Show that, for H a predictable process, $H \bullet M = 0$ (up to indistinguishability) if and only if $H = 0$ $d\langle M\rangle \times dP$-a.s.

Exercise 11.7.9. Let $X_t(\omega) = f(t)$ for some deterministic function f. Show that X is a semimartingale if and only if f is càdlàg and of finite variation on $[0, t]$ for all t, and X is a martingale if and only if f is constant.

Exercise 11.7.10. Consider a càdlàg process X which is constant except at the integers, $X_0 = 1/2$, and at each integer n takes the values $(n+1)^{-1}$ and $1-(n+1)^{-1}$. Create a measure such that X is a martingale, and give a formula for its quadratic variation. Hence show that not all bounded martingales are in $\mathcal{H}^\infty$, and the BDG inequality does not extend to the case $p = \infty$.

Exercise 11.7.11. Show that, if $X \in \mathcal{H}^2_{\mathrm{loc}}$, then X is a.s. constant on any stochastic interval $[\![S, T]\!]$ with $\langle X\rangle_S = \langle X\rangle_T$ a.s.

Exercise 11.7.12. Let X be a purely discontinuous local martingale. Suppose M is a purely discontinuous local martingale such that $\Delta M = \Delta X$ up to indistinguishability. Show that X and M are indistinguishable.

12

The Stochastic Integral

We have now established enough basic theory to construct the stochastic integral in full generality. In this chapter, we develop the integral with respect to semimartingales, and prove some of its properties. As in the previous chapters, we assume we have a filtered probability space, with filtration satisfying the usual conditions, $\mathcal{F}_\infty = \mathcal{F}_{\infty-}$, a martingale will always refer to its càdlàg version, and statements should be read as 'up to an evanescent set'.

12.1 The Itô Isometry

We shall first construct the stochastic integrals with respect to a general square integrable martingale. As with all classical integration theory, this is done by first defining the integral with respect to simple integrands, and then approximating general integrands using simple ones. To show that these integrals converge, we follow the classical construction of Itô in defining an isometry between the spaces of predictable integrands and the stochastic integrals.

Recall that, in Section 7.2, we defined the class of predictable processes. In particular, it was shown that the predictable σ-algebra (Σ_p) was generated by the left-continuous simple functions (Theorem 7.2.4). Therefore, if we define the integral for bounded left-continuous simple functions, we can expect to extend our integral to all predictable functions (which includes, for example, all left-continuous functions) with sufficient integrability.

Definition 12.1.1. *Write Λ for the space of bounded, left-continuous, simple predictable processes. That is, $H \in \Lambda \subset B(\Sigma_p, \mathbb{R})$ if there is a finite sequence of stopping times $0 = t_0 < t_1 < t_2 < \cdots < t_n < t_{n+1} = \infty$, and a family $\{H^i\}_{i\in\{0,1,2,\ldots,n\}}$ of bounded random variables such that H^i is $\mathcal{F}_{t_i}$-measurable for each i, and*

$$H_0 = H^0 \text{ and } H_t = H^i \text{ for } t \in]t_i, t_{i+1}].$$

S.N. Cohen, R.J. Elliott, *Stochastic Calculus and Applications*,
Probability and Its Applications, DOI 10.1007/978-1-4939-2867-5_12

Remark 12.1.2. Note that Λ generates the predictable σ-algebra, that is, $\Sigma_p = \sigma(\Lambda)$. Therefore, for any measure μ on Σ_p and any $q \in [1, \infty[$, by dominated convergence we know that Λ is a dense set of functions in $L^q(\Sigma_p, \mu)$.

Definition 12.1.3. *Suppose $M \in \mathcal{H}^2$. Then for all $H \in \Lambda$, considering each path, we define the stochastic integral to be the process*

$$\begin{aligned}(H \bullet M)_t = \int_{[0,t]} H_s dM_s &:= H^0 M_0 + \sum_i H^i (M_{t_{i+1}\wedge t} - M_{t_i \wedge t}).\\ &= H_0 M_0 + \sum_i H_{t_i} (M_{t_{i+1}\wedge t} - M_{t_i \wedge t}).\end{aligned}$$

The following fundamental lemma is why we restrict our attention to *left-continuous* simple processes.

Lemma 12.1.4 (Itô's Isometry). *For any $H \in \Lambda$ and $M \in \mathcal{H}_0^2$, we know $H \bullet M \in \mathcal{H}^2$, that is, $H \bullet M$ is a square integrable martingale, and we have the isometry*

$$E[(H \bullet M)_\infty^2] = E\Big[\int_{[0,\infty]} H_s^2 d\langle M\rangle_s\Big] = E\Big[\int_{[0,\infty]} H_s^2 d[M]_s\Big].$$

Proof. From the above definition, as $M_0 = 0$,

$$(H \bullet M)_t = \sum_{i=0}^n H_{t_i} (M_{t_{i+1}\wedge t} - M_{t_i \wedge t}).$$

To see that $H \bullet M$ is a martingale, note that for $s \leq t$,

$$\begin{aligned}E\Big[\sum_{i=0}^n H_{t_i} (M_{t_{i+1}\wedge t} - M_{t_i \wedge t})\Big|\mathcal{F}_s\Big] &= \sum_{i=0}^n E\Big[H_{t_i} E[M_{t_{i+1}\wedge t} - M_{t_i \wedge t}|\mathcal{F}_{s \vee t_i}]\Big|\mathcal{F}_s\Big]\\ &= \sum_{i=0}^n E\Big[H_{t_i} (M_{t_{i+1}\wedge s} - M_{t_i \wedge s})\Big|\mathcal{F}_s\Big]\\ &= \sum_{i=0}^n H_{t_i} (M_{t_{i+1}\wedge s} - M_{t_i \wedge s})\end{aligned}$$

and so $E[(H \bullet M)_t|\mathcal{F}_s] = (H \bullet M)_s$. We know that, for $i < j$ (so that $i+1 \leq j$),

$$\begin{aligned}&E[H_{t_i} H_{t_j} (M_{t_{i+1}\wedge t} - M_{t_i \wedge t})(M_{t_{j+1}\wedge t} - M_{t_j \wedge t})]\\ &= E\big[E[H_{t_i} H_{t_j} (M_{t_{i+1}\wedge t} - M_{t_i \wedge t})(M_{t_{j+1}\wedge t} - M_{t_j \wedge t})|\mathcal{F}_{t_j}]\big]\\ &= E\big[H_{t_i} H_{t_j} (M_{t_{i+1}\wedge t} - M_{t_i \wedge t}) E[(M_{t_{j+1}\wedge t} - M_{t_j \wedge t})|\mathcal{F}_{t_j}]\big] = 0\end{aligned}$$

and

$$\begin{aligned}
&E\big[(M_{t_{i+1}\wedge t} - M_{t_i\wedge t})^2\big|\mathcal{F}_{t_i\wedge t}\big] \\
&= E\big[M^2_{t_{i+1}\wedge t} + M^2_{t_i\wedge t} - 2E[M_{t_{i+1}\wedge t}|\mathcal{F}_{t_i}]M_{t_i\wedge t}\big|\mathcal{F}_{t_i\wedge t}\big] \\
&= E\big[M^2_{t_{i+1}\wedge t} - M^2_{t_i\wedge t}\big|\mathcal{F}_{t_i\wedge t}\big] \\
&= E\big[\langle M\rangle_{t_{i+1}\wedge t} - \langle M\rangle_{t_i\wedge t}\big|\mathcal{F}_{t_i\wedge t}\big].
\end{aligned}$$

As H is bounded, $M \in \mathcal{H}^2$ and all cross products vanish, we can therefore see that

$$\begin{aligned}
E[(H\bullet M)_t^2] &= E\Big[\sum_{i=0}^n H_t^2(M_{t_{i+1}\wedge t} - M_{t_i\wedge t})^2\Big] \\
&= E\Big[\sum_{i=0}^n H_{t_i}^2(\langle M\rangle_{t_{i+1}\wedge t} - \langle M\rangle_{t_i\wedge t})\Big] \\
&= E\Big[\int_{[0,t]} H_s^2 d\langle M\rangle_s\Big].
\end{aligned}$$

It follows that $H\bullet M \in \mathcal{H}^2$. Letting $t \to \infty$, by the monotone convergence theorem and the convergence of $\mathcal{H}^2$ martingales (as in Lemma 10.1.4), we have

$$E[(H\bullet M)_\infty^2] = E\Big[\int_{[0,\infty]} H_s^2 d\langle M\rangle_s\Big] < \infty.$$

The final equality of the lemma follows because $\langle M\rangle - [M]$ is a martingale of integrable variation (see Lemma 11.1.9), so $E\big[\int_{[0,t]} H_s d\langle M\rangle_s - \int_{[0,t]} H_s d[M]_s\big] = 0$ for any $H \in \Lambda$ and any $t < \infty$, and, after rearranging, we can take $t \to \infty$ by monotone convergence. □

Remark 12.1.5. As $H\bullet M \in \mathcal{H}^2$ for any $H \in \Lambda$ and $M \in \mathcal{H}^2$, simple calculations yield the general isometry

$$E[(H\bullet M)^2] = E\Big[(H_0M_0)^2 + \int_{]0,\infty]} H_s^2 d\langle M\rangle_s\Big].$$

Definition 12.1.6. *Write $L^2(M)$ for the space of predictable process H such that*

$$\|H\|^2_{\langle M\rangle} := E\Big[(H_0M_0)^2 + \int_{[0,\infty]} H_s^2 d\langle M\rangle_s\Big] < \infty.$$

Remark 12.1.7. For $M \in \mathcal{H}_0^2$, the space $L^2(M)$ is actually the L^2 space under the measure given by $d\langle M\rangle \times dP$ (Remark 11.3.7). By Lemma 11.3.4, we could equivalently define the space in terms of the optional variation, that is, with the norm $H \mapsto E\Big[(H_0M_0)^2 + \int_{[0,\infty]} H_s^2 d[M]_s\Big] < \infty$.

The result below follows Itô's construction of a stochastic integral. This is done by extending the integral defined for simple processes in Definition 12.1.3 to integrals of processes in $L^2(M)$ in such a way that the Itô isometry (Lemma 12.1.4) is preserved. Note that here we will take $\mathcal{H}^2$ to have the norm $M \mapsto E[M_\infty^2]^{1/2}$, as this is the norm which appears in the Itô isometry. By Lemma 10.1.3, this is an equivalent norm to that used to define $\mathcal{H}^2$.

Theorem 12.1.8. *The linear map $H \mapsto H\bullet M$ of Λ into $\mathcal{H}^2$ extends in a unique manner to an isometric linear map of $L^2(M) \mapsto \mathcal{H}^2$. This map is again denoted by*

$$H \mapsto H\bullet M = \Big\{ \int_{[0,t]} H_s dM_s \Big\}_{t\geq 0},$$

and is called the integral of H with respect to M.

Proof. We know that $L^2(M)$ is the L^2 space, with respect to the measure $|M_0|^2\delta_{t=0} + d\langle M\rangle \times dP$ (or equivalently, $|M_0|^2\delta_{t=0} + d[M] \times dP$), of the predictable σ-algebra. Therefore, the set of bounded simple functions Λ is dense in the space of square-integrable functions $L^2(M)$ (by dominated convergence). The map $H \to H\bullet M$ of Λ into $\mathcal{H}^2$ is an isometry by Lemma 12.1.4 above, if Λ is given the norm $\|\cdot\|_{\langle M\rangle}$.

Therefore, for any $H \in L^2(M)$, we can find a sequence $\{H^n\}_{n\in\mathbb{N}} \subset \Lambda$ such that $\|H - H^n\|_{\langle M\rangle} \to 0$, and hence the sequence $\{H^n\bullet M\}_{n\in\mathbb{N}}$ converges in $\mathcal{H}^2$. We define

$$H\bullet M = \lim_{n\to\infty} (H^n\bullet M).$$

If we have two approximating sequences $\{H^n\}_{n\in\mathbb{N}}$ and $\{\tilde{H}^n\}_{n\in\mathbb{N}}$, then $\|H^n - \tilde{H}^n\|_{\langle M\rangle} \to 0$, and so, by linearity, $H^n\bullet M - \tilde{H}^n\bullet M \to 0$ in $\mathcal{H}^2$. Therefore, the limit is independent of the choice of approximating sequence. □

Remark 12.1.9. From the definition, for any $H \in L^2(M)$ and $M \in \mathcal{H}^2$, it is clear that $(H\bullet M)$ is a càdlàg square-integrable martingale, that the Itô isometry

$$E[(H\bullet M)_\infty^2] = E[(H^2\bullet\langle M\rangle)_\infty] = \|H\|_{\langle M\rangle}$$

is satisfied, and that $H \mapsto (H\bullet M)$ is a linear map[1], that is

$$(aH + bK)\bullet M = a(H\bullet M) + b(K\bullet M).$$

From the isometry, we also observe that, if $H^n \to H$ in $L^2(M)$, then the integrals $H^n\bullet M \to H\bullet M$ in $\mathcal{H}^2$. In particular, this implies a dominated convergence theorem: if $|H^n| \leq G$ for some $G \in L^2(M)$ and $H^n \to H$ pointwise, then $H^n\bullet M \to H\bullet M$ in $\mathcal{H}^2$.

[1] We should be careful, as linearity holds up to an evanescent set, but this set may depend on a, b, H and K.

Lemma 12.1.10. *For $H \in L^2(M)$ the processes $\Delta(H \bullet M)$ and $H \Delta M$ are indistinguishable.*

Proof. For a simple function $H \in \Lambda$, clearly $\Delta(H \bullet M) = H \Delta M$. For a general process $H \in L^2(M)$ suppose $\{H^n\}_{n \in \mathbb{N}} \subset \Lambda$ is a sequence that approximates H in the $\| \cdot \|_{\langle M \rangle}$ seminorm. Then

$$E[(H^n \bullet M)^2_\infty] \to E[(H \bullet M)^2_\infty],$$

so by Lemma 10.1.3, when we consider the maximal process, we have

$$(H^n \bullet M)^*_\infty \to (H \bullet M)^*_\infty \quad \text{in} \quad L^2(\Omega).$$

From Theorem 10.1.10 there is, therefore, a subsequence $\{n_k\}_{k \in \mathbb{N}}$ such that $(H^{n_k} \bullet M)_t$ converges to $(H \bullet M)_t$, uniformly in t, for almost every ω. In particular, for almost all ω,

$$H^{n_k}_t \Delta M_t = \Delta(H^{n_k} \bullet M)_t \to \Delta(H \bullet M)_t \quad \text{uniformly in } t.$$

From the definition of $[M]$, with the convention $\Delta M_0 = M_0$,

$$\begin{aligned}
&E\Big[\sum_{u \geq 0} ((H^{n_k}_u - H_u)\Delta M_u)^2\Big] \\
&\leq E\Big[(H^{n_k}_0 - H_0)^2 M_0^2 + \int_{[0,\infty[} (H^{n_k}_u - H_u)^2 d[M]_u\Big] \\
&= \|H^{n_k} - H\|^2_{\langle M \rangle} \to 0,
\end{aligned}$$

so we know that $H^{n_k} \Delta M \to H \Delta M$ for almost all ω, uniformly in t. Therefore, $H \Delta M = \Delta(H \bullet M)$ except on an evanescent set. □

Example 12.1.11. For the sake of concreteness, we now outline how this construction works when the martingale M is a Brownian motion. In this case we have $\langle M \rangle_t = t$ and $M_0 = 0$, so for any sequence of bounded predictable simple functions $\{H^n\}_{n \in \mathbb{N}} \subset \Lambda$ such that, for some process H,

$$E\Big[\int_{[0,\infty[} (H^n_t - H_t)^2 dt\Big] \to 0,$$

we have

$$\int_{[0,t]} H^n_s dW_s \to \int_{[0,t]} H_s dW_s,$$

the convergence taking place in $\mathcal{H}^2$, that is,

$$E\Big[\sup_t \Big(\int_{[0,t]} H^n_s dW_s - \int_{[0,t]} H_s dW_s\Big)^2\Big] \to 0.$$

The process $H \bullet W$ is a continuous $\mathcal{H}^2$-martingale with quadratic variation

$$[H \bullet W]_t = \langle H \bullet W \rangle_t = \int_{[0,t]} H^2_s ds.$$

Remark 12.1.12. In the same way as for finite variation processes, for $s < t$, whenever M can jump there is a difference between the integrals $\int_{[s,t]} H_u dM_u$, $\int_{]s,t]} H_u dM_u$, $\int_{[s,t[} H_u dM_u$ and $\int_{]s,t[} H_u dM_u$. Many authors write $\int_s^t H_u dM_u$ for $\int_{]s,t]} H_u dM_u$, however we will avoid this notation to prevent confusion. Of course, when M is continuous, the integrals all agree, and the notation $\int_s^t$ is unproblematic.

Remark 12.1.13. When M is a *continuous* martingale, then $\langle M\rangle = [M]$, and we can see that the Itô isometry also holds for right-continuous adapted processes. As the simple right-continuous processes generate the optional σ-algebra, they are dense in the corresponding L^2 space (by dominated convergence), and it follows that we can consistently define the integral for any square-integral *optional* integrand (using the Itô isometry). In fact, one can also show that the simple processes are dense in $L^2(\Sigma_\pi, d\langle M\rangle \times dP)$, where Σ_π is the *progressive* σ-algebra, so the integral (with respect to a continuous martingale) can be well defined for any square-integrable progressive integrand. See Karatzas and Shreve [117, p.132ff.] for details.

12.2 Orthogonality and Integration

The above construction of Itô defines the stochastic integral in terms of an isometry from $L^2(M)$ into $\mathcal{H}^2$. The following result, due to Kunita and Wantabe [122], characterizes the stochastic integral in terms of the covariations of the integral process $H \bullet M$ with other martingales $N \in \mathcal{H}^2$.

Theorem 12.2.1. *Suppose $M \in \mathcal{H}^2_0$ and $H \in L^2(M)$.*

(i) For every $N \in \mathcal{H}^2$,

$$E\Big[\int_{[0,\infty]} |H_s|\,|d\langle M,N\rangle_s|\Big] < \infty,$$
$$E\Big[\int_{[0,\infty]} |H_s|\,|d[M,N]_s|\Big] < \infty.$$

(ii) The stochastic integral $L = H \bullet M$ is the unique element of $\mathcal{H}^2_0$ such that, for every $N \in \mathcal{H}^2$,

$$E[L_\infty N_\infty] = E\Big[\int_{[0,\infty]} H_s d\langle M,N\rangle_s\Big] = E\Big[\int_{[0,\infty]} H_s d[M,N]_s\Big].$$

(iii) For every $N \in \mathcal{H}^2$,

$$\langle L,N\rangle = H \bullet \langle M,N\rangle,$$
$$[L,M] = H \bullet [M,N],$$

the integrals on the right being Stieltjes integrals.

Proof. The inequalities of (i) are immediate consequences of the Kunita–Watanabe inequalities (Corollary 11.4.2).

Consider the linear functional on $L^2(M)$ defined by

$$H \mapsto E\Big[(H\bullet M)_\infty N_\infty - \int_{[0,\infty]} H_s d\langle M, N\rangle_s\Big].$$

This is continuous, again by the Kunita–Watanabe inequalities. Furthermore, it is easy to check that it is zero on the simple functions Λ, and so it is zero on all of $L^2(M)$ by continuity. The second identity of part (ii) follows because $\langle M, N\rangle - [M, N]$ is a martingale of integrable variation, so the identity holds for $H \in \Lambda$, and so for all $H \in L^2(M)$ by continuity.

To prove the final identities, note that the process J defined by

$$J_t = L_t N_t - \int_{[0,t]} H_s d\langle M, N\rangle_s$$

is bounded above by $L^*_\infty N^*_\infty + \int_{[0,\infty]} |H_s|\,|d\langle M, N\rangle_s|$ which is in L^1. Recall that $\langle M, N^T\rangle = \langle M, N\rangle^T$ for any stopping time T, then apply the identity of part (ii) to N^T to see that $E[J_T] = 0$. By Theorem 5.4.6, J is a uniformly integrable martingale. However, $\langle L, N\rangle$ is defined as the unique predictable process of integrable variation such that $LN - \langle L, N\rangle$ is a martingale. Therefore,

$$\langle L, N\rangle = H\bullet\langle M, N\rangle. \tag{12.1}$$

For the second identity in (iii), decompose M and N into the sum of their continuous and totally discontinuous parts,

$$M = M^c + M^d, \quad N = N^c + N^d.$$

Then, for any $K \in \mathcal{H}^{2,c}$, using (12.1) we have

$$\langle H\bullet M^d, K\rangle = H\bullet\langle M^d, K\rangle = 0.$$

It follows that $H\bullet M^d \in \mathcal{H}^{2,d}$. As $H\bullet M^c$ is certainly a continuous martingale (from Lemma 12.1.10), the decomposition of $H\bullet M$ into continuous and purely discontinuous martingales is given by

$$(H\bullet M)^c = H\bullet M^c \quad \text{and} \quad (H\bullet M)^d = H\bullet M^d.$$

Therefore by Remark 11.2.4, Lemma 12.1.10 and (12.1), we have

$$\begin{aligned}
[L, N]_t &= \langle L^c, N^c\rangle_t + \sum_{0<s\le t} \Delta L_s \Delta N_s \\
&= (H\bullet\langle M^c, N^c\rangle)_t + \sum_{0<s\le t} H_s \Delta M_s \Delta N_s \\
&= \int_{[0,t]} H_s d[M, N]_s\,.
\end{aligned}$$

□

Remark 12.2.2. The first identity in Theorem 12.2.1(ii) uniquely characterizes the stochastic integral $L = H \bullet M$. This is because the right-hand side is a continuous linear functional in N (given H and M), whilst the left-hand side is just the inner product of L and N in the Hilbert space $\mathcal{H}^2$. Consequently, by the Riesz–Fréchet representation theorem in Hilbert spaces (Theorem 1.5.22), given H and M there is a unique element $L = H \bullet M \in \mathcal{H}^2$ which gives this linear functional.

Let us explicitly note the following result obtained in the course of the above proof.

Lemma 12.2.3. *For $M \in \mathcal{H}^2$ and $H \in L^2(M)$, the continuous and purely discontinuous parts of the martingale $H \bullet M$ are $H \bullet M^c$ and $H \bullet M^d$, respectively.*

Corollary 12.2.4. *For $M \in \mathcal{H}^2$ and $H \in L^2(M)$, we have $\langle H \bullet M \rangle = H^2 \bullet \langle M \rangle$.*

Proof. By symmetry of $\langle \cdot, \cdot \rangle$ and Theorem 12.2.1(iii), we have

$$\langle H \bullet M, H \bullet M \rangle = H \bullet \langle M, H \bullet M \rangle = H^2 \bullet \langle M, M \rangle. \qquad \square$$

Corollary 12.2.5. *For $M \in \mathcal{H}^2$, $H \in L^2(M)$ and K a bounded predictable process,*

$$(KH) \bullet M = K \bullet (H \bullet M).$$

More generally, this holds whenever K is such that $KH \in L^2(M)$.

Proof. Suppose $M \in \mathcal{H}_0^2$. For any $N \in \mathcal{H}^2$,

$$\begin{aligned} E\big[\big((KH) \bullet M\big)_\infty N_\infty\big] &= E[(KH) \bullet \langle M, N \rangle_\infty] \\ &= E[K \bullet \langle H \bullet M, N \rangle_\infty] \\ &= E\big[\big(K \bullet (H \bullet M)\big)_\infty N_\infty\big], \end{aligned}$$

and the result follows by Theorem 12.2.1(ii). The general case $M \in \mathcal{H}^2$ follows by addition of the initial value $K_0 H_0 M_0$. $\square$

The following 'stability' results have been seen in the context of finite variation processes in Exercise 11.7.4.

Theorem 12.2.6. *Suppose $\mathcal{Y} \subseteq \mathcal{H}^2$ is a stable subspace (see Definition 10.1.18). Then for $M \in \mathcal{Y}$ and $H \in L^2(M)$ we know $H \bullet M \in \mathcal{Y}$.*

Conversely, if $\mathcal{Y}$ is closed in $\mathcal{H}^2$ and for each $M \in \mathcal{Y}$ and $H \in L^2(M)$ we have $H \bullet M \in \mathcal{Y}$, then $\mathcal{Y}$ is stable.

Proof. To prove the direct statement of the theorem, it is enough to show that $H \bullet M \in \mathcal{Y}^{\perp\perp}$. Suppose $N \in \mathcal{Y}^\perp$. Then $M_0 N_0 = 0$ and $\langle M, N \rangle = 0$. Hence $(H \bullet M)_0 N_0 = H_0 M_0 N_0 = 0$ and $\langle H \bullet M, N \rangle = H \bullet \langle M, N \rangle = 0$, by Theorem 12.2.1. Therefore $(H \bullet M)N$ is a local martingale, and so $H \bullet M \in \mathcal{Y}^{\perp\perp}$.

The converse is proven by observing that if $H = I_{[\![0,T]\!]}$ then $H \in L^2(M)$ and $H \bullet M = M^T$. Also, if $A \in \mathcal{F}_0$, then $H = I_{[0,\infty[\times A} \in L^2(M)$ and $H \bullet M = I_A M$. $\square$

Theorem 12.2.7. *Consider $M \in \mathcal{H}^2$. The stable subspace generated by M (that is, the smallest stable subspace containing M) is the set of stochastic integrals $\{H \bullet M\}_{H \in L^2(M)}$.*

For $N \in \mathcal{H}^2_0$ the projection of N on this subspace is $D \bullet M$, where D is a predictable density (or Radon–Nikodym derivative) of $\langle M, N\rangle$ with respect to $\langle M\rangle$, considered as measures on Σ_p. In other words, the process $\tilde{N}$ defined by

$$\tilde{N}_t = N_t - \int_{[0,t]} \Big(\frac{d\langle M, N\rangle}{d\langle M\rangle}(\omega, s)\Big) dM_s$$

is a martingale orthogonal to M.

Proof. With M fixed, the map $H \mapsto H \bullet M$ of $L^2(M)$ into $\mathcal{H}^2$ is an isometry (where $L^2(M)$ has norm $\|\cdot\|_{\langle M\rangle}$). Write $\mathcal{K}$ for its image. Then $\mathcal{K}$ is closed, stable and contained in every stable subspace containing M, by Theorem 12.2.6. Therefore $\mathcal{K}$ is the stable subspace generated by M.

For $N \in \mathcal{H}^2_0$, write $N = N_1 + N_2$, where N_1 is the projection of N on $\mathcal{K}$ and $N_2 \in \mathcal{K}^\perp$. Then $\langle M, N\rangle = \langle M, N_1 + N_2\rangle = \langle M, N_1\rangle$, because $\langle M, N_2\rangle = 0$. Because $N_1 \in \mathcal{K}$ it can be written as $N_1 = D \bullet M$, for some $D \in L^2(M)$. Therefore $\langle M, N_1\rangle = \langle M, D \bullet M\rangle = D \bullet \langle M\rangle$, and so D is a predictable density of $\langle M, N\rangle$ with respect to $\langle M\rangle$, that is,

$$D_t(\omega) = \frac{d\langle M, N\rangle}{d\langle M\rangle}(\omega, t),$$

as measures on the space $(\Omega \times [0, \infty], \Sigma_p)$. □

The following result relates stochastic and Stieltjes integrals, whenever both are defined, ensuring that our notation is consistent.

Theorem 12.2.8. *Suppose $M \in \mathcal{H}^2$ is a process of integrable variation. Furthermore, suppose $H \in L^2(M)$ is such that $E\big[\int_{[0,\infty]} |H_s|\,|dM_s|\big] < \infty$. Then the stochastic integral $H \bullet M$ and the pathwise Stieltjes integral $\{\int_{[0,t]} H_u(\omega) dM_u(\omega)\}_{t \geq 0}$ (see Definition 8.1.1) are indistinguishable processes.*

Proof. As M is of integrable variation, by Theorem 10.2.6 we know that M is of the form

$$M = M_0 + A - \Pi_p^* A,$$

where $A_t = \sum_{0<s\leq t} \Delta M_s$ and $\Pi_p^* A$ is continuous, in particular, $M \in \mathcal{H}^{2,d}$. By Theorem 12.2.6, we see that $H \bullet M \in \mathcal{H}^{2,d}$. As H is predictable, using Theorem 8.2.11(iii) (see Exercise 8.4.2) we can check that the Stieltjes integral $\{\int_{[0,t]} H_u dM_u\}_{t\geq 0}$ is a martingale, and it is straightforward to verify that it is orthogonal to any continuous martingale. Hence $\{\int_{[0,t]} H_u dM_u\}_{t\geq 0} \in \mathcal{H}^{2,d}$.

By Lemma 12.1.10,

$$\Delta\Big(\int_{[0,\cdot]} H_u dM_u\Big)_t = H_t \Delta M_t = \Delta(H\bullet M)_t,$$

and so, by Theorem 10.2.22, as the jumps are indistinguishable, the two processes are indistinguishable. □

12.3 Local Martingales and Semimartingales

12.3.1 The Local Martingale Integral

So far in this chapter the stochastic integrals have been defined with respect to square integrable martingales in $\mathcal{H}^2$. We now wish to extend the definition to allow integration with respect to local martingales and semimartingales. This allows a far larger class of integrals to be considered.

We shall use the Burkholder–Davis–Gundy inequality. Note that, from dominated convergence, for $p \geq 1$, the processes in Λ are also dense with respect to the seminorm given by $H \mapsto E[(H^2 \bullet [M])^{p/2}]^{1/p}$.

Definition 12.3.1. *For M a local martingale and H a predictable process, a stochastic integral $X = H\bullet M$ is a local martingale with the following properties.*

(i) $X_0 = H_0 M_0$,
(ii) X has continuous martingale part $X^c = H\bullet M^c$, defined locally by Theorem 12.1.8,
(iii) the processes $\Delta(H\bullet M)$ and $H\Delta M$ are indistinguishable.

We shall allow a greater range of integrands in Definition 12.3.10, by loosening the requirement that X is also a local martingale. To show the existence and uniqueness of this integral, for a large class of processes H, we first prove the following result.

Theorem 12.3.2. *For $p \in [1, \infty[$, suppose M is a local martingale, $M_0 = 0$ and $\{T_n\}_{n\in\mathbb{N}}$ is a sequence of stopping times with $T_n \to \infty$ and $\{H^n\}_{n\in\mathbb{N}} \subset \Lambda$ is a sequence of processes such that for all $m \in \mathbb{N}$,*

$$E\Big[\Big(\int_{[0,T_m]} (H^n_s - H^{n'}_s)^2 d[M]_s\Big)^{p/2}\Big]^{1/p} \to 0 \text{ as } n, n' \to \infty$$

Then the integrals $H^n\bullet M$ converge locally in $\mathcal{H}^p$.

Proof. From the Burkholder–Davis–Gundy inequality (Theorem 11.5.5) we see that $M \mapsto E\big[[M]_\infty^{p/2}\big]^{1/p}$ and $M \mapsto E\big[|M^*_\infty|^p\big]^{1/p}$ are equivalent norms on $\mathcal{H}^p_0$. Localizing to $[0, T_m]$, we know $H^n\bullet M$ is a Cauchy sequence, and the result holds by completeness of $\mathcal{H}^p_0$. □

This allows us to generalize the class of stochastic integrals.

Theorem 12.3.3. *For $p \in [1,\infty[$, let $L^p(M)$ denote the space of predictable processes H with*

$$E\Big[\Big(H_0^2 M_0^2 + \int_{[0,\infty]} H_s^2 d[M]_s\Big)^{p/2}\Big] < \infty,$$

and $L^p_{\mathrm{loc}}(M)$ the space of processes locally in $L^p(M)$. (That is, $L^1_{\mathrm{loc}}(M)$ is the space of processes H where $(H^2 \bullet [M])^{1/2}$ is an increasing process of locally integrable variation.)

For any local martingale M and any $H \in L^1_{\mathrm{loc}}(M)$ satisfying the integrability assumption, the stochastic integral $H \bullet M$ is a uniquely defined local martingale (up to indistinguishability).

Remark 12.3.4. Note that, by Lemma 11.4.6, any locally bounded predictable process H is in $L^1_{\mathrm{loc}}(M)$ for all M.

Proof. We begin by constructing the integral. First localize, so that we can assume $(H^2 \bullet [M])^{1/2} \in \mathcal{A}^+$ and $[M]^{1/2} \in \mathcal{A}^+$ (see Lemma 11.4.6). Then, apply Theorem 12.3.2 (with $p = 1$) to an approximating sequence $\{H^n\}_{n\in\mathbb{N}} \subset \Lambda$ such that H^n is a Cauchy sequence in $L^1(M)$ with $H^n \to H$, that is, such that

$$E\Big[\Big((H_0^n - H_0)^2 M_0^2 + \int_{[0,\infty]} (H^n - H)_s^2 d[M]_s\Big)^{1/2}\Big] \to 0$$

as $n \to \infty$. By dominated convergence, such a sequence can be found using Corollary 7.4.3. Taking the limit in $\mathcal{H}^1$ of $H^n \bullet M$ and then pasting (to undo our localization), we have defined a local martingale X. We now need to show that X satisfies Definition 12.3.1.

That X is a local martingale is true by construction. Also, $(H^n \bullet M)_0 = H_0^n M_0 \to H_0 M_0$, so clearly $X_0 = H_0 M_0$, guaranteeing (i). To check (ii), first observe that, as M^c and M^d are orthogonal, we know that, locally,

$$E\Big[\Big(\int_{[0,\infty]} (H^n - H)_s^2 d[M^c]_s\Big)^{1/2}\Big] \le E\Big[\Big(\int_{[0,\infty]} (H^n - H)_s^2 d[M]_s\Big)^{1/2}\Big] \to 0$$

and similarly for M^d. Therefore $H \bullet M^c$ and $H \bullet M^d$ are both well defined. Hence the identity

$$H \bullet M = H \bullet M^c + H \bullet M^d$$

holds for all simple processes, and it follows that it holds for general H.

The process $H^2 \bullet [M^c] = H^2 \bullet \langle M^c \rangle$ is continuous, so locally integrable. Therefore, after localization, $H \bullet M^c$ can be defined by Theorem 12.1.8, and uniqueness implies the definitions coincide.

That $\Delta X = H \Delta M$ and $[X^d] = H^2 \bullet [M^d] = [H \bullet M^d]$ can be shown exactly as in Lemma 12.1.10, as convergence in $\mathcal{H}^1$ implies convergence uniformly in

t for almost all ω. This implies $[H \bullet M^d]$ increases only at its discontinuities, and so $H \bullet M^d$ is a pure jump martingale. It follows that $X^d - H \bullet (M^d)$ is a pure jump local martingale with $[X^d - H \bullet (M^d)] = 0$, and Lemma 11.4.5 then implies that X^d and $H \bullet M^d$ are indistinguishable. Consequently, $X^c = H \bullet M^c$ is the continuous part of X.

To see that this is the unique local martingale with these properties, we observe that (iii), Theorem 11.5.11 and Lemma 11.4.5 uniquely determine the value of the purely discontinuous local martingale X^d, and (ii) uniquely determines the value of X^c. The result follows. □

Corollary 12.3.5. *The following hold up to an evanescent set (which may depend on α, H, G, M and N).*

(i) The integral (in the sense of local martingales) is linear in the integrand, that is, for any local martingale M and any M-integrable H and G, any $\alpha \in \mathbb{R}$, we know $(\alpha H + G)$ is also M-integrable, and

$$(\alpha H + G) \bullet M = \alpha (H \bullet M) + G \bullet M.$$

(ii) The integral is linear in the integrator, that is, for any local martingales M and N and any H which is both M and N-integrable, any $\alpha \in \mathbb{R}$, we know H is also $(\alpha M + N)$-integrable, and

$$H \bullet (\alpha M + N) = \alpha (H \bullet M) + H \bullet N.$$

Proof. Clearly these results hold when the integrands are in Λ, as the integral is equal to a finite sum. By Theorem 12.3.2, a simple approximation argument (applied locally in $\mathcal{H}^1$) gives the result in the general case. □

An alternative characterization of the stochastic integral is as follows.

Corollary 12.3.6. *For any $M \in \mathcal{M}_{\mathrm{loc}}$ and $H \in L^1_{\mathrm{loc}}$, the integral $H \bullet M$ is the unique local martingale such that $(H \bullet M)_0 = H_0 M_0$ and, for any local martingale N,*

$$[H \bullet M, N] = H \bullet [M, N].$$

In particular, if $H \in L^p(M)$, then $H \bullet M \in \mathcal{H}^p$.

Proof. We know that $H \bullet M = H_0 M_0 + H \bullet M^c + H \bullet M^d$, and that this is its decomposition into continuous and purely discontinuous local martingale parts. By Theorem 12.2.1(iii), $H \bullet M^c$ is the unique local martingale such that $\langle H \bullet M^c, N\rangle = H \bullet \langle M^c, N\rangle$ for every $N \in \mathcal{H}^2$. By localization, as $\langle M^c, N\rangle$ is continuous, this implies the result holds with N a local martingale. We also know that $\Delta(H \bullet M)$ and $H \Delta M$ are indistinguishable, therefore

$$\begin{aligned}[H \bullet M^d, N]_t = [H \bullet M^d, N^c]_t + [H \bullet M^d, N^d]_t &= \sum_{s \leq t} H_s \Delta M_s \Delta N_s \\ &= H \bullet [M^d, N^d]_t.\end{aligned}$$

The result follows by linearity of $[\cdot,\cdot]$.

That $H \in L^p(M)$ implies $H\bullet M \in \mathcal{H}^p$ follows from the fact $H^2\bullet[M] = [H\bullet M]$ and the BDG inequality. □

Lemma 12.3.7. *If the local martingale M is also in $\mathcal{A}_{\text{loc}}$, and H is locally $|dM|\times\mathbb{P}$-integrable, then $H\bullet M$ can be calculated as the Stieltjes integral along each sample path, up to some evanescent set.*

Proof. As in Theorem 12.2.8, we first observe that M is a pure jump local martingale (Theorem 10.2.6) and the Stieltjes integral defines a local martingale (Exercise 8.4.2), which is orthogonal to any continuous martingale. Therefore the Stieltjes integral defines a process satisfying Definition 12.3.1, and so equals the stochastic integral. □

The following theorem allows us to say, for many examples, that the stochastic integral with respect to a martingale is a *martingale* (and not simply a local martingale).

Theorem 12.3.8. *Let $p \in [1,\infty]$ and $M \in \mathcal{H}^p_0$. Then there exists a constant C such that, for any bounded predictable process H with bound $|H| \leq k$,*

$$\|H\bullet M\|_{\mathcal{H}^p} = \|(H\bullet M)^*_\infty\|_p \leq Ck\|M^*_\infty\|_p = Ck\|M\|_{\mathcal{H}^p}$$

and hence the process $H\bullet M$ is in $\mathcal{H}^p$, in particular, it is a true martingale.

Proof. We know that $[H\bullet M] = H^2\bullet[M] \leq k^2[M]$. Applying the Burkholder–Davis–Gundy inequality (Theorem 11.5.5) we have

$$\|(H\bullet M)^*_\infty\|_p \leq C'\|[H\bullet M]^{1/2}_\infty\|_p \leq C'k\|[M]^{1/2}_\infty\|_p \leq Ck\|M^*_\infty\|_p$$

for some constants C', C. Therefore $H\bullet M$ is a local martingale bounded in $\mathcal{H}^p$ norm, and so is a martingale in $\mathcal{H}^p$ by Exercise 10.4.6. □

Remark 12.3.9. Even for pure jump martingales, the Itô construction can be more or less general than the pathwise integral. The pathwise integral assumes $M \in \mathcal{V}$ and $\sum_s |H_s \Delta M_s| < \infty$ for each ω, but needs no integrability with respect to ω. On the other hand, the Itô construction needs the process $\big\{(H^2\bullet[M])^{1/2}_t = \big(\sum_{s\leq t}(H_s\Delta M_s)^2\big)^{1/2}\big\}_{t\geq 0}$ to be locally integrable. As

$$\Big(\sum_s (H_s\Delta M_s)^2\Big)^{1/2} \leq \sum_s |H_s\Delta M_s|,$$

in cases where M may have many small jumps (so may not be in $\mathcal{V}$, and the sum on the right-hand side may not be finite), the stochastic construction allows for a more general integral.

12.3.2 The Semimartingale Integral

We can now state the natural extension of stochastic integration to semimartingales, by combining the pathwise and Itô approaches.

Definition 12.3.10. *Suppose X is a semimartingale. Let H be a predictable process such that there exists a decomposition $X = X_0 + M + A$ with*

- *$M \in \mathcal{M}_{0,\text{loc}}$ and $A \in \mathcal{V}_0$,*
- *$(H^2 \bullet [M])^{1/2} \in \mathcal{A}^+_{\text{loc}}$ (that is, $H \in L^1_{\text{loc}}(M)$) and*
- *for almost all ω, the path $H_{(\cdot)}(\omega)$ is locally $|dA|(\omega)$-integrable.*

*Such a process H is called X-*integrable *(and we write $H \in L(X)$), and we define*

$$H \bullet X = H_0 X_0 + H \bullet M + H \bullet A$$

and call this a stochastic integral *of H with respect to the semimartingale X. (Here the integral $H \bullet M$ is as in Theorem 12.3.3, while $H \bullet A$ is as in Theorem 8.1.3.)*

Remark 12.3.11. As $H \bullet M \in \mathcal{M}_{0,\text{loc}}$ and $H \bullet A \in \mathcal{V}_{0,\text{loc}} = \mathcal{V}_0$, we see that the integral $H \bullet X$ is also a semimartingale.

Remark 12.3.12. Clearly, if X is a local martingale, then $L^p_{\text{loc}}(X) \subseteq L(X)$. However, if X is discontinuous, we shall see that these spaces are generally not equal. In particular, if $H \in L^1_{\text{loc}}(X)$, then $H \bullet X$ is a local martingale (this follows from the following theorem and our previous results), however this may not be true for $H \in L(X)$.

Theorem 12.3.13. *For X, H as in Definition 12.3.10, the integral $H \bullet X$ is independent of the choice of decomposition of X (among those decompositions satisfying the integrability requirements), so we can call it* the *stochastic integral of H with respect to X.*

Proof. Suppose $X = X_0 + \overline{M} + \overline{A}$, is a second decomposition of X, with $\overline{M} \in \mathcal{M}_{0,\text{loc}}$, $\overline{A} \in \mathcal{V}_0$, $(H^2 \bullet [\overline{M}])^{1/2} \in \mathcal{A}^+_{\text{loc}}$ and $H_{(\cdot)}(\omega)$ almost surely $|d\overline{A}|(\omega)$ locally integrable. Then $M - \overline{M} = \overline{A} - A \in \mathcal{M}_{0,\text{loc}} \cap \mathcal{V}_0$, and so $M - \overline{M}$ is a local martingale which is locally of integrable variation.

As Λ is dense in all the metrics involved in the stochastic integral, there exists a sequence $\{H^n\}_{n\in\mathbb{N}} \subset \Lambda$ such that both $H^n \bullet M \to H \bullet M$ and $H^n \bullet \overline{M} \to H \bullet \overline{M}$ in $\mathcal{H}^1$, and both $H^n \bullet A \to H \bullet A$ and $H^n \bullet \overline{A} \to H \bullet \overline{A}$ except possibly on an evanescent set. As H^n is locally bounded, by Lemma 12.3.7 the stochastic integral $H^n \bullet (M - \overline{M})$ is equal to the Stieltjes integral $H^n \bullet (A - \overline{A})$. Taking the limit as $n \to \infty$ we see

$$H_0 X_0 + H \bullet M + H \bullet A = H_0 X_0 + H \bullet \overline{M} + H \bullet \overline{A}.$$

The integrals $H \bullet M$ and $H \bullet A$ are uniquely defined, and it follows that the integral $H \bullet X$ is also unique. □

Corollary 12.3.14. *For X, H as in Definition 12.3.10,*

(i) $H\bullet X^c = (H\bullet X)^c$ and $H\Delta X = \Delta(H\bullet X)$ up to indistinguishability and
(ii) if T is any stopping time, the processes $H\bullet X^T$ and $(H\bullet X)^T$ are indistinguishable.

Proof. For any decomposition $X = X_0 + M + A$, we know $X^c = M^c$. Therefore, the processes $H\bullet X^c$ and $(H\bullet X)^c$ are both defined using the local martingale integral construction (Theorem 12.3.3), and so are indistinguishable by Definition 12.3.1(ii). Similarly,

$$\begin{aligned} H\Delta X &= H(\Delta M + \Delta A) \\ &= \Delta(H\bullet M) + \Delta(H\bullet A) = \Delta(H\bullet X)\,. \end{aligned}$$

For the second statement, we have

$$\begin{aligned} H\bullet X^T &= H\bullet(M^T + A^T) = H\bullet M^T + H\bullet A^T \\ &= (H\bullet M)^T + (H\bullet A)^T = (H\bullet X)^T. \end{aligned}$$

□

Corollary 12.3.15. *For any semimartingale X and any X-integrable H,*

$$[H\bullet X] = H^2\bullet[X]$$

and when each of the terms exists,

$$\langle H\bullet X\rangle = H^2\bullet\langle X\rangle.$$

Proof. Decomposing $H\bullet X = H\bullet M^c + H\bullet M^d + H\bullet A$, we see that the result is true for the continuous component $X\bullet M^c$, by definition of the stochastic integral. For $H\bullet(M^d + A)$, we have

$$[H\bullet(M^d + A)]_t = \sum_{0<s\leq t} H_s^2(\Delta X_s)^2 = H^2\bullet[M^d + A]_t.$$

The first result follows. The second result is obtained by applying Π_p^* to the first. □

Remark 12.3.16. The previous lemma, along with the linearity of the integral, allows us to manipulate stochastic integrals in an easy way. In particular, for the integral $Y = Y_0 + H\bullet X$, we can write $dY = HdX$, where the objects 'dY' and 'dX' are simply a notational convenience, whose meaning is defined by reference to the stochastic integral. Using Lemma 12.3.23 we shall see that $KdY = KHdX$ for any Y-integrable process K, so the vector-space operations of multiplication (by a Y-integrable process) and addition are well defined *in the differential notation.* This can simplify calculations significantly.

Remark 12.3.17. For X a local martingale, it is important to note that there may be processes H such that $H \bullet X$ is well defined (as a semimartingale integral) but is *not* a local martingale. An example of this is given in Exercise 12.6.12. These questions are explored more generally by Émery [76] and Cherny and Shiryaev [32]. However, if H is locally bounded (for example, if H is left-continuous, cf. Lemma 7.3.20), then, by Theorem 12.3.3, these complications cannot arise.

The next theorem shows that, if we can restrict our attention to processes known to be special semimartingales, then the natural simplification of the integral into two components (one martingale, one predictable finite variation) is possible.

Theorem 12.3.18. *Suppose X is a special semimartingale with canonical decomposition $X = M + A$ and H is X-integrable. Then $H \bullet X$ is a special semimartingale if and only if H is M-integrable in the sense of local martingales (Definition 12.3.1) and H is a.s. locally $|dA|$-integrable, in which case the canonical decomposition of $H \bullet X$ is*

$$H \bullet X = H \bullet M + H \bullet A.$$

Proof. Sufficiency of the conditions is clear, so we consider necessity only. Suppose $H \bullet X$ is a special semimartingale.

As H is X-integrable, we know that there is some semimartingale decomposition $X = N + B$ such that $(H^2 \bullet [N])^{1/2} \in \mathcal{A}^+_{\text{loc}}$ and $|H|$ is $|dB|$-integrable. From Theorem 11.6.10, B is locally integrable, and hence it is easy to show that $\Pi^*_p B = A$. As $H \bullet X$ is also a special semimartingale and $H \bullet X = H \bullet N + H \bullet B$ is a semimartingale decomposition, again from Theorem 11.6.10 we know $H \bullet B$ is locally integrable. Therefore, we can define the process $\Pi^*_p(H \bullet B)$. Using Theorem 8.2.9 to pass the projection through the integral, which Theorem 8.2.19 guarantees is well defined,

$$\Pi^*_p(H \bullet B) = H \bullet (\Pi^*_p B) = H \bullet A.$$

It follows that the canonical decomposition of $H \bullet X$ has finite variation term $H \bullet A$, and H is a.s. locally $|dA|(\omega)$-integrable.

Now consider

$$H\Delta M = H\Delta X - H\Delta A = H\Delta N + H\Delta(B - A).$$

The terms on the right are both locally integrable. Therefore, by sublinearity of the square root, one can see that

$$(H^2 \bullet [M])^{1/2} = \Big(H^2 \bullet \langle M^c \rangle + \sum_{0<t\leq(\cdot)} (H_t \Delta M_t)^2\Big)^{1/2}$$

is also locally integrable. It follows that $H \bullet M$ is a well-defined local martingale integral. By Theorem 12.3.13, we conclude $H \bullet X = H \bullet M + H \bullet A$, and the term on the right is the canonical decomposition. □

Remark 12.3.19. Jacod [108] gives an alternative proof, where H is first approximated by a bounded process (in which case the result is easy), and then a limit is taken using convergence locally in $\mathcal{H}^1$ and $\mathcal{A}$.

Corollary 12.3.20. *Let X be a local martingale and H an X-integrable process. Then $H\bullet X$ is a local martingale if and only if $H\bullet X$ is a special semimartingale. (Necessary and sufficient conditions for this are given by Theorem 11.6.10.)*

Proof. Every local martingale is a special semimartingale, so it is enough to prove the converse. Clearly, X is a special semimartingale, so, if $H\bullet X$ is also a special semimartingale, then, from Theorem 12.3.18, we know H is X-integrable in the sense of local martingales; hence $H\bullet X$ is a local martingale. □

Corollary 12.3.21. *(i) The space of X-integrable processes is a vector space.*

(ii) The integral is linear in the integrand, that is, for any semimartingale X and any X-integrable H and G, any $\alpha \in \mathbb{R}$, we know $(\alpha H + G)$ is also X-integrable, and

$$(\alpha H + G)\bullet X = \alpha(H\bullet X) + G\bullet X.$$

(iii) The integral is linear in the integrator, that is, for any semimartingales X and Y and any H which is both X and Y-integrable, any $\alpha \in \mathbb{R}$, we know H is also $(\alpha X + Y)$-integrable, and

$$H\bullet(\alpha X + Y) = \alpha(H\bullet X) + H\bullet Y.$$

Proof. For the case where X is a special semimartingale, all of these statements follow from Theorem 12.3.18, together with Corollary 12.3.5 and linearity of the (classical) Stieltjes integral (with respect to a finite variation process). Statement (iii) is straightforward in general. In the case where X is a general semimartingale, statements (i) and (ii) form a significantly more difficult result, which we prove as part of the next theorem, see Appendix A.6.1. □

We now state a result which implies that the class of integrands for which we have defined the integral is the largest class satisfying some natural conditions. Together with the Bichteler–Dellacherie–Mokobodzki theorem (Theorem 12.3.26), this gives us confidence that the theory of stochastic integration with respect to semimartingales is, in some sense, exhaustive. A proof can be found in Appendix A.6.1.

Theorem 12.3.22. *The class $L(X)$ of X-integrable processes is a vector space, equal to the largest class $L'(X)$ of predictable processes H such that one can define a bilinear*[2] *map $\mathfrak{I}$, with $\mathfrak{I}(H,X)$ defined for all semimartingales X and all $H \in L'(X)$, which satisfies*

(i) $\mathfrak{I}(H,X)$ *is a semimartingale,*
(ii) $\mathfrak{I}(H,X)^c = H \bullet X^c$, *where X^c is the continuous martingale part of X (with $\bullet$ denoting the Itô integral of Theorem 12.3.3 with respect to a continuous martingale),*
(iii) $H \Delta X = \Delta \mathfrak{I}(H,X)$ *and*
(iv) $\mathfrak{I}(H,X) = H \bullet X$ *whenever X has finite variation and H is locally $|dX|$-integrable, in the sense of Stieltjes integrals.*

Furthermore, $\mathfrak{I}$ is uniquely defined by these properties and is given by the stochastic integral $\mathfrak{I}(H,X) = H \bullet X$ as defined in Definition 12.3.10.

The following lemma extends Corollary 12.2.5 to the space of semimartingales. The setting where K is locally bounded is an important special case.

Lemma 12.3.23. *For X a semimartingale, H an X-integrable process and K a predictable process,*

(i) KH *is X-integrable if and only if K is $(H \bullet X)$-integrable and*
(ii) *in this case,* $(HK) \bullet X = K \bullet (H \bullet X)$.

Proof. Suppose first that K is $(H \bullet X)$-integrable. Then we can define the process $Y = K \bullet (H \bullet X)$, which is a semimartingale satisfying

$$Y^c = K \bullet (H \bullet M)^c = K \bullet (H \bullet M^c)$$

and

$$\Delta Y = K \Delta (H \bullet M) = KH \Delta M.$$

As M^c is continuous, we know that $(KH) \bullet M^c$ is locally square integrable, so Corollary 12.2.5 shows that

$$K \bullet (H \bullet M^c) = (KH) \bullet M^c = Y^c.$$

If X and $H \bullet X$ are of finite variation, then we can regard $H = d(H \bullet X)/dX$ as a Radon–Nikodym derivative in the sense of measures on Σ_p. It follows that $K \bullet (H \bullet X) = (KH) \bullet X$ whenever K is locally $|H||dX|$-integrable. The linearity of the integrals shows that this definition of Y is linear in X, H and K.

[2] We mean that the map is bilinear on its domain of definition, that is,

- $\mathfrak{I}(\alpha H + K, X) = \alpha \mathfrak{I}(H,X) + \mathfrak{I}(K,X)$ for all $\alpha \in \mathbb{R}$, $H, K \in L(X)$ and
- $\mathfrak{I}(H, \alpha X + Y) = \alpha \mathfrak{I}(H,X) + \mathfrak{I}(H,Y)$ for all $\alpha \in \mathbb{R}$, $H \in L(X) \cap L(Y)$,

up to evanescent sets depending on the arguments.

By Theorem 12.3.22, Y satisfies the requirements to be the stochastic integral $(KH)\bullet X$, so we must have $KH \in L(X)$, that is KH is X-integrable, and $(KH)\bullet X = Y = K\bullet(H\bullet X)$. The converse statement (assuming KH is X-integrable) can be proven in a similar way. $\square$

We now state a significant result, due to Bichteler, Dellacherie and Mokobodzki, which characterizes semimartingales in terms of stochastic integrals.

Definition 12.3.24. *A process X is called a* good integrator *if for any sequence $\{H^n\}_{n\in\mathbb{N}} \in \Lambda$ converging uniformly to a process H (that is, $\sup_t \|H_t^n - H_t\|_\infty \to 0$) we know that $(H^n\bullet X)_t$ converges in probability for all $t < \infty$.*

An alternative characterization is given by the following lemma, whose proof is left as an exercise (Exercise 12.6.2).

Lemma 12.3.25. *A process X is a good integrator if and only if the set*

$$\mathcal{J} = \{H\bullet X; H \in \Lambda, |H| \leq 1\}$$

is bounded in probability, that is, for any $\epsilon > 0$, there exists $k > 0$ such that $\sup_{J\in\mathcal{J}} P(|J| > k) < \epsilon$.

We can now state the theorem of interest, which states that semimartingales are the only good integrators. A proof is given in Appendix A.6.2.

Theorem 12.3.26 (Bichteler–Dellacherie–Mokobodzki Theorem). *A càdlàg adapted process X is a good integrator if and only if it is a semimartingale.*

12.4 Émery's Semimartingale Topology

We will make much use of the following notion of convergence, which we already came across briefly in Lemma 5.5.6.

Definition 12.4.1. *A sequence of processes $\{X^n\}_{n\in\mathbb{N}}$ will be said to converge* uniformly on compacts in probability *(or* converge ucp*) if*

$$P\big((X^n - X)_t^* > \epsilon\big) \to 0 \qquad \text{as } n \to \infty$$

for every $t > 0$ and $\epsilon > 0$.

Before proceeding further, we observe that the following is a simple consequence of the 'uniform' convergence of ucp.

Lemma 12.4.2. *If $\{X^n\}_{n\in\mathbb{N}}$ is a sequence of càdlàg processes with $X^n \to X$ in ucp for some process X, then X is càdlàg and $\Delta X^n \to \Delta X$ in ucp.*

Proof. By definition, for any $\delta > 0$ we can find a set A such that $P(A) \geq 1-\delta$ and $\sup_{t\leq T} |X_t^n - X_t| I_A \to 0$ a.s. As in Lemma 5.5.6, this implies X is càdlàg on A, so by sending $\delta \to 0$ we see X is càdlàg a.s. Furthermore,

$$\sup_{t\leq T}\{|X_{t-} - X_{t-}^n|\} \leq \sup_{t\leq T}\Big\{\sup_{s<t}\{|X_s - X_s^n|\}\Big\} \to 0$$

in probability, so $\Delta X^n = X^n - \{X_{t-}^n\}_{t\geq 0} \to X - \{X_{t-}\}_{t\geq 0} = \Delta X$ in ucp. □

The Bichteler–Dellacherie–Mokobodzki theorem gives a characterization of semimartingales as 'good integrators'. This perspective, particularly with the characterization of Lemma 12.3.25, leads us to define a topology for semimartingales slightly stronger than that of ucp convergence, which will allow us to prove several useful convergence results. This topology is due to Émery [75]. To motivate this definition, recall that if $A : U \to V$ is a linear operator between Banach spaces, then the operator norm of A is defined by

$$\|A\|_{\mathrm{op}} = \sup\big\{\|A(u)\|_V : u \in U, \|u\|_U \leq 1\big\}.$$

Thinking of the stochastic integral as a linear map between processes, the following definition becomes natural.

Definition 12.4.3. *The norm-like-map*[3]

$$X \mapsto \|X\|_{\mathrm{ucp}} := \sum_{n=1}^{\infty} 2^{-n} E[1 \wedge X_n^*]$$

defines the topology of ucp convergence. We then define the Émery 'norm' of a semimartingale to be

$$X \mapsto \|X\|_{\mathcal{S}} := \sup\big\{\|H \bullet X\|_{\mathrm{ucp}} :\ H \in \Lambda, |H| \leq 1\big\}. \tag{12.2}$$

The topology so defined will be called the semimartingale topology (or Émery topology) and we write $X^n \to X$ in $\mathcal{S}$ when $\|X^n - X\|_{\mathcal{S}} \to 0$.

Remark 12.4.4. Under the metric induced by $\|\cdot\|_{\mathcal{S}}$, we shall see that $\mathcal{S}$, the space of semimartingales, is complete. (See Theorem 12.4.15.)

[3] These maps are not norms, as they do not satisfy the property of positive homogeneity. Yosida [189] calls such objects quasinorms; however this terminology does not appear widespread. Nevertheless, the distance functions $d_{\mathrm{ucp}}(X,Y) = \|X-Y\|_{\mathrm{ucp}}$ and $d_{\mathcal{S}}(X,Y) = \|X-Y\|_{\mathcal{S}}$ are metrics, as they satisfy the triangle inequality, and define the desired topologies in the usual way. Furthermore, these metrics are translation invariant and, for any $\lambda > 1$, we know $\|\lambda X\|_{\mathrm{ucp}} \leq \lambda \|X\|_{\mathrm{ucp}}$ and

$$\|\lambda X\|_{\mathcal{S}} = \sup\big\{\|H \bullet X\|_{\mathrm{ucp}} : H \in \Lambda, |H| \leq \lambda\big\} \leq \lambda \|X\|_{\mathcal{S}}.$$

If $|X| < m$ for some $m \in \mathbb{R}$, then, for $\lambda \leq 1/m$, we know $\|\lambda X\|_{\mathrm{ucp}} = \lambda \|X\|_{\mathrm{ucp}}$.

Remark 12.4.5. It is easy to check that the semimartingale topology is weaker than the $\mathcal{H}^p$ topology, when restricted to the martingales. In particular, if $X^n \to X$ in $\mathcal{H}^p$, then $X^n \to X$ in $\mathcal{S}$.

The key point of this topology is that convergence of $X^n \to X$ in $\mathcal{S}$ will guarantee convergence of the integrals $H \bullet X^n \to H \bullet X$ in ucp, for any simple H with $|H| \leq 1$. In fact a stronger result holds, see Theorem 12.4.13. We now give an alternative useful characterization of convergence.

Lemma 12.4.6. *Let $\{X^n\}_{n\in\mathbb{N}}$ be a sequence of semimartingales and X a semimartingale. Then $\{X^n\}_{n\in\mathbb{N}}$ converges to X in the semimartingale topology if and only if, for every $t \geq 0$ and every sequence $\{H^n\}_{n\in\mathbb{N}} \subset \Lambda$ with $|H^n| \leq 1$,*

$$\|H^n \bullet (X^n - X)\|_{\text{ucp}} \to 0.$$

Proof. First suppose $\|X^n - X\|_{\mathcal{S}} \to 0$. Then, as H^n is of the class in (12.2),

$$\|H^n \bullet (X^n - X)\|_{\text{ucp}} \leq \|X^n - X\|_{\mathcal{S}} \to 0$$

Conversely, suppose there exists $\epsilon > 0$ such that $\|X^n - X\|_{\mathcal{S}} > \epsilon$ for infinitely many n. Then, for such n, there exists a process $H^n \in \Lambda$ with $|H^n| \leq 1$ such that

$$\|H^n \bullet (X^n - X)\|_{\text{ucp}} > \epsilon$$

and we see that the stated convergence does not hold. □

Lemma 12.4.7. *If $\{X^n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in $\|\cdot\|_{\mathcal{S}}$, then the sequence is also Cauchy in $\|\cdot\|_{\text{ucp}}$, that is, the semimartingale topology is stronger than the ucp topology on the semimartingales.*

Proof. Taking $H = 1$ we see from the definitions that

$$\|X^n - X^m\|_{\text{ucp}} = \|H \bullet (X^n - X^m)\|_{\text{ucp}} \leq \|X^n - X^m\|_{\mathcal{S}}. \qquad \square$$

We now see that local semimartingale convergence implies global semimartingale convergence (and similarly for ucp).

Lemma 12.4.8. *Let $\{X^n\}_{n\in\mathbb{N}}$ be a sequence in $\mathcal{S}$, $X \in \mathcal{S}$ and $\{T_n\}_{n\in\mathbb{N}}$ be an increasing sequence of stopping times with $T_n \to \infty$ a.s. such that*

$$(X^n)^{T_m} \to X^{T_m} \text{ in } \mathcal{S} \text{ for all } m.$$

Then $X^n \to X$ in $\mathcal{S}$.

Similarly if $\{X^n\}_{n\in\mathbb{N}}$ and X are càdlàg adapted processes and the convergence is ucp.

Proof. Fix $t > 0$. As $T_m \to \infty$ a.s., for any $\delta > 0$ we can find m such that $P(T_m < t) < \delta$. Then, as $(X^n)^{T_m} \to X^{T_m}$ in $\mathcal{S}$, for any $\epsilon > 0$ and any sequence $\{H^n\}_{n\in\mathbb{N}} \subset \Lambda$ with $|H^n| \leq 1$, we can find n sufficiently large that

$$P\big((H^n \bullet (X^n - X))^*_t > \epsilon\big) \leq P\big((H^n \bullet (X^n - X)^{T_m})^*_t > \epsilon\big) + \delta \leq 2\delta,$$

so $X^n \to X$ in $\mathcal{S}$, by Lemma 12.4.6. The same argument holds for ucp convergence, taking $H^n \equiv 1$. □

Remark 12.4.9. It will be useful to note that this result holds even if we only assume $(X^n)^{T_m} I_{[\![0,T_m[\![} \to X^{T_m} I_{[\![0,T_m[\![}$ in $\mathcal{S}$ for all m, and the proof is unchanged.

Using this topology, we obtain a very powerful convergence theorem for stochastic integrals, where the integrands are only assumed to converge pointwise (and be dominated by an X-integrable process).

Theorem 12.4.10 (Stochastic Dominated Convergence Theorem). *Let X be a semimartingale and $\{H^n\}_{n\in\mathbb{N}}$ be a sequence of predictable processes with $H^n_t \to H_t$ a.s. for every $t \geq 0$. Suppose $|H^n| \leq G$ for some X-integrable process G (in the sense of Definition 12.3.10). Then H is X-integrable and $H^n \bullet X$ converges to $H \bullet X$ in $\mathcal{S}$ (and hence ucp).*

Proof. We have $|H^n - H| \leq 2G$, so $\{H^n\}_{n\in\mathbb{N}}$ and H are all X-integrable processes. Let $Y^n = H^n \bullet X$ and $Y = H \bullet X$. For any sequence $\{K^n\}_{n\in\mathbb{N}} \subset \Lambda$ with $|K| \leq 1$, we have, from Lemma 12.3.23,

$$K^n \bullet (Y^n - Y) = K^n(H^n - H) \bullet X$$

Let $Z^n = K^n(H^n - H)$ and $X = X_0 + M + A$ be a decomposition of X such that G satisfies the integrability requirements of Definition 12.3.10 with this decomposition. We know $|Z^n| < 2G$, so Z^n is X-integrable with the same decomposition, and $Z^n_t \to 0$ a.s. for every $t \geq 0$.

We know that Z^n is locally $|dA|$-integrable, so, by Exercise 3.4.16 and Lebesgue's dominated convergence theorem (Theorem 1.3.34), for any $t > 0$ and $\delta > 0$ we can find a stopping time T with $P(T \leq t) < \delta$ and $(Z^n \bullet A)^*_T \leq \int_{[0,T]} |Z^n_t||dA_t| \to 0$ a.s. Therefore, for any $\epsilon > 0$, we can find n sufficiently large that

$$P\big((Z^n \bullet A)^*_t \geq \epsilon\big) \leq P\big((Z^n \bullet A)^*_T > \epsilon\big) + \delta \leq 2\delta,$$

so $Z^n \bullet A \to 0$ in ucp, which implies $H^n \bullet A \to H \bullet A$ in $\mathcal{S}$, by Lemma 12.4.6.

Now let S be any stopping time such that $E[(G^2 \bullet [M])^{1/2}_S] < \infty$. Again by Lebesgue's dominated convergence theorem, we know that

$$E\Big[\Big(\int_{[0,S\wedge t]} (Z^n_s)^2 d[M]_s\Big)^{1/2}\Big] \to 0$$

which implies $Z^n \bullet M \to 0$ locally in $\mathcal{H}^1$. By Exercise 3.4.16, we see $Z^n \bullet M \to 0$ in ucp, and so $H^n \bullet M \to H \bullet M$ in $\mathcal{S}$. As $\|\cdot\|_{\mathcal{S}}$ satisfies the triangle inequality, the result follows. □

Corollary 12.4.11. *Let X be a semimartingale and $\{H^n\}_{n\in\mathbb{N}}$ be a sequence of uniformly locally bounded predictable processes, that is, for a localizing sequence $\{T_n\}_{n\in\mathbb{N}}$ we have $\sup_m |(H^m)^{T_n}| < K_n$, for some constants K_n. Suppose that $H^n_t \to H_t$ a.s. for every t. Then H is X-integrable and $H^n \bullet X \to H \bullet X$ in $\mathcal{S}$.*

Proof. First localize with T_n, and apply Theorem 12.4.10 to see $(H^m)^{T_n} \bullet X \to (H)^{T_n} \bullet X$ in $\mathcal{S}$ as $m \to \infty$. By Lemma 12.4.8, this implies $H^n \bullet X \to H \bullet X$ in $\mathcal{S}$. □

Corollary 12.4.12. *(i) For any càdlàg process X, any sequence $\lambda_n \to 0$, we know $\|\lambda_n X\|_{\text{ucp}} \to 0$.*

(ii) For any semimartingale X and any sequence of predictable processes $\{H^n\}_{n\in\mathbb{N}} \subset \Lambda$ with $\|(H^n)^*_\infty\|_\infty \to 0$, we know $\|H^n \bullet X\|_{\mathcal{S}} \to 0$.

Proof. (i) For any càdlàg process X, let $T_m = \inf\{t : |X_t| > m\}$. Then $|X^* I_{[\![0,T_m[\![}| \leq m$, so for all $\lambda_n \leq 1/m$,

$$\|\lambda_n X I_{[\![0,T_m[\![}\|_{\text{ucp}} = \lambda_n \|X I_{[\![0,T_m[\![}\|_{\text{ucp}} \to 0.$$

By the previous lemma and remark, using the fact $T_m \to \infty$, it follows that $\|\lambda_n X\|_{\text{ucp}} \to 0$, as desired.

(ii) This follows directly from Theorem 12.4.10. □

We can also obtain a strong convergence result in terms of the semimartingale integrator, for locally bounded integrands.

Theorem 12.4.13. *Let H be a locally bounded predictable process and $\{X^n\}_{n\in\mathbb{N}}$ a sequence of semimartingales converging to X in $\mathcal{S}$. Then $H \bullet X^n \to H \bullet X$ in $\mathcal{S}$.*

Proof. By localizing and rescaling, we can assume $|H| \leq 1$. Then suppose $H \in \Lambda$. By definition,

$$\begin{aligned}\|H \bullet (X^n - X)\|_{\mathcal{S}} &= \sup\left\{\|KH \bullet (X^n - X)\|_{\text{ucp}} : K \in \Lambda, |K| \leq 1\right\}\\ &\leq \sup\left\{\|K \bullet (X^n - X)\|_{\text{ucp}} : K \in \Lambda, |K| \leq 1\right\}\\ &= \|X^n - X\|_{\mathcal{S}} \to 0\end{aligned}$$

uniformly in H.

For $H \notin \Lambda$, we can approximate H pointwise by $\{H^m\}_{m\in\mathbb{N}} \in \Lambda$ with $|H^m| \leq 1$ (Corollary 7.4.3). Therefore,

$$\|H \bullet (X^n - X)\|_{\mathcal{S}} \leq \|H^m \bullet (X^n - X)\|_{\mathcal{S}} + \|(H^m - H) \bullet (X^n - X)\|_{\mathcal{S}}.$$

For any $\epsilon > 0$, as $H^m \in \Lambda$, using our earlier result we can take n sufficiently large that

$$\|H^m \bullet (X^n - X)\|_{\mathcal{S}} \leq \epsilon \quad \text{for all } m.$$

Then, leaving n fixed, by Theorem 12.4.10 we can take m sufficiently large that

$$\|(H^m - H) \bullet (X^n - X)\|_{\mathcal{S}} \leq \epsilon.$$

It follows that $\|H \bullet (X^n - X)\|_{\mathcal{S}} \leq 2\epsilon$, and taking $\epsilon \to 0$ gives the result. □

Remark 12.4.14. The previous results, along with the translation invariance of the metric, is enough to show that $\mathcal{S}$ is a topological vector space under the semimartingale topology. In particular, the operations of addition and scalar multiplication are continuous.

Theorem 12.4.15. *The space of semimartingales is complete under the semimartingale topology, that is, if $\{X^n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in $\|\cdot\|_{\mathcal{S}}$, then there exists a semimartingale X such that $X^n \to X$ in $\mathcal{S}$.*

Proof. By Lemma 12.4.7, we see that X^n converges ucp to a process X, and furthermore that this process X is almost surely càdlàg. We know that the sequence $H \bullet X^n$ converges ucp, uniformly in $H \in \Lambda$ with $|H| \leq 1$, and we call this limit $I(H, X)$. For $H \in \Lambda$, this agrees with the simple integral, that is $I(H, X) = H \bullet X$. Hence, if X is a semimartingale, then we know $X^n \to X$ in the semimartingale topology.

To show X is a semimartingale, we use the Bichteler–Dellacherie–Mokobodzki theorem (Theorem 12.3.26). For any sequence $\{H^m\}_{m\in\mathbb{N}} \subset \Lambda$ converging uniformly to a process H, we know that

$$\|H^m \bullet X - I(H, X)\|_{\mathrm{ucp}} \leq \|H^m \bullet X^n - H \bullet X^n\|_{\mathrm{ucp}} + \|H \bullet X^n - I(H, X)\|_{\mathrm{ucp}}.$$

Taking $m \to \infty$, as $\|H^m - H\|_\infty \to 0$, we see that

$$\lim_m \|H^m \bullet X - I(H, X)\|_{\mathrm{ucp}} \leq \|H \bullet X^n - I(H, X)\|_{\mathrm{ucp}}$$

and the right-hand side can be made arbitrarily small by taking n sufficiently large. Therefore, $H^m \bullet X$ converges ucp. It follows that X is a good integrator, and so is a semimartingale, by Theorem 12.3.26. □

The proof of the following theorem we leave to Appendix A.6.3, as it uses an approach we shall develop in Chapter 16 (however one can check that no intermediate result depends on this theorem).

Theorem 12.4.16. *For any semimartingale X, the space $\{H \bullet X\}_{H\in L(X)}$ is complete in the semimartingale topology.*

We now give one final useful result, which allows us to exchange the order of integration in stochastic integrals.

Lemma 12.4.17. *Let $(Y, \mathcal{Y})$ be a measurable space and $X_n : Y \times \Omega \to \mathbb{R}$ a sequence of $\mathcal{Y} \otimes \mathcal{F}$-measurable functions such that $\{X_n(y, \cdot)\}_{n\in\mathbb{N}}$ converges in probability for every y. Then there exists a $\mathcal{Y} \otimes \mathcal{F}$-measurable function X such that $X(y, \cdot)$ is the limit in probability of $X_n(y, \cdot)$ for every y.*

Proof. For any $k > 0$, any $y \in \mathcal{Y}$, we know that $P\big(|X_n(y, \cdot) - X_m(y, \cdot)| > 1/k\big) \to 0$ as $m, n \to \infty$. Therefore, we define the sequence

$$n_k(y) = \inf\Big\{m > n_{k-1}(y) : \quad P\big(|X_a(y, \cdot) - X_b(y, \cdot)| > 1/k\big) \leq 1/k \\ \text{for all } a, b > m\Big\}.$$

Then $n_k(\cdot)$ is a $\mathcal{Y}$-measurable function, and

$$P\Big(\sup_y |X_{n_k(y)}(y,\cdot) - X_{n_{k'}(y)}(y,\cdot)| > \frac{1}{k}\Big) \le \frac{1}{k}$$

for every $k' > k$. Therefore, $X_{n_k(y)}(y,\cdot)$ converges in probability, uniformly in y. Let $X(y,\cdot) = \lim_{k\to\infty} X_{n_k(y)}(y,\cdot)$, and we see that X is $\mathcal{Y}\otimes\mathcal{F}$-measurable and so is the desired limit. □

Theorem 12.4.18 (Stochastic Fubini Theorem). *Let X be a semimartingale, $(Y,\mathcal{Y})$ be a measurable space, $\{H_t^y\}_{t\ge 0, y\in\mathcal{Y}}$ be a family of predictable processes such that $\sup_y |H^y| \le G$ for some X-integrable process G, and $(y,t,\omega)\mapsto H_t^y(\omega)$ is $\mathcal{Y}\otimes\Sigma_p$-measurable. Then there exists a $\mathcal{Y}\otimes\Sigma_o$-measurable function $K : Y\times[0,\infty[\times\Omega \to \mathbb{R}$ such that*

$$K(y,\cdot,\cdot) = H^y\bullet X \qquad \text{for all } y$$

up to indistinguishability. Furthermore, if ν is a finite measure on $(Y,\mathcal{Y})$, then, up to indistinguishability,

$$\int_Y K(y,t,\cdot)d\nu_y = \int_{[0,t]} \Big(\int_Y H_s^y d\nu_y\Big) dX_s.$$

Proof. First suppose that $H_t^y(\omega) = h(y)g(t,\omega)$ for some $\mathcal{Y}$-measurable h and predictable g. Then we write $K(y,\cdot,\cdot) = h(y)(g\bullet X)$, and the result is clear.

By Lemma 12.4.17, if we have a sequence $H^{y,(n)}$ such that $K^{(n)}(y,\cdot,\cdot) = H^{y,(n)}\bullet X$ converges in probability for each y, then we can find a suitably measurable function K such that $K(y,\cdot,\cdot) = \lim_{n\to\infty} H^{y,(n)}\bullet X$, (the limit being taken in probability). Therefore, by stochastic dominated convergence (Theorem 12.4.10) and the function-space monotone class theorem (Theorem 7.4.1), we see that the result holds true for every uniformly bounded H.

Finally, for general H satisfying the integrability properties of the theorem, by truncating H^y with $(-n\vee H^y\wedge n)$, we define

$$K^{(n)}(y,\cdot,\cdot) = (-n\vee H^y\wedge n)\bullet X.$$

By stochastic dominated convergence, as $n\to\infty$, we see that $K^{(n)}(y,\cdot,\cdot)$ converges ucp . So, by Lemma 12.4.17 we can construct the desired function K as its limit.

Furthermore, let $n_k(y)$ be the sequence constructed in the proof of Lemma 12.4.17 when applied to $K^{(n)}$. As ν is a finite measure, for any $\epsilon > 0$, as $k\to\infty$ we see

$$\begin{aligned} &P\Big(\int_Y |K^{(n_k(y))}(y,t,\cdot) - K(y,t,\cdot)|d\nu_y > \epsilon\Big) \\ &\le P\Big(\sup_y |K^{(n_k(y))}(y,t,\cdot) - K(y,t,\cdot)| > \frac{\epsilon}{\nu(Y)}\Big) \to 0. \end{aligned}$$

Therefore, by stochastic dominated convergence, taking limits in ucp, we have that, up to indistinguishability,

$$\begin{aligned}\int_Y K(y,t,\cdot)d\nu_y &= \lim_{k\to\infty}\int_Y K(y,t,\cdot)d\nu_y\\ &= \lim_{k\to\infty}\int_{[0,t]}\Big(\int_Y (-n_k(y)\vee H_s^y\wedge n_k(y))d\nu_y\Big)dX_s\\ &= \int_{[0,t]}\Big(\int_Y H_s^y d\nu_y\Big)dX_s.\end{aligned}$$

□

12.5 Vector Integration

The theory of integration we have constructed is now fairly complete in one-dimension. However, when we consider vector semimartingales, a gap remains, as is made clear by the following example. This section is based on Jacod [108], Mémin [132] and Cherny and Shiryaev [32].

Example 12.5.1. Let X be a semimartingale and H a predictable process. For any $H\in L(X)$, we naturally define the vector stochastic integral componentwise

$$Y := \begin{bmatrix}H\\H\end{bmatrix}^\top \bullet \begin{bmatrix}X\\-X\end{bmatrix} = H\bullet X + H\bullet(-X) = 0.$$

However, this definition cannot be used when $H\notin L(X)$, even though the 'integral' $Y=0$ is clearly still natural in this setting.

While this example may seem trivial, similar concerns lead to more delicate problems, as we now shall see.

Example 12.5.2. Let B^1 and B^2 be independent Brownian motions in a filtered probability space. Let $H_t = t$ and define $X^1 = B^1$ and $X^2 = (1-H)\bullet B^1 + H\bullet B^2$. Then the space

$$\big\{K^1\bullet X^1 + K^2\bullet X^2 : K^1\in L(X^1), K^2\in L(X^2)\big\}$$

is not closed in the semimartingale topology.

To see this, observe that, for any $\epsilon>0$, we know $1-(H+\epsilon)^{-1}\in L(X^1)$ and $(H+\epsilon)^{-1}\in L(X^2)$. Therefore, we can calculate

$$\begin{aligned}Y^\epsilon &:= (1-(H+\epsilon)^{-1})\bullet X^1 + (H+\epsilon)^{-1}\bullet X^2\\ &= B^1 - \frac{1}{H+\epsilon}\bullet B^1 + \frac{1-H}{H+\epsilon}\bullet B^1 + \frac{H}{H+\epsilon}\bullet B^2\\ &= \frac{\epsilon}{H+\epsilon}\bullet\big(B^1-B^2\big) + B^2.\end{aligned}$$

The quadratic variation

$$[Y^\epsilon - B^2]_t = \int_{[0,t]} \frac{2\epsilon^2}{(s+\epsilon)^2} ds = \frac{2\epsilon t}{\epsilon + t} \to 0$$

uniformly as $\epsilon \to 0$ and, therefore, $Y^\epsilon \to B^2$ in the semimartingale topology.

On the other hand, we can show that B^2 cannot be written in the form $K^1 \bullet X^1 + K^2 \bullet X^2$. To see this, suppose such a representation exists. Then

$$B^2 = K^1 \bullet B^1 + K^2(1-H) \bullet B^1 + K^2 H \bullet B^2. \tag{12.3}$$

Therefore, as B^1 and B^2 are orthogonal

$$t = [B^2]_t = [K^2(1-H) \bullet B^2, B^2] = \int_{]0,t]} K_t^2 H_t dt$$

which implies $K_t^2 = 1/H_t = 1/t$. Similarly, taking the covariation of B^2 with B^1, we have $0 = K_t^1 + K_t^2(1-t)$, so $K_t^1 = 1 - 1/t$. This implies that $K^1 \bullet B^1$ is not well defined (as K^1 is not locally Itô integrable with respect to the continuous martingale B^1, which would be implied by (12.3) and Theorem 12.3.18). Hence we have a contradiction.

Therefore, we have a sequence $\{Y^{1/n}\}_{n\in\mathbb{N}}$, which can be written in terms of stochastic integrals defined componentwise, which converges to a point with no such representation.

In order to avoid these problems, it is important to define the vector stochastic integral more carefully. Essentially, we wish to allow terms in the integrand to cancel *before* taking a stochastic integral. We shall present this theory fairly briefly, as many of the results follow with only mild modifications from the approach given in the scalar case. As we did before, we start by supposing our integrator is a local martingale. In this section, all vectors will be thought of as column vectors, so $x^\top y$ is the inner product of x and y. We first construct a convenient representation of the quadratic covariation matrix.

Lemma 12.5.3. *Let M be a d-dimensional vector local martingale with components M^i. Then there exists $C \in \mathcal{V}^+$ and an optional process π taking values in the positive semidefinite symmetric real matrices (i.e. such that $x^\top \pi x \geq 0$ for all $x \in \mathbb{R}^d$ and $\pi^\top = \pi$) such that*

$$[M^i, M^j] = \pi^{ij} \bullet C$$

for any $i, j \in \{1, 2, \dots, d\}$.

Proof. We know that, as signed measures on Σ_p, $[M^i, M^j]$ is absolutely continuous with respect to $[M^i]+[M^j]$. Therefore, taking $C = \sum_i [M^i]$ (for example), we can define a symmetric-matrix-valued optional process ρ with components

$$\rho^{ji} = \rho^{ij} = d[M^i, M^j]/dC \text{ for } i \leq j.$$

Taking a countable dense subset $\{\lambda_k\}_{k\in\mathbb{N}} \subset \mathbb{R}^d$, define the optional set $D_k := \{(\omega, t) : \lambda_k^\top \rho \lambda_k \geq 0\}$. As $\{\lambda_k\}_{k\in\mathbb{N}}$ is dense, the Hahn–Banach theorem shows that

$$\bigcap_k D_k = D := \{(\omega, t) : \lambda^\top \rho\lambda \geq 0 \text{ for all } \lambda \in \mathbb{R}^d\}.$$

Given $[\lambda_k^\top M] = (\lambda_k^\top \rho\lambda_k)\bullet C \geq 0$, we know that D_k^c and hence D^c are evanescent sets. Therefore, we can define $\pi = \rho I_D$, which has all the desired properties. □

Remark 12.5.4. If $[M]$ is the covariation matrix of M, then the result of Lemma 12.5.3 is simply that $[M] = \pi\bullet C$, as a matrix-valued integral. In general, we note that $[X, Y]$ can be defined to be the compensator of the outer product matrix $XY^\top$, and many calculations are simplified by using this notation.

Definition 12.5.5. *For M a d-dimensional semimartingale with $M_0 = 0$, H a d-dimensional predictable process, and $p \in [1, \infty[$, define the norm*

$$H \mapsto E\big[\big((H^\top \pi H)\bullet C\big)_\infty^{p/2}\big]^{1/p} =: \|H\|_{L^p(M)}.$$

We write $H \in L^p(M)$ if $\|H\|_{L^p(M)} < \infty$ and $H \in L^1_{\text{loc}}(M)$ if H is locally in $L^1(M)$. As usual, we define elements of $L^p(M)$ to be equivalent if $\|H - H'\|_{L^p(M)} = 0$, and do not distinguish between a process and its equivalence class.

As in the scalar case, it is easy to show that $\|H\|_{L^p(M)}$ is a norm, as it is made up of the composition of a Hilbert space norm and an L^p-norm. Similarly, a dominated convergence argument shows that the simple integrands are dense in the space $\{H : \|H\|_{L^p(M)} < \infty\}$ for each p.

Consideration of the definitions yields

$$\{H : \|H\|_{L^p(M)} < \infty\} = \{H : ((H^\top \pi H)\bullet C)^{p/2} \in \mathcal{A}^+\},$$

and as in the scalar case, for $p = 2$ we can modify the above construction by changing $[M^i, M^j]$ to $\langle M^i, M^j\rangle$. The space $L^2_{\text{loc}}(M)$ corresponds precisely to the space of predictable processes H such that $(H^\top \pi H)\bullet C$ is locally integrable.

Definition 12.5.6. *For M as in Definition 12.5.5 and a predictable process $H \in L^1_{\text{loc}}(M)$, we define the stochastic integral (in the sense of local martingales) to be the local martingale $X = H\bullet M$ such that, for any local martingale N,*

$$[X, N] = (H^\top K)\bullet C$$

where K is an optional vector process such that $[M^i, N] = K^i\bullet C$.

As before, we should at this stage only state that this defines 'a' stochastic integral, but the next theorem shows that the integral is uniquely defined.

Theorem 12.5.7. *For any $H \in L^1_{\text{loc}}(M)$, the stochastic integral (in the sense of local martingales) is uniquely defined.*

Proof. The construction of the integral for simple processes H is easy. First observe that simple processes are locally bounded, and hence $H^\top \bullet M = \sum_{i=1}^d H^i \bullet M^i$ is well defined. It is easy to show that these integrals satisfy the isometry

$$E\big[[H^\top \bullet X]_\infty^{1/2}\big] = \|H\|_{L^1(M)}$$

As the left-hand side is an equivalent norm on $\mathcal{H}^1$, approximating a given H locally by simple functions, we have a sequence of local martingales converging in $\mathcal{H}^1_{\text{loc}}$. By completeness, the limit exists. Furthermore, as the isometry is preserved, we know that the limit must be uniquely defined. □

Remark 12.5.8. From uniqueness and linearity, it is clear that if H^i is M^i-integrable for each i, then $H^\top \bullet M = \sum_i (H^i \bullet M^i)$. In particular, this is the case if H is locally bounded. Similarly, if M has uncorrelated components (i.e. $\langle M^i, M^j\rangle = 0$ for all $i \neq j$), then the componentwise sum must agree with the vector integral (as π is diagonal, so the isometry implies each component H^i must be M^i integrable).

We now move to considering the appropriate integrals with respect to vector processes of finite variation.

Definition 12.5.9. *Let A be a $\mathbb{R}^d$-valued càdlàg process with components of finite variation. Taking $V_t = \sum_i \int_{[0,t]} |dA^i|$, we know there is a càdlàg process such that $A^i = v^i \bullet V$, and that V and v^i are predictable whenever A is predictable. We write $L^{\text{FV}}(A)$ for the space of predictable processes H such that*

$$H^\top \bullet A := (H^\top v) \bullet V$$

is a process of finite variation. Note that $L^{\text{FV}}(A) = L^{\text{FV}}_{\text{loc}}(A)$.

Lemma 12.5.10. *(i) The notation $H^\top \bullet X$ is not ambiguous, that is, if we take a vector martingale of finite variation X, and H such that both integrals are defined, then the integrals agree.*

(ii) If B is the compensator of a locally integrable finite variation process X (defined componentwise), then the compensator of $H^\top \bullet X$ is $H^\top \bullet B$, and $H^\top \bullet B$ is well defined.

Proof. We simply sketch the proof, details can be found in Jacod [108] and Mémin [132]. For both statements, the technique is the same, we consider only the first case. Approximate H with a bounded process $H^{(n)} = HI_{\{\|H\| \leq n\}}$. As all terms are well defined for bounded processes, the vector integral coincides with the componentwise sum $(H^{(n)})^\top \bullet A = \sum_i (H^{i,(n)} \bullet M^i)$. As we know the result holds in the scalar case, it must hold for the componentwise sum, that is $(H^n)^\top \bullet_{\text{FV}} A = (H^n)^\top \bullet_{\mathcal{M}} A$ (where $\bullet_{\text{FV}}$ denotes the finite variation Stieltjes integral and $\bullet_{\mathcal{M}}$ the integral in the sense of martingales). Applying dominated convergence, we take the limit $n \to \infty$, and see that both the stochastic and Stieltjes vector integrals must agree. □

Naturally, we now proceed to the case of a general semimartingale.

Definition 12.5.11. *We say that a vector process H is integrable with respect to a vector semimartingale X, and write $H \in L(X)$, if there exists a decomposition $X = M + A$ such that $H^\top \bullet M$ and $H^\top \bullet A$ are well defined, as vector local martingale and vector Stieltjes integrals respectively. Naturally, we then define $H^\top \bullet X = H^\top \bullet M + H^\top \bullet A$.*

Remark 12.5.12. From Lemma 12.5.10(i), as in the scalar case, we see that the integral does not depend on the choice of decomposition (among those decompositions where the integrals are well defined). By effectively the same argument as in Theorem 12.3.18, we also see that for vector special semimartingales (that is, where X has components which are special semimartingales) the vector stochastic integral is special if and only if H is integrable with respect to the canonical decomposition $X = M+A$ (i.e. $H \in L^1(M) \cap L^{\mathrm{FV}}(A)$), and then has canonical decomposition $H^\top \bullet X = H^\top \bullet M + H^\top \bullet A$.

We state the following extensions of the scalar results without proof.

Theorem 12.5.13. *(i) The space $L(X)$ of X-integrable processes is a vector space.*

(ii) The integral is linear in the integrand, that is, for any semimartingale X and any $H, G \in L(X)$, any $\alpha \in \mathbb{R}$, we know $(\alpha H + G) \in L(X)$ and

$$(\alpha H + G)^\top \bullet X = \alpha(H^\top \bullet X) + G^\top \bullet X.$$

(iii) The integral is linear in the integrator, that is, for any semimartingales X and Y and any $H \in L(X) \cap L(Y)$, any $\alpha \in \mathbb{R}$, we know $H \in L(\alpha X + Y)$ and

$$H^\top \bullet (\alpha X + Y) = \alpha(H^\top \bullet X) + H^\top \bullet Y.$$

Here (ii) and (iii) are up to evanescent sets, which may depend on the arguments.

As in Corollary 12.3.21, it is easy to show (iii), and that (i) and (ii) hold in the case when all terms are special semimartingales. The general case can be proven in much the same way as Theorem 12.3.22 (which one can show still holds), this is done in full in Jacod [108]. The following result then follows in essentially the same way as Lemma 12.3.23.

Lemma 12.5.14. *For X a semimartingale, H an X-integrable process and K a predictable scalar process, KH is X-integrable if and only if K is $(H^\top \bullet X)$-integrable and, in this case, $(HK)^\top \bullet X = K \bullet (H^\top \bullet X)$.*

We can also generalize the (scalar) statement $[H \bullet X, K \bullet Y] = HK \bullet [X, Y]$.

Lemma 12.5.15. *Let X and Y be vector semimartingales. Let $H \in L(X)$ and $K \in L(Y)$. Let C be an increasing process, and π, ρ, σ optional matrix valued processes, such that*

$$[X] = \pi \bullet C, \qquad [X, Y] = \rho \bullet C, \qquad [Y] = \sigma \bullet C.$$

Then $H^\top \rho K$ is C-integrable, and

$$[H^\top \bullet X, K^\top \bullet Y] = H^\top \rho K \bullet C.$$

Proof. First consider the case when Y is scalar and $K = 1$. We know that we can write $X = M + A$, where $H \in L^1(M) \cap L^{\mathrm{FV}}(A)$. Furthermore, we can assume that C is chosen such that $[M]$ and $[A]$ are both absolutely continuous with respect to C, and so we can find optional processes κ and λ with

$$[M] = \kappa \bullet C, \qquad [A] = \lambda \bullet C.$$

From Definition 12.5.5, we know that

$$\begin{aligned}
[H^\top \bullet X, Y]_t &= [H^\top \bullet M + H^\top \bullet A, X]_t \\
&= H^\top \bullet [M, Y]_t + \sum_{s \leq t} \Delta Y_s H_s \Delta A_s \\
&= H^\top \bullet [M, Y]_t + H^\top \bullet [A, Y]_t = H^\top \bullet [X, Y]_t.
\end{aligned}$$

Integrability of H with respect to $[Y, X]$ is guaranteed by the existence of quadratic variation, together with the above formula. (To see this, one can think of approximating H with a bounded process, then taking a limit.) Hence the result is proven in this case.

Now note that all terms are well defined, in the sense that the dimensions agree whenever a product is taken (no matter what the dimensions of X and Y), and that $H^\top \bullet X$ and $K^\top \bullet Y]$ are scalar semimartingales. Using our above argument, and the fact $[X, Y]^\top = [Y, X]$, we have

$$\begin{aligned}
[H^\top \bullet X, K^\top \bullet Y] &= H^\top \bullet [X, K^\top \bullet Y] = H^\top \bullet (K^\top \bullet [Y, X])^\top \\
&= H^\top \bullet ((K^\top \rho)^\top \bullet C) = (H^\top \rho K) \bullet C.
\end{aligned}$$
□

Finally, we note that our key concern with the componentwise sum has been resolved, that is, the space of integrals is complete. The proof is almost identical to the scalar case, which can be found in Appendix A.6.3.

Theorem 12.5.16. *For any semimartingale X, the space $\{H \bullet X\}_{H \in L(X)}$ is complete in the semimartingale topology.*

12.5.1 The Infinite Dimensional Case

Before concluding, we make a couple of comments about the case when X is an infinite dimensional process whose components are semimartingales. One approach to this theory is given by Mikulevicius and Rozovskii [137], see also De Donno, Guasoni and Pratelli [50] and references therein. Alternatively, one can consider these issues through an extension of the theory of Random Measures, the basic case of which we shall consider in the coming chapter, see Bichteler [14] for details of this approach.

However, a significant case which can be easily treated is when the components of X are independent Brownian motions (some generalizations of this are easily obtained; we shall content ourselves with the simplest case). This arises naturally in many settings, and can be thought of as the simplest case of a 'Brownian motion in a Hilbert space' (the space here being ℓ_2), which is considered in more detail by Carmona and Teranchi [30], Da Prato and Zabczyk [39] or the lecture notes of Hairer [91].

Suppose X is such a process, that is, X^i is a Brownian motion for each $i \in \mathbb{N}$, and X^i and X^j are independent for $i \neq j$. This is called a 'cylindrical Brownian motion in ℓ_2'; however it is easy to see that

$$P(X_t \in \ell_2) = P\Big(\sum_{i\in\mathbb{N}} (X^i)_t^2 < \infty\Big) = 0.$$

Nevertheless, if H is a constant in ℓ_2, then

$$E\Big[\Big(\sum_{i\leq n} H^i X_t^i\Big)^2\Big] = \sum_{i\leq n} (H^i)^2 t \leq \|H\|_{\ell_2} t,$$

so we can define $H^\top X$ as the $\mathcal{H}^2$-limit of $\sum_{i\leq n} H^i X^i$. This simple fact allows us to define the stochastic integral with respect to a cylindrical Brownian motion.

Definition 12.5.17. *Let X be a cylindrical Brownian motion in ℓ_2, that is, a sequence of independent Brownian motions. For any process H taking values in ℓ_2, and such that*

$$\|H\|_{L^1(X)} := E\Big[\Big(\int_{[0,\infty[} \|H\|_{\ell_2} dt\Big)^{1/2}\Big] < \infty,$$

we take a limit in $\mathcal{H}^1$ to define

$$H^\top \bullet X := \lim_n \Big(\sum_{i\leq n} H^i \bullet X^i\Big).$$

Remark 12.5.18. In the usual way, we localize to define the integral for processes H locally in $L^1(X)$. As X is continuous, this is certainly true whenever H is locally ℓ_2-bounded. Furthermore, as X has independent components, this definition agrees with the definition of the vector stochastic integral in a natural way (cf. Remark 12.5.8).

12.6 Exercises

Exercise 12.6.1. For W a Brownian motion starting at zero, for what $\alpha \in \mathbb{R}$ is the integral $(H \bullet W)$ well defined, when $H_t = t^\alpha$? For what t is the integral $\int_{[0,t]} (s-1)^{-1} dW_s$ well defined? Find $E\big[\big(\int_{]0,t]} W_s dW_s\big)^2\big]$.

Exercise 12.6.2. Prove Lemma 12.3.25, namely that a process X is a good integrator if and only if the set

$$\mathcal{J} = \{H \bullet X; H \in \Lambda, |H| \leq 1\}$$

is bounded in probability, that is, for any $\epsilon > 1$, there exists $k > 0$ such that $\sup_{J \in \mathcal{J}} P(|J| > k) < \epsilon$.

Exercise 12.6.3. Let W be a Brownian motion, and suppose $X_t = W_t(\omega^*)$ for some fixed $\omega^* \in \Omega$. Show that, for almost all choices of ω^*, the deterministic path X is not a good integrator, and explain why this does not contradict the fact that W is a good integrator.

Exercise 12.6.4. Let N be a Poisson process, X the associated martingale $X_t = N_t - \lambda t$ and $\{\mathcal{F}_t\}_{t \geq 0}$ the filtration generated by N. Using the result of Exercise 8.4.9 or otherwise, show that there exists no nontrivial martingale orthogonal to X in this space.

Exercise 12.6.5. For M a martingale with $M_0 = 0$, we can define the 'integral' $\int_{]0,t]} M_t dM_s = M_t \int_{]0,t]} dM_s = M_t^2$. Show that this is not, generally, a local martingale and explain why.

Exercise 12.6.6. For W a Brownian motion and H a predictable process with $H \neq 0$, show that $H \bullet W$ cannot have finite variation.

Exercise 12.6.7. For W a Brownian motion and H a deterministic process in $L^2(W)$, show that $(H \bullet W)_t$ is normally distributed for every deterministic t and find its mean and variance.

Exercise 12.6.8. Let M be a martingale in a probability space with filtration $\{\mathcal{F}_t\}_{t \geq 0}$ and suppose $M_t - M_s$ is independent of $\mathcal{F}_s$ for any $t > s$. Let $\{\mathcal{G}_t\}_{t \geq 0}$ be the completed filtration generated by M, that is, $\mathcal{G}_t = \sigma(\{M_s\}_{s \leq t}) \vee \mathcal{N}$, where $\mathcal{N}$ are the null sets of $\mathcal{F}$. By Exercise 5.7.2, M is a $\{\mathcal{G}\}_{t \geq 0}$-martingale.

Let H be a $\{\mathcal{F}_t\}_{t \geq 0}$-predictable M-integrable process which admits an M-integrable $\{\mathcal{G}_t\}_{t \geq 0}$-predictable projection $\hat{H}$. Show that, for $t \geq 0$,

$$E[(H \bullet M)_t | \mathcal{G}_t] = \int_{[0,t]} \hat{H}_s dM_s \quad a.s.,$$

(Hint: First show $E[H_s | \mathcal{G}_t] = E[H_s | \mathcal{G}_s]$, then assume H is a simple process of the form implied by Corollary 7.2.5.)

Exercise 12.6.9. Let $\{X^n\}_{n\in\mathbb{N}}$ be a sequence of $\mathcal{H}^1$ martingales converging in $\mathcal{H}^1$. Show that the sequence converges in the semimartingale topology.

Exercise 12.6.10. For W a Brownian motion, show that the process $X_t = (1/W_t)I_{\{W_t\neq 0\}}$ is not a semimartingale.

Exercise 12.6.11. Prove Stricker's Theorem: Let X be a semimartingale with respect to the filtration $\{\mathcal{F}_t\}_{t\geq 0}$. Let $\{\mathcal{G}_t\}_{t\geq 0}$ be a subfiltration of $\{\mathcal{F}_t\}_{t\geq 0}$ (i.e. a filtration with $\mathcal{G}_t \subseteq \mathcal{F}_t$ for all t). Given that X is adapted to $\{\mathcal{G}_t\}_{t\geq 0}$, show that X is a semimartingale with respect to $\{\mathcal{G}_t\}_{t\geq 0}$.

Exercise 12.6.12. A semimartingale X is called a σ-martingale if there exists a predictable process H and a local martingale M such that $X = H\bullet M$, in the sense of the semimartingale integral. Consider the following example, due to Émery [76].

Let T and S be independent exponential random variables with parameter $\lambda = 1$.

(i) Show that the process M defined by

$$M_t = I_{\{t\geq T\}} - I_{\{t\geq S\}} - \int_{T\wedge S\wedge t}^{(T\vee S)\wedge t} (I_{\{T\geq S\}} - I_{\{T\geq S\}})ds$$

is a square integrable martingale (in its natural filtration).

(ii) Show that $H_t = 1/t$ is $|dM(\omega)|$-integrable for almost all ω, so $H\bullet M$ is well defined as a semimartingale integral.

(iii) Show that $E[|(H\bullet M)_{(T\wedge S\wedge t)}|] = \infty$ for any $t > 0$.

(iv) For any stopping time $R > 0$, show that R is constant on the set $R < T\wedge S$, and hence that $H\bullet M$ is not locally integrable (and so is *not* a local martingale).

Exercise 12.6.13. Let X be a σ-martingale, as defined in the previous question. Show that there exists a countable family $\{D_n\}_{n\in\mathbb{N}} \subset \Sigma_p$ with $\cup_n D_n = \Omega \times [0,\infty[$, such that $I_{D_n}\bullet X$ is a martingale for each n. (This is the origin of the term 'σ-martingale'.)

13

Random Measures

When dealing with jump processes, it is sometimes useful to have a theory of integration which distinguishes between jumps of different sizes. Particularly for processes with many jumps, this is most easily accomplished by treating the jump process as generating a 'random measure', that is a stochastic measure over time *and the sizes of the jumps*, such that the integrals with respect to this measure correspond, in some sense, to the stochastic integrals with respect to the original process. Formalizing this idea, in a general setting, is the purpose of this chapter.

To illustrate and motivate some of the ideas of the general situation (and, indeed, some concepts presented in earlier chapters), the first section of the chapter discusses, in some detail, a very basic stochastic process which has just one random jump, in a general space. Random measures are associated with such a process in an elementary way, and related martingales can be considered. This section is based on the work of Chou and Meyer [33], Davis [46], Elliott [67, 69] and Jacod [105]. We will then discuss general random measures following Jacod [107].

13.1 The Single Jump Process

In this section, we shall consider a process $\{X_t\}_{t\geq 0}$ which takes its values in a Blackwell[1] space $(\mathcal{Z}, \mathfrak{Z})$ and which remains at its initial point $z_0 \in \mathcal{Z}$ until a random time $T(\omega)$, when it jumps to a new random position $z(\omega)$. The underlying probability space can be taken to be

[1]We use the term Blackwell space in the sense of Dellacherie and Meyer [54], as discussed in Section 2.6. The important facts which we need are that the σ-algebra is separable, that is, it is generated by a countable algebra, and one can define regular conditional distributions. In applications, our space is usually $\mathbb{R}^d$ or $\overline{\mathbb{R}}^d$, or possibly a Polish space with its Borel σ-algebra.

S.N. Cohen, R.J. Elliott, *Stochastic Calculus and Applications*,
Probability and Its Applications, DOI 10.1007/978-1-4939-2867-5_13

$$\Omega = [0, \infty] \times \mathcal{Z},$$

with the σ-algebra $\mathcal{F} = \mathcal{B} \otimes \mathfrak{Z}$. (As usual, $\mathcal{B}$ denotes the Borel σ-algebra on $[0, \infty[$). A sample path of the process is

$$X_t(\omega) = \begin{cases} z_0 & \text{if} \quad t < T(\omega), \\ z(\omega) & \text{if} \quad t \geq T(\omega). \end{cases}$$

Suppose a probability measure P is given on $(\Omega, \mathcal{B} \otimes \mathfrak{Z})$. To prevent confusion in notation, we will write $\hat{P}$ when we think of this as a measure on $[0, \infty] \times \mathcal{Z}$, and P when we think of it as a measure on the 'abstract' space Ω. That is, we write

$$P(T \geq t, z \in A) = \hat{P}([t, \infty] \times A).$$

For convenience, assume that

$$\hat{P}([0, \infty] \times \{z_0\}) = 0 = \hat{P}(\{0\} \times \mathcal{Z}),$$

so that the probabilities of a zero sized jump and a jump at time zero are zero (Fig. 13.1).

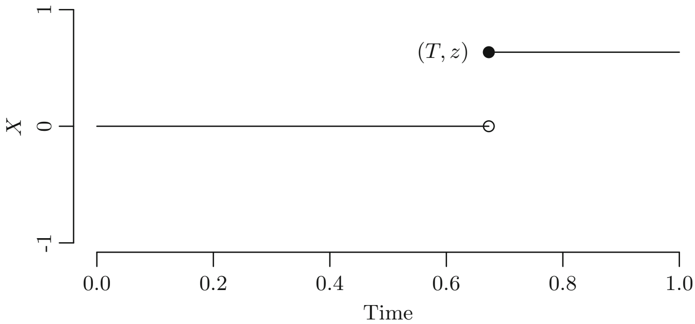

Fig. 13.1. A path of the single jump process X, with $z_0 = 0$.

Write $\{\mathcal{F}_t\}_{t \geq 0}$ for the completed σ-algebra generated by $\{X_s\}_{s \leq t}$. Note that $]t, \infty] \times \mathcal{Z}$ is an atom in $\mathcal{F}_t$ whenever $P(T > t) > 0$.

For $A \in \mathfrak{Z}$ write

$$F_t^A := \hat{P}(]t, \infty] \times A),$$

so that F_t^A is the probability that $T > t$ and $z \in A$. Furthermore, write

$$F_t := F_t^{\mathcal{Z}}$$

and

$$c := \inf\{t : F_t = 0\}.$$

Lemma 13.1.1. *Suppose τ is an $\{\mathcal{F}_t\}_{t\geq 0}$ stopping time. Then there is a $t_0 \in [0,\infty]$ such that $\tau \wedge T = t_0 \wedge T$ a.s.*

Proof. As observed above, $]t,\infty] \times \mathcal{Z}$ is an atom in $\mathcal{F}_t$ for every t with $P(T > t) > 0$. Suppose τ takes two values $t_1 \neq t_2$ on $\{\tau \leq T\}$ with positive probability (or values in disjoint neighbourhoods of $t_1 \neq t_2$). Then, for $t \in]t_1, t_2[$,

$$\{\tau \leq t\} \cap (]t,\infty] \times \mathcal{Z}) \subsetneq]t,\infty] \times \mathcal{Z},$$

so $\{\tau \leq t\} \notin \mathcal{F}_t$, contradicting the assumption that τ is a stopping time. Therefore, for some $t_0 \in [0,\infty]$, we know $\tau = t_0$ on $\{\tau \leq T\}$, as desired. □

Remark 13.1.2. The deterministic function F_t is right continuous and monotonic decreasing, so there are only countably many points of discontinuity

$$D := \{u : \Delta F_u = F_u - F_{u-} \neq 0\}.$$

Any constant time is a predictable stopping time, so each time u where $\Delta F_u \neq 0$ is predictable. By Lemma 13.1.1, the only predictable stopping times which can equal T are deterministic. It follows that, in the notation of Theorem 6.2.9, T_D is the accessible part of T. Clearly F is continuous if and only if T is totally inaccessible.

The Stieltjes measure on $([0,\infty], \mathcal{B})$ generated by F^A is absolutely continuous with respect to that generated by F, so there is a Radon-Nikodym derivative $\lambda(A,s)$ such that

$$F_t^A - F_0^A = \int_{]0,t[} \lambda(A,s) dF_s\,.$$

Remark 13.1.3. As $(\mathcal{Z}, \mathfrak{Z})$ is a Blackwell space, $\lambda(\cdot, s)$ can be constructed so as to be a regular conditional probability measure on $(\mathcal{Z}, \mathfrak{Z})$; that is, for each s, $\lambda(\cdot, s)$ is a probability measure on $\mathcal{Z}$ and, for each $A \in \mathfrak{Z}$, $\lambda(A, \cdot)$ is a measurable function of time (cf. Section 2.6, Theorem 13.3.7).

Definition 13.1.4. *The pair (λ, Λ) is called the* Lévy system *for the jump process, where*

$$\Lambda(t) = -\int_{]0,t]} \frac{dF_s}{F_{s-}}\,.$$

We also define the process $\tilde{\Lambda}(t) = \Lambda(t \wedge T)$.

Remark 13.1.5. Roughly speaking, $d\tilde{\Lambda}(t)$ is the probability that the jump occurs in the interval $]t, t+dt]$, given that it has not happened before time t.

For $A \in \mathfrak{Z}$, write

$$\begin{aligned}
\mu(\omega, t, A) &= I_{\{t \geq T\}} I_{\{z \in A\}},\\
\mu_p(\omega, t, A) &= -\int_{]0,t\wedge T]} \lambda(A,s) \frac{dF_s}{F_{s-}} = \int_{]0,t]} \lambda(A,s) d\tilde{\Lambda}(s),
\end{aligned}$$

and note that, for fixed A, these are both nonnegative processes. As usual, we will often omit the ω for ease of notation. Clearly $t \mapsto t \wedge T$ is continuous and $t \mapsto \int_{]0,t]} \lambda(A,s) d\tilde{\Lambda}(s)$ is a Borel-measurable function, so for any $A \in \mathfrak{Z}$ the finite variation process $\mu_p(\cdot, A)$ is predictable. In fact $\mu_p(\cdot, A)$ is the dual predictable projection of $\mu(\cdot, A)$, as the following result shows (cf. Corollary 8.2.12).

Theorem 13.1.6. *For any $A \in \mathfrak{Z}$, the process*

$$\tilde{\mu}(\cdot, A) := \mu(\cdot, A) - \mu_p(\cdot, A)$$

is an $\{\mathcal{F}_t\}_{t\geq 0}$-martingale.

Proof. For $t > s$, as $E[I_{\{T\geq u\}}|\mathcal{F}_s] = I_{\{T>s\}} F_{u-}/F_s$ for all $u > s$, we have

$$\begin{aligned}
E[\mu(t,A) - \mu(s,A)|\mathcal{F}_s] &= I_{\{T>s\}} \frac{F_s^A - F_t^A}{F_s}, \\
E[\mu_p(t,A) - \mu_p(s,A)|\mathcal{F}_s] &= I_{\{T>s\}} E\Big[\int_{]s,t]} \lambda(A,u) I_{\{T\geq u\}} \frac{dF_u}{F_{u-}} \Big| \mathcal{F}_s\Big] \\
&= -I_{\{T>s\}} \int_{]s,t]} \lambda(A,u) \frac{F_{u-}}{F_s} \frac{dF_u}{F_{u-}} \\
&= -I_{\{T>s\}} \frac{1}{F_s} \int_{]s,t]} \lambda(A,u) dF_u \\
&= -I_{\{T>s\}} \frac{F_t^A - F_s^A}{F_s}.
\end{aligned}$$

Therefore,

$$E[\tilde{\mu}(t,A) - \tilde{\mu}(s,A)|\mathcal{F}_s] = 0 \text{ a.s.},$$

and for any $t > 0$, we know $E\big[|\tilde{\mu}(t,A)|\big] \leq 2E\big[|\mu(t,A)|\big] \leq 2$. □

Given this, we now seek a formula for the predictable quadratic variation of $\tilde{\mu}(\cdot, A)$.

Remark 13.1.7. The jump of $\tilde{\mu}(t,A)$ at a discontinuity u of F_t is

$$\Delta\tilde{\mu}(u,A) = I_{\{T=u\}} I_{\{z\in A\}} + \lambda(A,u) \, \frac{\Delta F_u}{F_{u-}} \, I_{\{T\geq u\}} \, .$$

However,

$$\begin{aligned}
E[I_{\{T=u\}} I_{\{z\in A\}}|\mathcal{F}_{u-}] &= E[I_{\{z\in A\}}|T=u] P(T=u|\mathcal{F}_{u-}) \\
&= -\lambda(A,u) \frac{\Delta F_u}{F_{u-}} I_{\{T\geq u\}} \, ,
\end{aligned}$$

so $E[\Delta\tilde{\mu}(u,A)|\mathcal{F}_{u-}] = 0$. Therefore, from Theorem 10.2.11, for fixed $u \in D$,

$$\tilde{\mu}^{\Delta u}(t,A) := \Delta\tilde{\mu}(u,A)I_{\{t\geq u\}} \tag{13.1}$$

is a square integrable martingale orthogonal to every square integrable martingale which is continuous at u. Furthermore, applying Theorem 10.2.11 again, the predicable quadratic variation of $\tilde{\mu}^{\Delta u}(\cdot,A)$ is

$$\begin{aligned}\langle\tilde{\mu}^{\Delta u}\rangle_t &= E[(\Delta\tilde{\mu}(u,A))^2|\mathcal{F}_{u-}]I_{\{t\geq u\}}\\ &= -\lambda(A,u)\frac{\Delta F_u}{F_{u-}}I_{\{T\geq u\}}I_{\{t\geq u\}} - \Big(\lambda(A,u)\frac{\Delta F_u}{F_{u-}}\Big)^2 I_{\{T\geq u\}}I_{\{t\geq u\}}.\end{aligned}$$

Theorem 13.1.8. *For any $A \in \mathfrak{Z}$, the predictable quadratic variation of $\tilde{\mu}(\cdot,A)$ is given by*

$$\langle\tilde{\mu}(\cdot,A)\rangle_t = \mu_p(t,A) - r(t,A),$$

where

$$r(t,A) = \sum_{0<u\leq(t\wedge T)}\Big(\lambda(A,u)\,\frac{\Delta F_u}{F_{u-}}\Big)^2.$$

Proof. Decompose F into the sum of its continuous part F^c and the sum of its jumps $F^d_t = \sum_{0<u\leq t}\Delta F_u$. In a similar way, we can decompose $\mu_p(t,A)$ as a sum of

$$\mu^c_p(t,A) = -\int_{]0,t\wedge T]}\lambda(A,s)\frac{dF^c_s}{F_{s-}}\,,$$

and

$$\mu^d_p(t,A) = -\sum_{0<u\leq t\wedge T}\lambda(A,u)\frac{\Delta F_u}{F_{u-}}\,.$$

As above, let $D = \{u : \Delta F_u \neq 0\}$ and write

$$\mu^d(t,A) = \sum_{0<u\leq t\wedge T} I_{\{T=u\}}I_{\{z\in A\}}I_{\{u\in D\}} \qquad \text{and} \qquad \mu^c = \mu - \mu^d.$$

Then we have the decomposition

$$\tilde{\mu}(t,A) = \tilde{\mu}^c(t,A) + \tilde{\mu}^d(t,A),$$

where $\tilde{\mu}^c = \mu^c - \mu^c_p$ and

$$\tilde{\mu}^d(t,A) = \mu^d(t,A) - \mu^d_p(t,A) = \sum_{\substack{u\in D\\ u\leq t}}\tilde{\mu}^{\Delta u}.$$

Note that $\tilde{\mu}^c(\cdot,A)$ is not a continuous martingale, rather it corresponds to the totally inaccessible part of T.

If $u \neq u'$ the martingale $\tilde{\mu}^{\Delta u}$ defined in (13.1) is orthogonal to $\tilde{\mu}^{\Delta u'}$, and $\tilde{\mu}^d(t, A)$ is orthogonal to $\tilde{\mu}^c(t, A)$. By Remark 11.2.5, as $\tilde{\mu}^c$ and $\tilde{\mu}^d$ also a.s. share no jumps,

$$\langle \tilde{\mu}(\cdot, A) \rangle = \langle \tilde{\mu}^c(\cdot, A) \rangle + \langle \tilde{\mu}^d(\cdot, A) \rangle$$

and

$$\langle \tilde{\mu}^d(\cdot, A) \rangle = \sum_{u \in D} \langle \tilde{\mu}^{\Delta u}(\cdot, A) \rangle = \mu_p^d(t, A) - r(t, A).$$

To find $\langle \tilde{\mu}^c(\cdot, A) \rangle$, as we know that $\tilde{\mu}^c$ is a pure jump martingale, we have

$$[\tilde{\mu}^c(\cdot, A)]_t = \sum_{s \leq t} (\Delta \tilde{\mu}^c(s, A))^2 = I_{\{t \geq T\}} I_{\{z \in A\}} I_{\{t \notin D\}} = \mu^c(t, A).$$

As in Theorem 13.1.6, one can directly verify that μ_p^c is a continuous process (and hence predictable) such that $\mu^c(t, A) - \mu_p^c(t, A)$ is a martingale, and therefore $\langle \tilde{\mu}^c(\cdot, A) \rangle_t = \mu_p^c(t, A)$.

Combining these results, we have

$$\langle \tilde{\mu}(\cdot, A) \rangle_t = \mu_p^c(t, A) + \mu_p^d(t, A) - r(t, A) = \mu_p(t, A) - r(t, A). \qquad \square$$

Remark 13.1.9. The above form of predictable quadratic variation, with second-order terms arising from the accessible part $D = \{u : \Delta F_u \neq 0\}$, is typical of what happens in more complicated situations (cf. Theorem 13.3.16).

Notation 13.1.10. Because $\mu(t, A)$ and $\mu_p(t, A)$ are countably additive with respect to A, we can define, for $g \in L^1(\Omega, \mathcal{B} \otimes \mathfrak{Z}, \mu_p)$, the Stieltjes integrals

$$\begin{aligned}
\int_\Omega g(s, \zeta) \mu(ds, d\zeta) &= g(T, z), \\
\int_\Omega g(s, \zeta) \mu_p(ds, d\zeta) &= \int_{]0,T]} \int_{\mathcal{Z}} g(s, \zeta) \lambda(s, d\zeta) d\Lambda(s) \\
&= \int_{]0,T] \times \mathcal{Z}} g(s, \zeta) \frac{\hat{P}(ds, d\zeta)}{F_{s-}}
\end{aligned}$$

and hence, writing $d\mu$ for $\mu(ds, d\zeta)$ etc.,

$$\int_\Omega g d\tilde{\mu} = \int_\Omega g d\mu - \int_\Omega g d\mu_p.$$

In this way, we have transformed our original process X into a measure $\mu(\omega, \cdot, \cdot) : [0, \infty] \times \mathcal{Z} \to \mathbb{R}$. The measure μ is a probability measure on $[0, \infty[\times \mathcal{Z}$ with $\mu(\omega, T(\omega), \{z(\omega)\}) = 1$. We can also write

$$X_t = z_0 + \int_{]0,t] \times \mathcal{Z}} (\zeta - z_0) \mu(dt, d\zeta)$$

and so we see that no information has been lost by representing X in this way. The function μ is a 'random measure' and provides a useful way of understanding the dynamics of X.

We will consider the following spaces of measurable functions:

$$L^1(\mu \times P) := \Big\{g : \Omega \to R : E\Big[\int_\Omega |g| d\mu\Big] = E\Big[|g(T,z)|\Big] < \infty\Big\},$$
$$L^1(\mu_p \times P) := \Big\{g : \Omega \to R : E\Big[\int_\Omega |g| d\mu_p\Big] < \infty\Big\}.$$

We will also write $L^1_{\text{loc}}(\mu \times P)$ for the set of measurable functions $g : \Omega \to R$ such that $I_{\{s \leq t\}} g(s,\zeta) \in L^1(\mu \times P)$ for all $t < c$, and similarly $L^1_{\text{loc}}(\mu_p \times P)$.

Note that, as μ and μ_p assign no weight to the set $]T,\infty] \times \mathcal{Z}$, by Lemma 13.1.1 this agrees with the usual stochastic notion of localization of a space (see Section 3.3).

Lemma 13.1.11. *Writing* $\|g\|_1$ *for the* L^1 *norm* $\|g\|_1 = E\big[|g(T,z)|\big]$ *of* $g \in L^1(P)$ *we have*

$$\|g\|_1 = E\Big[\int_\Omega |g| d\mu\Big] = E\Big[\int_\Omega |g| d\mu_p\Big]$$

and hence

$$L^1(\mu \times P) = L^1(\mu_p \times P) = L^1(\hat{P}).$$

Proof. Recalling that $\Omega = [0,\infty] \times \mathcal{Z}$, we have

$$\int_\Omega |g| d\mu = |g(T,z)|$$

so the first identity is immediate. Now

$$\int_\Omega |g| d\mu_p = \int_{]0,T] \times \mathcal{Z}} F_{s-}^{-1} |g(s,\zeta)| d\hat{P},$$

so

$$\begin{aligned} E\Big[\int_\Omega |g| d\mu_p\Big] &= -\int_{]0,\infty]} \int_{]0,t] \times \mathcal{Z}} F_{s-}^{-1} |g(s,\zeta)| d\hat{P}(s,\zeta) dF_t \\ &= \int_{]0,\infty] \times E} \Big(-\int_{[s,\infty]} dF_t\Big) F_{s-}^{-1} |g(s,\zeta)| d\hat{P} \\ &= \int_\Omega |g| d\hat{P} = \|g\|_1. \end{aligned}$$

The identity of $L^1(\mu \times P)$, $L^1(\mu_p \times P)$ and $L^1(\hat{P})$ follows. $\square$

Remark 13.1.12. It will be useful to see explicitly how localization in $L^1(\hat{P})$ behaves.

Suppose $g \in L^1_{\text{loc}}(\mu \times P)$. Then there is an increasing sequence of stopping times $\{T_k\}_{k\in\mathbb{N}}$ such that $\lim_k T_k = \infty$ a.s. and $I_{\{s<T_k\}}g \in L^1(\mu \times P)$ for each k. Let $\{t_k\}_{k\in\mathbb{N}}$ be the corresponding sequence, given by Lemma 13.1.1 such that $T_k \wedge T = t_k \wedge T$. Because $\lim_k T_k = \infty$ a.s. we see that $\lim t_k = c = \inf\{t : F_t = 0\}$ a.s.

Now

$$\int_\Omega I_{\{s<T_k\}}|g|d\mu = g(T,z)I_{T<T_k}\,.$$

However, $\{T_k > T\} = [0, t_k[\, \times \mathcal{Z}$, so

$$E\Big[\int_\Omega I_{\{s<T_k\}}|g|d\mu\Big] = \int_{[0,t_k[\times\mathcal{Z}} |g|d\hat{P}.$$

On the other hand, suppose $g \in L^1_{\text{loc}}(P)$ and recall $c = \inf\{t : F_t = 0\}$. Construct the following sequence of stopping times $\{T_k\}_{k\in\mathbb{N}}$:

(i) if $c = \infty$ put $T_k = k$,
(ii) if $c < \infty$ and $F_{c-} > 0$ put $T_k = \infty$ for all k,
(iii) if $c < \infty$ and $F_{c-} = 0$ consider a sequence $\{t_k\}_{k\in\mathbb{N}}$ such that $\lim_k t_k = c$ and put
$$T_k = kI_{\{T\le t_k\}} + t_k I_{\{T>t_k\}}\,.$$

Clearly $\lim_k T_k = \infty$ a.s. and $I_{\{s<T_k\}}g \in L^1(\mu \times P)$, so

$$L^1_{\text{loc}}(\mu \times P) = L^1_{\text{loc}}(\hat{P}).$$

Lemma 13.1.13. *Suppose M is a uniformly integrable $\{\mathcal{F}_t\}_{t\ge 0}$-martingale such that $M_0 = 0$ a.s. Then there is an $\mathcal{F}_{\infty-} = \bigvee_{t\ge0}\mathcal{F}_t$ measurable function $h : \Omega \to \mathbb{R}$ such that $h \in L^1(\mu \times P)$ and*

$$M_t = h(T,z)I_{\{t\ge T\}} - I_{\{t<T\}}\frac{1}{F_t}\int_{]0,t]\times\mathcal{Z}} h(s,\zeta)d\hat{P}(s,\zeta) \quad \text{a.s.}$$

Proof. If M is a uniformly integrable martingale, then, from Corollary 5.2.4,

$$M_t = E[h|\mathcal{F}_t] \quad \text{a.s.}$$

for some $\mathcal{F}$-measurable random variable h. From the definition of $(\Omega, \mathcal{F})$ and the Doob–Dynkin lemma (Lemma 1.3.12), h is of the form $h(T,z)$. However,

$$E[h(T,z)|\mathcal{F}_t] = h(T,z)I_{\{t\ge T\}} + I_{\{t<T\}}\frac{1}{F_t}\int_{]t,\infty]\times\mathcal{Z}} h(s,\zeta)d\hat{P}(s,\zeta).$$

Because $M_0 = E[h] = \int_\Omega h d\hat{P} = 0$ the result follows by rearrangement. $\square$

Lemma 13.1.14. *Suppose M is a local martingale of $\{\mathcal{F}_t\}_{t\geq 0}$, and recall $c = \inf\{t : F_t = 0\}$.*

(i) If $c = \infty$, or $c < \infty$ and $F_{c-} = 0$, then M is a martingale on $[0, c[$.
(ii) If $c < \infty$ and $F_{c-} > 0$ then M is a uniformly integrable martingale.

Proof. Let $\{T_k\}_{k\in\mathbb{N}}$ be an increasing sequence of stopping times such that $\lim_k T_k = \infty$ a.s. and M^{T_k} is a uniformly integrable martingale for each k. We now separate into the two cases.

Case (i). Either we know that M is uniformly integrable, or for each k we have $P(T_k < T) > 0$. Then by Lemma 13.1.1 there is a sequence $\{t_k\}_{k\in\mathbb{N}}$ such that $T_k \wedge T = t_k \wedge T$ for each k, and because $P(T > t_k) > 0$ we have $t_k < c$. As $\lim_k T_k = \infty$, we see that $\lim_k P(T > t_k) = 0$, so $\lim_k t_k = c$. Now M is stopped at time T, so $M_{t\wedge T_k} = M_{t\wedge t_k}$. Consequently $\{M_t\}_{t\leq t_k}$ is a uniformly integrable martingale, and M is certainly a martingale on $[0, c[$.

Case (ii). Suppose now that $c < \infty$ and $F_{c-} > 0$. Hence $T \leq c$ a.s. and $P(T = c) > 0$. By Lemma 13.1.1, we again know there is a sequence $\{t_k\}_{k\in\mathbb{N}}$ such that $T \wedge t_k = T \wedge T_k$. As $\lim_k T_k = \infty$ a.s., looking at the set $\{\omega : T = c\}$ we see that $t_k \geq c$ for some k. Consequently, for such a k, $T_k \geq T$ a.s. and the process $M = M^{T_k}$ is a uniformly integrable martingale. □

We now obtain a 'martingale representation theorem', which allows us to write any local martingale in terms of the random measure $\tilde{\mu}$. A related result was considered in Exercise 8.4.9; this will be generalized in Section 14.5.

Theorem 13.1.15. *An adapted process M is a local $\{\mathcal{F}_t\}_{t\geq 0}$-martingale with* $M_0 = 0$ a.s. *if and only if $M = M^g$ for some $g \in L^1_{\text{loc}}(\mu \times P)$, where*

$$M^g_t := \int_{]0,t]\times\mathcal{Z}} g(s, \zeta)\tilde{\mu}(ds, d\zeta).$$

Furthermore, this function g is unique in $L^1_{\text{loc}}(\mu \times P)$.

Proof. Suppose $g \in L^1_{\text{loc}}(\mu \times P)$. If $\{T_k\}_{k\in\mathbb{N}}$ is the sequence of stopping times introduced in Remark 13.1.12, calculations similar to those in the proof of Theorem 13.1.6 show that $M^g_{t\wedge T_k}$ is a uniformly integrable $\{\mathcal{F}_t\}_{t\geq 0}$-martingale. (Alternatively, Theorem 13.1.6 shows directly that the statement holds for functions of the form $g(t, \zeta) = I_{\{t\leq a\}} I_{\{\zeta\in A\}}$, and then a monotone class argument gives the result.)

Conversely, suppose that M is uniformly integrable. Lemma 13.1.13 implies that

$$M_t = h(T, z)I_{\{t\geq T\}} - I_{\{t<T\}} \frac{1}{F_t} \int_{]0,t]\times\mathcal{Z}} h(s, \zeta) d\hat{P}(s, \zeta), \qquad (13.2)$$

where $h(T, z) = M_\infty$.

Now, for any $g \in L^1_{\text{loc}}(\mu \times P)$,

$$M_t^g = \int_{]0,t]\times\mathcal{Z}} g\,d\tilde{\mu} = I_{\{t\geq T\}}g(T,z) - \int_{]0,t\wedge T]\times\mathcal{Z}} g(s,\zeta)\frac{1}{F_{s-}}d\hat{P}.$$

Define, with h as in (13.2),

$$g(t,\zeta) = h(t,\zeta) + \frac{1}{F_t}\int_{]0,t]\times\mathcal{Z}} h(s,\zeta')d\hat{P}(s,\zeta'),$$

so Fubini's theorem yields

$$\begin{aligned}
&\int_{]0,t]\times\mathcal{Z}} g(s,\zeta)\frac{1}{F_{s-}}d\hat{P} \\
&= \int_{]0,t]\times\mathcal{Z}} h(s,\zeta)\frac{1}{F_{s-}}d\hat{P}(s,\zeta) - \int_{]0,t]}\Big(\frac{1}{F_sF_{s-}}\int_{]0,s]\times\mathcal{Z}} h(u,\zeta)\hat{P}(du,d\zeta)\Big)dF_s \\
&= \int_{]0,t]\times\mathcal{Z}} \frac{h}{F_{s-}}d\hat{P} + \int_{]0,t]\times\mathcal{Z}}\Big(-\int_{[u,t]}\frac{1}{F_sF_{s-}}dF_s\Big)h(u,\zeta)d\hat{P} \\
&= \int_{]0,t]\times\mathcal{Z}} \frac{h}{F_{s-}}d\hat{P} + \int_{]0,t]\times\mathcal{Z}}\Big(\frac{1}{F_t} - \frac{1}{F_{u-}}\Big)h(u,\zeta)d\hat{P} \\
&= \frac{1}{F_t}\int_{]0,t]\times\mathcal{Z}} h\,d\hat{P}.
\end{aligned}$$

Therefore $M^g = M$, as desired.

Now suppose M is a local martingale. By Lemma 13.1.14, if $c < \infty$ and $F_{c-} > 0$, then M is uniformly integrable, and so the result is proven. Otherwise, M is a martingale on $[0,c[$, and therefore is uniformly integrable on $[0,c-\epsilon]$ for any $\epsilon > 0$, and so by pasting together the functions obtained above, we can find a g such that $M_t^g = M_t$ a.s. for all $t < c$. If $c = \infty$ this is sufficient, otherwise we also know that $F_{c-} = 0$, that is, $P(T < c) = 1$. As M is stopped at T, this implies that $M_t^g = M_t$ almost surely for all t.

It remains to show that $g \in L^1_{\mathrm{loc}}(\mu \times P)$ and is unique. If $E\big[|M_t|\big] < \infty$, then our construction gives

$$\begin{aligned}
\int_{]0,t]\times\mathcal{Z}} |g|d\hat{P} &\leq \int_{]0,t]\times\mathcal{Z}} |h|d\hat{P} - \int_{]0,t]}\Big(\frac{1}{F_s}\int_{]0,s]\times\mathcal{Z}}|h|d\hat{P}\Big)dF_s \\
&\leq \int_{]0,t]\times\mathcal{Z}} |h|d\mu - F_t^{-1}\int_{]0,t]}\int_{]0,s]\times\mathcal{Z}}|h|d\hat{P}dF_s \\
&= \int_{]0,t]\times\mathcal{Z}} |h|d\mu + F_t^{-1}\int_{]0,t]\times\mathcal{Z}}(F_s - F_t)|h|d\hat{P} \\
&\leq \Big(1 + \frac{1}{F_t}\Big)\int_{]0,t]\times\mathcal{Z}}|h|d\hat{P} \\
&= \Big(1 + \frac{1}{F_t}\Big)E\big[|M_T|I_{\{T\leq t\}}\big].
\end{aligned}$$

As $E\big[|M_T|I_{\{T\leq t\}}\big] \leq E\big[|M_t|\big] < \infty$, this implies that $g(s,\zeta)I_{\{s<t\}} \in L^1(\mu\times P)$ for each t such that $F_t > 0$. For $\{T_k\}_{k\in\mathbb{N}}$ the sequence of stopping times constructed in Remark 13.1.12, we deduce $g(s,\zeta)I_{\{s<T_k\}} \in L^1(\mu \times P)$, and hence $g \in L^1_{\text{loc}}(\mu \times P)$. □

Remark 13.1.16. With the notation of Remark 13.1.12,

$$\tilde{\mu}(t,A) = \tilde{\mu}^c(t,A) + \sum_{u \in D} \tilde{\mu}^{\Delta u},$$

so, for $g \in L^1(\mu)$,

$$M_t^g = \int_{]0,\infty]\times\mathcal{Z}} I_{\{s\le t\}} g \, d\tilde{\mu}^c + \sum_{u\in D} \int_{]0,\infty]\times\mathcal{Z}} I_{\{s\le t\}} g \, d\tilde{\mu}^{\Delta u}.$$

For example, if $u \in D = \{u : \Delta F_u \neq 0\}$, then

$$\begin{aligned}\int_{]0,\infty]\times\mathcal{Z}} I_{\{s\le t\}} g \, d\tilde{\mu}^{\Delta u} &= I_{\{t\ge u\}} I_{\{T=u\}} g(u,z) \\ &\quad + I_{\{T\wedge t\ge u\}} \int_{\mathcal{Z}} g(u,\zeta)\lambda(d\zeta,u) F_{u-}^{-1} \Delta F_u\end{aligned}$$

and

$$E[g(u,\zeta)I_{\{T=u\}}|\mathcal{F}_{u-}] = -\Big(\int_{\mathcal{Z}} g(u,\zeta)\lambda(d\zeta,u)\Big) F_{u-}^{-1} \Delta F_u \, I_{\{T\ge u\}}.$$

The $\tilde{\mu}^c$ and $\tilde{\mu}^{\Delta u}$ martingales are orthogonal and either direct calculation, or Theorem 13.1.8 together with a monotone class argument, establishes the following result.

Theorem 13.1.17. *The predictable quadratic variation of M^g is given by*

$$\langle M^g \rangle_t = \int_{]0,t]\times\mathcal{Z}} g^2 \, d\mu_p - \sum_{\substack{0<u\le t\wedge T \\ u\in D}} \Big(\int_E g(u,\zeta)\lambda(d\zeta,u)\Big)^2 \frac{\Delta F_u^2}{F_{u-}^2}.$$

Remark 13.1.18. The maps $\mu(ds,d\zeta)$, $\mu_p(ds,d\zeta)$ and $\tilde{\mu}(ds,d\zeta)$ are all examples of random measures. In the terminology of the coming section, μ_p is the dual predictable projection, or compensator, of μ. The results above are taken from the work of Chou and Meyer [33], Davis [46] and Elliott [69, 68]. The results are extended to a sequence of jumps with a single time accumulation point in [46], and to general right constant processes with values in $\mathcal{Z}$ in [67]. Jacod, discussing multivariate point processes, obtained related theorems in [105]. Further results for the single jump process are given in Chapter 20, where the optimal control of jump processes is discussed.

13.2 General Random Measures

Having motivated some of the ideas by discussing the basic example of a single jump process, we now define a general random measure. Jacod was the first to discuss such measures in a detailed and rigorous manner, and our presentation is based on his work. For further results, see [107] or [110].

Convention 13.2.1. As before, we suppose we are working on a probability space $(\Omega, \mathcal{F}, P)$ which has a complete, right continuous filtration $\{\mathcal{F}_t\}_{t\geq 0}$. We also have an auxiliary Blackwell space $(\mathcal{Z}, \mathfrak{Z})$; however, such generality will not usually be required as the applications we have in mind are when $\mathcal{Z} \subseteq \mathbb{R}^n$.

Notation 13.2.2. Write

$$\begin{aligned}\tilde{\Omega} &= \Omega \times [0,\infty] \times \mathcal{Z},\\ \tilde{\mathcal{F}} &= \mathcal{F} \otimes \mathcal{B}([0,\infty]) \otimes \mathfrak{Z},\\ \tilde{\Sigma}_x &= \Sigma_x \otimes \mathcal{Z}, \qquad \text{for } x = o, p.\end{aligned}$$

We shall write $\mathcal{L}^0(\tilde{\Sigma}_x)$ for the space of $\tilde{\Sigma}_x$ measurable process.

Definition 13.2.3. *A* nonnegative random measure μ *is an $\mathcal{F}$-measurable family $\{\mu(\omega,\cdot)\}_{\omega\in\Omega}$ of σ-finite measures on $\big([0,\infty] \times \mathcal{Z}, \mathcal{B}([0,\infty]) \otimes \mathfrak{Z}\big)$. A function which can be written as the difference of two nonnegative random measures is called a* random measure.

Remark 13.2.4. Philosophically, if μ is a random measure, then we think of $\mu(\omega,\cdot)$ as a signed measure on $[0,\infty] \times \mathcal{Z}$. However, we do *not* assume that $\mu(\omega,\cdot)$ can take only one of the values $+\infty$ and $-\infty$.

As we assume that $\mu(\omega,\cdot)$ is the difference of σ-finite measures, this does not pose too many problems. If $\mu = \mu_1 - \mu_2$, and $\{B_n\}_{n\in\mathbb{N}}$ is a family of sets such that $[0,\infty] \times \mathcal{Z} = \cup_n B_n$ and $\mu_1(\omega, B_n) + \mu_2(\omega, B_n) < \infty$, then the map $B \mapsto \mu(\omega, B \cap B_n)$ is a signed measure, and we can write

$$\mu(\omega, B) = \sum_n \mu(\omega, B \cap B_n)$$

for any $B \in \mathcal{B}([0,\infty]) \otimes \mathfrak{Z}$ such that at least one of the sums $\sum_n \big(\mu(\omega, B \cap B_n)\big)^+$ and $\sum_n \big(\mu(\omega, B \cap B_n)\big)^-$ is convergent. Nevertheless, one should be aware that there is no guarantee that $\mu(\omega, [0,\infty] \times \mathcal{Z})$ is well defined.

Example 13.2.5. The processes μ, μ_p and $\tilde{\mu}$ constructed in the first part of this chapter are all random measures, while μ and μ_p are also nonnegative random measures. More generally, a random measure can be associated with any measurable process A of locally finite variation by taking $\mathcal{Z}$ to be a one point set $\{\Delta\}$ and defining

$$\mu(\omega, dt \times \{\Delta\}) = dA_t(\omega).$$

Example 13.2.6. Another classic example of a random measure is when N is a Poisson process, we define $\mathcal{Z} = \{1\}$ and then take

$$\mu(\omega, dt \times \{1\}) = dN_t.$$

Just as in the single jump case, this is naturally paired with the predictable measure defined by

$$\mu_p(\omega, dt \times \{1\}) = \lambda dt,$$

and, given our interest in martingales (via Theorem 5.5.18), we will naturally be interested in their difference $\tilde{\mu} = \mu - \mu_p$. This example also demonstrates why we do not assume our measures to be finite, as $\mu(\omega, [0, \infty] \times \{1\}) = \infty$ almost surely and $\mu_p(\omega, [0, \infty] \times \{1\}) = \infty$ surely, so $\tilde{\mu}(\omega, \cdot)$ is not a signed measure and $\tilde{\mu}(\omega, [0, \infty] \times \{1\})$ is not defined. (In some sense, this is because the martingale $N_t - \lambda t$ has unbounded total variation on $[0, \infty]$.)

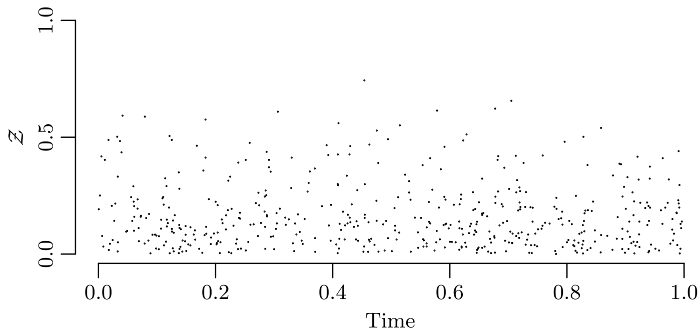

Fig. 13.2. A 'path' of a random measure μ, with $\mathcal{Z} = [0, 1]$. Each point shown corresponds to a unit point-mass (i.e. a point with $\mu\big(\{(t, z)\}\big) = 1$), and these are the only sets charged by the measure (i.e. it is an integer valued random measure, in the terminology of Section 13.3). The compensator (cf. Theorem 13.2.21) of this example is $\mu_p(dt, dz) = 500 \times 5z^4 dz\, dt$ (i.e. points are independently Beta(5,1) distributed in space, and occur at a total average rate of 500/unit time).

Remark 13.2.7. For each ω, as $\mu(\omega, \cdot)$ is the difference of σ-finite measures, μ has a Jordan–Hahn decomposition (see Section 1.7)

$$\mu(\omega, \cdot) = \mu^+(\omega, \cdot) - \mu^-(\omega, \cdot)$$

and an absolute value $|\mu|(\omega, \cdot) = \mu^+(\omega, \cdot) + \mu^-(\omega, \cdot)$. One can verify that all of these are also measurable with respect to ω, and so are also random measures.

Notation 13.2.8. Suppose $W : \tilde{\Omega} \to \mathbb{R}$ is a map such that each section $W(\omega, \cdot) : [0, \infty] \times \mathcal{Z} \to \mathbb{R}$ is a Borel measurable function. Then write

$$(W * \mu)_t(\omega) = \int_{[0,t]\times\mathcal{Z}} W(\omega, s, \zeta)\mu(\omega, ds, d\zeta),$$

if this integral exists (and may equal $+\infty$ or $-\infty$).

If $(W * \mu)_t$ exists for all $t \in [0, \infty[$ and is measurable with respect to ω, one can talk of the *process* $W * \mu$. This process has jumps

$$\Delta(W * \mu)_t = \int_{\mathcal{Z}} W(\omega, t, \zeta)\mu(\omega, \{t\} \times d\zeta).$$

The random measure $W \cdot \mu$ is defined by

$$(W \cdot \mu)(\omega, dt, d\zeta) = W(\omega, t, \zeta)\mu(\omega, dt, d\zeta).$$

Clearly, for each ω, the Radon–Nikodym derivative of $W \cdot \mu$ with respect to μ is $W(\omega, \cdot)$, and, if the appropriate integrals exist,

$$W' * (W \cdot \mu) = (W'W) * \mu. \tag{13.3}$$

If H is a $W * \mu$-integrable process, then a simple approximation argument shows that

$$H \bullet (W * \mu) = (HW) * \mu = W * (H \cdot \mu).$$

Definition 13.2.9. *The random measure μ is called* optional *(resp.* predictable*) if, for each positive process $W \in \mathcal{L}^0(\tilde{\Sigma}_o)$ (resp. $\mathcal{L}^0(\tilde{\Sigma}_p)$), the processes $W * \mu^+$ and $W * \mu^-$ are optional (resp. predictable).*

Remark 13.2.10. If μ is optional (resp. predictable), then clearly the random measures μ^+, μ^- and $|\mu|$ are optional (resp. predictable). Furthermore, for each $W \in \mathcal{L}^0(\tilde{\Sigma}_o)$ (resp. $\mathcal{L}^0(\tilde{\Sigma}_p)$) for which the process $W * \mu$ exists,

$$W * \mu = W^+ * \mu^+ + W^- * \mu^- - W^+ * \mu^- - W^- * \mu^+,$$

and so $W * \mu$ is then an optional (resp. predictable) process.

Lemma 13.2.11. *Suppose μ is an optional (resp. predictable) random measure and $W \in \mathcal{L}^0(\tilde{\Sigma}_o)$ (resp. $\mathcal{L}^0(\tilde{\Sigma}_p)$). Then, if the random measure $W \cdot \mu$ is defined, it is optional (resp. predictable).*

Proof. Clearly

$$(W \cdot \mu)^+ = W^+ \cdot \mu^+ + W^- \cdot \mu^-,$$
$$(W \cdot \mu)^- = W^- \cdot \mu^+ + W^+ \cdot \mu^-,$$

and, as observed above, for any $W' \in \mathcal{L}^0(\tilde{\Sigma}_o)$ (resp. $W' \in \mathcal{L}^0(\tilde{\Sigma}_p)$) we have

$$W' * (W \cdot \mu) = (W'W) * \mu.$$

As $W'W \in \mathcal{L}^0(\tilde{\Sigma}_o)$ (resp. $\mathcal{L}^0(\tilde{\Sigma}_p)$), the result follows. $\square$

We now consider spaces of random measures which generalize the spaces $\mathcal{A}$, $\mathcal{A}_{\mathrm{loc}}$ and $\mathcal{V}$. Again, we should consider the equivalence classes of random measures which are P-almost surely equal, but we use the same notation, μ, for an equivalence class and a member of the equivalence class, that is, a *version* of μ. A random measure is then *optional* if its equivalence class contains an optional member.

It is worth noting that we do not need to speak of 'indistinguishable' random measures, as a random measure is defined on the entire space $[0,\infty] \times \mathcal{Z}$. Therefore, if two random measures μ, μ' are almost surely equal, then the set where they differ does not depend on t. Consequently, for any measurable W, the processes $W * \mu$ and $W * \mu'$ are indistinguishable.

Definition 13.2.12. *A random measure μ is said to be* integrable *if, with* 1 *denoting the process which is identically* 1,

$$E[1 * |\mu|_\infty] < \infty.$$

Notation 13.2.13. Define the following sets of (equivalence classes of) random measures

- $\tilde{\mathcal{A}}$ will denote the set of optional and integrable random measures.
- $\tilde{\mathcal{V}}$ will denote the set of random measures μ for which there exists a $\tilde{\Sigma}_o$-measurable partition $\{B_n\}_{n\in\mathbb{N}}$ of $\tilde{\Omega}$ such that $I_{B_n} \cdot \mu \in \tilde{\mathcal{A}}$ for each n.
- $\tilde{\mathcal{A}}_\sigma$ will denote the set of random measures μ for which there exists a $\tilde{\Sigma}_p$-measurable partition $\{B_n\}_{n\in\mathbb{N}}$ of $\tilde{\Omega}$ such that $I_{B_n} \cdot \mu \in \tilde{\mathcal{A}}$ for each n.
- $\tilde{\mathcal{A}}^+$ (resp. $\tilde{\mathcal{A}}^+_\sigma$, $\tilde{\mathcal{V}}^+$) will denote the positive measures in $\tilde{\mathcal{A}}$ (resp. $\tilde{\mathcal{A}}_\sigma$, $\tilde{\mathcal{V}}$).

The measures in $\tilde{\mathcal{A}}_\sigma$ are the 'predictably σ-integrable' random measures, while the measures in $\tilde{\mathcal{V}}$ are 'optionally σ-integrable'. Note, however, that a measure in $\tilde{\mathcal{A}}_\sigma$ is not necessarily predictable. Clearly each measure in $\tilde{\mathcal{V}}$ is optional and $\tilde{\mathcal{A}} \subset \tilde{\mathcal{A}}_\sigma \subset \tilde{\mathcal{V}}$.

Remark 13.2.14. It is easy to see from the definition that $\mu \in \tilde{\mathcal{A}}$ if and only if $1 * \mu \in \mathcal{A}$.

The process $1 * \mu \in \mathcal{V}$ (resp. $\mathcal{A}_{\mathrm{loc}}$) if and only if $\mu \in \tilde{\mathcal{V}}$ (resp. $\tilde{\mathcal{A}}_\sigma$) and there is an increasing sequence $\{T_n\}_{n\in\mathbb{N}}$ of stopping times, with $\lim_n T_n = \infty$ a.s., such that $I_{D_n} \cdot \mu \in \tilde{\mathcal{A}}$ (or equivalently, $I_{D_n} * \mu \in \mathcal{A}$) where $D_n = [\![0, T_n[\![$ (resp. $[\![0, T_n]\!]$). When $1 * \mu \in \mathcal{V}$, one can take for $\{T_n\}_{n\in\mathbb{N}}$ the sequence of stopping times which localizes $1 * \mu$.

Example 13.2.15. Suppose that $\mathcal{Z} = \{1/n : n \in \mathbb{N}\}$, and that for each n we have a Poisson process $N^{(n)}$ with parameter n. Define the random measure μ by

$$\mu(\omega, [0,t] \times \{1/n\}) = N^{(n)}_t.$$

We then have

$$E\Big[\int_{[0,t]\times\mathcal{Z}} I_{\{\zeta=1/n\}}\mu(ds,d\zeta)\Big] = nt$$

so, for any $t \geq 0$,

$$E[(1*\mu)_t] = E\Big[\int_{[0,t]\times\mathcal{Z}} 1\mu(ds,d\zeta)\Big] = t + 2t + 3t + \ldots = \infty.$$

Therefore, we see that $\mu \notin \tilde{\mathcal{A}}$ (and this holds even if we localize in time). However, the sets $\{[nk, n(k+1)] \times \{1/n\}\}_{k,n\in\mathbb{N}}$ give a predictable (indeed, deterministic) decomposition of $[0,\infty] \times \mathcal{Z}$ with

$$E\Big[\int_{[0,\infty]\times\mathcal{Z}} I_{[nk,n(k+1)]\times\{1/n\}}\mu(ds,d\zeta)\Big] = n^2 < \infty,$$

and so $\mu \in \tilde{\mathcal{A}}_\sigma$.

Definition 13.2.16. *Suppose $\mu \in \tilde{\mathcal{V}}$. Let $\{B_n\}_{n\in\mathbb{N}}$ be a $\tilde{\Sigma}_o$-measurable partition of $\tilde{\Omega}$ such that $I_{B_n} \cdot \mu \in \tilde{\mathcal{A}}$ for each n, and let W be a bounded $\tilde{\mathcal{F}}$-measurable process. For $B \in \tilde{\Sigma}_o$, write*

$$M_\mu(B) = \sum_n M_\mu(B \cap B_n) = \sum_n E\big[I_B\big(I_{B_n} * \mu\big)_\infty\big].$$

This defines a σ-finite measure M_μ on $(\tilde{\Omega}, \tilde{\Sigma}_o)$ which is independent of the partition $\{B_n\}$. M_μ is called the Doléans measure *associated with μ.*

Remark 13.2.17. If μ is associated with $A \in \mathcal{A}_{\text{loc}}$ as in Example 13.2.5, then

$$M_\mu(B) = E\Big[\int_{[0,\infty]} I_{\{(\omega,t,\Delta)\in B\}} dA\Big].$$

We now give a general result which allows us to construct a random measure given its Doléans measure (cf. Theorem 8.3.2).

Theorem 13.2.18. *Suppose m is a measure on $(\tilde{\Omega}, \tilde{\mathcal{F}})$. There is an optional random measure $\mu \in \tilde{\mathcal{V}}$ (resp. a predictable random measure $\mu \in \tilde{\mathcal{A}}_\sigma$), such that $m = M_\mu$ if and only if, writing $x = o$ (resp. $x = p$),*

(i) m is $\tilde{\Sigma}_x$-σ-finite,
(ii) $m(N \times \mathcal{Z}) = 0$ for every P-evanescent set N,
(iii) for every $A \in \tilde{\Sigma}_x$ such that $|m|(A) < \infty$, and every bounded $\mathcal{F}$-measurable process X,

$$\int_{\tilde{\Omega}} I_A X dm = \int_{\tilde{\Omega}} I_A \Pi_x(X) dm.$$

The associated random measure μ is then unique. Furthermore $\mu \in \tilde{\mathcal{A}}$ if and only if m is finite, and is positive if and only if m is positive.

Proof. We prove the optional case, the predicable case follows in the same way. First suppose $\mu \in \tilde{\mathcal{V}}$, and $m = M_\mu$. Properties (i) and (ii) are immediate. If $A \in \tilde{\Sigma}_o$ is such that $|m|(A) < \infty$, then the processes $\alpha^+ = I_A * \mu^+$ and $\alpha^- = I_A * \mu^-$ are in $\mathcal{A}$. For any bounded $\mathcal{B} \otimes \mathcal{F}$-measurable process X,

$$\begin{aligned}\int_{\tilde{\Omega}} I_A X dm &= E[XI_A * \mu^+_\infty] - E[XI_A * \mu^-_\infty] \\ &= E\Big[\int_{[0,\infty]} X\, d\alpha^+\Big] - E\Big[\int_{[0,\infty]} X\, d\alpha^-\Big].\end{aligned}$$

Property (iii) then follows from Theorem 8.2.5. The necessity of conditions (i – iii) is, therefore, established.

To show the converse, we proceed in a sequence of steps.

Step 1. We demonstrate that $\mu \in \tilde{\mathcal{V}}$ is determined by its Doléans measure M_μ. Suppose $\{B_n\}_{n\in\mathbb{N}}$ is a $\tilde{\Sigma}_o$-measurable partition of $\tilde{\Omega}$ such that $I_{B_n} \cdot \mu \in \tilde{\mathcal{A}}$. It is sufficient to show that each measure $I_{B_n} \cdot \mu$ is determined by its Doléans measure $I_{B_n} \cdot M_\mu$; that is, it is sufficient to prove the result when $\mu \in \tilde{\mathcal{A}}$.

In this case, because $\mathfrak{Z}$ is separable (as $(\mathcal{Z}, \mathfrak{Z})$ is a Blackwell space), the measure $\mu(\omega, \cdot)$ is determined by the values of $\mu(\omega, [0,t] \times D)$, for t a positive rational number, and $D \in \mathfrak{Z}$ a member of a countable family which generates $\mathfrak{Z}$. However, for any set $G \in \mathcal{F}$,

$$M_\mu(G \times [0,t] \times D) = E\big[I_G\big(\mu([0,t] \times D)\big)\big].$$

Therefore, P-almost surely, the measure M_μ determines each random variable $\mu([0,t] \times D)$, and so, P-almost surely, the random measure μ itself.

Step 2. We now show that conditions (i – iii) are sufficient to show that $m = M_\mu$ for some $\mu \in \tilde{\mathcal{V}}$.

Step 2a. Initially we suppose that, m is a positive, finite measure and that, for every set $H \in \tilde{\Sigma}_o$ and every bounded measurable process X,

$$\int_{\tilde{\Omega}} I_H \Pi_o(X) dm = \int_{\tilde{\Omega}} I_H X dm.$$

Write $\hat{m}$ for the measure on $([0,\infty] \times \Omega, \mathcal{B} \otimes \mathcal{F})$ defined by

$$\hat{m}(B \times G) = m(B \times G \times \mathcal{Z}).$$

Then $\hat{m}$ satisfies the conditions of Theorem 8.1.17, and so there is an optional process $A \in \mathcal{A}^+$ such that $\hat{m} = \mu_A$. Because $(\mathcal{Z}, \mathfrak{Z})$ is a Blackwell space, by taking Radon–Nikodym derivatives, one can factorize the measure m with respect to $\hat{m}$, that is, there is a family of regular random probability measures $\{n(t,\omega,\cdot)\}$ on $(\mathcal{Z}, \mathfrak{Z})$, which is Σ_x-measurable with respect to (t,ω), such that on $(\tilde{\Omega}, \tilde{\Sigma}_x)$,

$$m(dt, d\omega, dz) = \widehat{m}(dt, d\omega) n(t, \omega, dz).$$

Therefore,

$$\mu(\omega, dt, dz) = dA_t(\omega) n(\omega, t, dz)$$

is a random measure in $\tilde{\mathcal{A}}^+$. Suppose X is a positive, bounded measurable process and $D \in \mathcal{Z}$. Then

$$\int_{\tilde{\Omega}} X I_D dM_\mu = E\big[\big((X\, n(D)) \bullet A\big)_\infty\big] = \int_{[0,\infty]\times\Omega} \big(X\, n(D)\big) d\hat{m}.$$

The process $n(D)$ is optional, so from Theorem 8.2.2 and Corollary 7.6.7 we observe that, on $\tilde{\Sigma}_x$,

$$\int_{\tilde{\Omega}} X I_D dM_\mu = \int_{[0,\infty]\times\Omega} \big(\Pi_o(X) n(D)\big) d\hat{m} = \int_{\tilde{\Omega}} \big(\Pi_o(X) I_D\big) dm = \int_{\tilde{\Omega}} X I_D dm.$$

By a monotone class argument, the measures m and M_μ are seen to be equal. (In the predictable case, μ is also seen to be a predictable random measure.)

Step 2b. We now suppose m is now a signed finite measure satisfying conditions (i – iii) of the theorem, and that $m = m^+ - m^-$ is its Jordan–Hahn decomposition.

Define finite positive measures n^+, n^- on $(\tilde{\Omega}, \tilde{\mathcal{F}})$ by putting, for every $D \in \mathcal{Z}$ and positive bounded measurable process X,

$$\int_{\tilde{\Omega}} (X I_D) dn^+ = \int_{\tilde{\Omega}} \big(\Pi_o(X) I_D\big) dm^+,$$
$$\int_{\tilde{\Omega}} (X I_D) dn^- = \int_{\tilde{\Omega}} \big(\Pi_o(X) I_D\big) dm^-.$$

With $n := n^+ - n^-$, we have

$$\int_{\tilde{\Omega}} (X I_D) dn = \int_{\tilde{\Omega}} \big(\Pi_o(X) I_D\big) dm,$$

so $n = m$ by Theorem 8.2.2 and property (iii).

If $N \in \mathcal{F}$ is a P-evanescent set, then $\Pi_o(I_N)$ is indistinguishable from the zero process, so from their definition and property (iii) we have

$$n^+(N \times \mathcal{Z}) = m^+(N \times \mathcal{Z}) = 0 = m^+(N \times \mathcal{Z}) = n^-(N \times \mathcal{Z}).$$

Consequently, the measures n^+ and n^- satisfy condition (iii) of the theorem (which was not guaranteed for the initial decomposition $m^+ - m^-$). Therefore, by Step 2a there are random measures $\nu, \eta \in \tilde{\mathcal{A}}^+$ such that $M_\nu = n^+$ and $M_\eta = n^-$. The measure $\mu = \nu - \eta$ then belongs to $\tilde{\mathcal{A}}$ and satisfies $M_\mu = m$.

Step 2c. Suppose m is now a general measure satisfying conditions (i – iii) of the theorem. Let $\{B_n\}_{n\in\mathbb{N}}$ be a Σ_o-measurable partition of $\tilde{\Omega}$ such that m restricted to each B_n is finite. If $m_n = I_{B_n} \cdot m$, then m_n is a finite measure satisfying the conditions of Step 2b. Consequently, for each n there is a random measure $\mu_n \in \tilde{\mathcal{A}}$ such that $M_{\mu_n} = m_n$. Because m_n is concentrated on B_n, we know $m_n(B_n^c) = 0$ and $I_{B_n} \cdot \mu_n = \mu_n$, so the random measure $\mu = \sum_n \mu_n$ is in $\tilde{\mathcal{V}}$. By construction, $M_\mu = m$.

Step 3. Clearly, $\mu \in \tilde{\mathcal{A}}$ if and only if M_μ is finite. From the construction of μ in Step 2 and the uniqueness established in Step 1 we see $\mu \geq 0$ if and only if $M_\mu \geq 0$. □

Corollary 13.2.19. *Suppose μ and ν are two random measures in $\tilde{\mathcal{V}}$ (resp. $\tilde{\mathcal{A}}_\sigma$). If the measures M_μ, M_ν are equal on $\tilde{\Sigma}_o$ (resp. $\tilde{\Sigma}_p$), then $\mu = \nu$.*

We also observe from the theorem the following result.

Corollary 13.2.20. *The spaces $\tilde{\mathcal{A}}$, $\tilde{\mathcal{A}}_\sigma$ and $\tilde{\mathcal{V}}$ of random measures (resp. the predictable random measures in each of these spaces) form vector spaces of random measures (or more formally, vector spaces of equivalence classes of random measures).*

We now construct the 'dual predictable projections' or 'compensators' of random measures. As in the case of finite variation processes, this will allow us to construct and describe those random measures which correspond to martingales.

Theorem 13.2.21. *Suppose $\mu \in \tilde{\mathcal{A}}_\sigma$. There is a unique predictable random measure $\mu_p \in \tilde{\mathcal{A}}_\sigma$ which satisfies the following equivalent conditions:*

(i) the measures M_μ and M_{μ_p} coincide on $\tilde{\Sigma}_p$,
(ii) for every function $W \in \mathcal{L}^0(\tilde{\Sigma}_p)$ such that $W * \mu \in \mathcal{A}_{\text{loc}}$ we have

$$\Pi_p^*(W * \mu) = W * \mu_p.$$

Proof. Suppose $\{B_n\}_{n\in\mathbb{N}}$ is a $\tilde{\Sigma}_p$-measurable partition of $\tilde{\Omega}$ such that $I_{B_n} \cdot \mu \in \tilde{\mathcal{A}}$ for each n. Define a $\tilde{\Sigma}_p$-σ-finite measure m on $(\tilde{\Omega}, \tilde{\Sigma}_p)$ by putting, for each $D \in \mathfrak{Z}$ and each bounded measurable process X,

$$\int_{\tilde{\Omega}} (X I_D I_{B_n}) dm = \int_{\tilde{\Omega}} \big(\Pi_p^*(X) I_D I_{B_n}\big) dM_\mu.$$

Then m satisfies the conditions of Theorem 13.2.18, and

$$\int_{\tilde{\Omega}} I_A X dm = \int_{\tilde{\Omega}} \big(I_A \Pi_p^*(X)\big) dm$$

for every $A \in \tilde{\Sigma}_p$. Therefore, there is a predictable random measure $\mu_p \in \tilde{\mathcal{A}}_\sigma$ such that $m = M_{\mu_p}$. Clearly M_μ and M_{μ_p} are equal on $\tilde{\Sigma}_p$, so we have established (i).

We now prove that (i) implies (ii). Suppose μ_p is a predictable random measure in $\tilde{\mathcal{A}}_\sigma$ and $W \in \mathcal{L}^0(\tilde{\Sigma}_p)$ is such that $W * \mu \in \mathcal{A}_{\text{loc}}$. Then

$$E\big[\big(X\bullet(W*\mu)\big)_\infty\big] = \int_{\tilde{\Omega}} XW dM_\mu$$

and

$$E\big[\big(X\bullet(W*\mu_p)\big)_\infty\big] = \int_{\tilde{\Omega}} XW dM_{\mu_p}$$

for every bounded measurable process X. By (i), the measures M_μ and M_{μ_p} coincide on Σ_p, so if X is also predictable, $XW \in \mathcal{L}^0(\tilde{\Sigma}_p)$ and

$$E\big[(X\bullet(W*\mu))_\infty\big] = \int_{\tilde{\Omega}} XW dM_\mu = \int_{\tilde{\Omega}} XW dM_{\mu_p} = E\big[(X\bullet(W*\mu_p))_\infty\big].$$

By Theorem 8.2.6, as $W*\mu_p$ is predictable, we see that $\Pi_p^*(W*\mu) = W*\mu_p$, that is, μ_p satisfies (ii).

Finally, we prove that (ii) implies (i). With $\{B_n\}_{n\in\mathbb{N}}$ as defined at the start of the proof, for every $A \in \tilde{\Sigma}_p$ the process $I_{A\cap B_n} * \mu$ is in $\mathcal{A} \subset \mathcal{A}_{\text{loc}}$. Therefore, from (ii),

$$\Pi_p^*(I_{A\cap B_n} * \mu) = (I_{A\cap B_n}) * \mu_p.$$

Consequently, by Theorem 8.2.6,

$$M_\mu(A\cap B_n) = E[(I_{A\cap B_n} * \mu)_\infty] = E\big[(I_{A\cap B_n} * \mu_p)_\infty\big] = M_{\mu_p}(A\cap B_n)$$

and so μ_p satisfies (i).

□

Definition 13.2.22. *The measure μ_p is called the* dual predictable projection *or* compensator *of μ. By connection with the finite-variation processes, we may write $\mu_p = \Pi_p^*(\mu)$.*

Corollary 13.2.23. *Suppose $\mu \in \tilde{\mathcal{A}}_\sigma$ and $W \in \mathcal{L}^0(\tilde{\Sigma}_p)$ are such that $W*\mu \in \mathcal{A}_\sigma$. Then $\Pi_p^*(W*\mu) = W*\Pi_p^*(\mu)$. In particular, if $\mu \in \tilde{\mathcal{A}}_\sigma$ is predictable, then $\Pi_p^*(\mu) = \mu$.*

Remark 13.2.24. From Theorem 10.2.11 and the characterization of the dual predictable projection in Corollary 8.2.12, we see that, if T is a predictable stopping time and $A \in \mathcal{A}_{\text{loc}}$, then

$$\Delta\Pi_p^*(A)_T = E[\Delta A_T | \mathcal{F}_{T-}] \quad \text{on} \quad \{T<\infty\}.$$

Therefore, if $W \in \tilde{\Sigma}_p$ is such that

$$\int_{\mathcal{Z}} W(T,\zeta)\mu(\{T\}\times d\zeta) I_{\{T<\infty\}}$$

exists, it follows that

$$\int_{\mathcal{Z}} W(T,\zeta)\Big(\Pi_p^*(\mu)(\{T\}\times d\zeta)\Big) = E\Big[\int_{\mathcal{Z}} W(T,\zeta)\mu(\{T\}\times d\zeta)\Big|\mathcal{F}_{T-}\Big], \quad (13.4)$$

on $\{T<\infty\}$, if the conditional expectation exists.

Definition 13.2.25. *A random measure $\mu \in \tilde{\mathcal{A}}_\sigma$ is said to be a* martingale random measure *if the restriction of M_μ to $\tilde{\Sigma}_p$ is zero.*

Remark 13.2.26. We see that, if $\mu \in \tilde{\mathcal{A}}_\sigma$, then μ_p is the unique predictable measure in $\tilde{A}_\sigma$ such that $\mu - \mu_p$ is a martingale random measure.

Theorem 13.2.27. *The following are equivalent:*

(i) μ is a martingale random measure,
(ii) the compensator of μ is $\Pi_p^*(\mu) \equiv 0$,
(iii) for all $W \in \mathcal{L}^0(\tilde{\Sigma}_p)$ such that $W * \mu \in \mathcal{A}_{\mathrm{loc}}$, the process $W * \mu$ is a local martingale.

Proof. By Theorem 13.2.21, we know that a random measure μ is a martingale random measure if and only if $\Pi_p^*(\mu)$ is zero on $\tilde{\Sigma}_p$, and as $\Pi_p^*(\mu)$ is predictable, this implies it is zero everywhere. Equivalently, this states that $\Pi_p^*(W * \mu)$ is indistinguishable from zero for every $W \in \mathcal{L}^0(\tilde{\Sigma}_p)$ such that $W * \mu \in \mathcal{A}_{\mathrm{loc}}$. By Theorem 8.2.11 this is equivalent to stating that $W * \mu$ is a local martingale. $\square$

13.3 Integer Valued Random Measures

We now consider a particular class of random measures, which correspond in a more precise way to jump processes.

Definition 13.3.1. *A random measure μ is said to be* integer valued *if*

(i) $\mu(\omega, \{t\} \times \mathcal{Z}) \leq 1$ for all (ω, t),
(ii) for every $A \in \mathcal{B}([0,\infty]) \otimes \mathfrak{Z}$, the random variable $\mu(\omega, A)$ takes values in the set $\overline{\mathbb{Z}}^+ = \{0, 1, 2, \dots\} \cup \{\infty\}$.

Note that μ is then a nonnegative random measure.

Notation 13.3.2. We will denote by $\tilde{\mathcal{V}}^1$ the members of $\tilde{\mathcal{V}}$ which have a representation which is an integer valued random measure. Similarly, we will write $\tilde{\mathcal{A}}_\sigma^1 = \tilde{\mathcal{V}}^1 \cap \tilde{\mathcal{A}}_\sigma$ and $\tilde{\mathcal{A}}^1 = \tilde{\mathcal{V}}^1 \cap \tilde{\mathcal{A}}$. When discussing elements in $\tilde{\mathcal{V}}^1$ we always consider a representative which has integer values.

Remark 13.3.3. If μ is an integer valued random measure, we write

$$D = \{(\omega, t) : \mu(\omega, \{t\} \times \mathcal{Z}) = 1\}.$$

From Definition 13.3.1(ii), if $(\omega, t) \in D$ there is a unique point $z_t(\omega) \in \mathcal{Z}$ such that

$$\mu(\omega, \{t\} \times d\zeta) = \delta_{z_t(\omega)}(d\zeta).$$

Here $\delta_{z_t(\omega)}(d\zeta)$ denotes the unit mass at $z_t(\omega)$.

If $(\omega, t) \notin D$, define $z_t(\omega) = \eta$, where η is an additional point not in $\mathcal{Z}$. Then we have a process $z = \{z_t\}_{t\geq 0}$ with values in $\mathcal{Z} \cup \{\eta\}$ and the measure μ can be written

$$\begin{aligned}\mu(\omega, dt, d\zeta) &= \sum_{(\omega,s)\in D} \delta_{(s,z_s(\omega))}(dt, d\zeta) \\ &= \sum_{s\geq 0} I_{\{z_s\in\mathcal{Z}\}}\delta_{(s,z_s(\omega))}(dt, d\zeta).\end{aligned} \tag{13.5}$$

Because μ is σ-finite, each section of D,

$$D_\omega = \{t \geq 0 : (\omega, t) \in D\}$$

is at most countable. If μ is integrable, then its sections are a.s. finite. If $\mu \in \tilde{\mathcal{V}}^1$, then the set D is optional and so, in the terminology of Definition 7.5.1, D is thin. (This follows by applying Theorem 7.3.17 for $\mu \in \tilde{\mathcal{A}}^1$, and then extending to $\mu \in \tilde{\mathcal{V}}^1$, see Exercise 13.7.3.)

Conversely, if z is a process with values in $\mathcal{Z} \cup \{\eta\}$ such that the set $\{(\omega, t) : z_t(\omega) \in \mathcal{Z}\}$ is thin, then the above expression (13.5) defines a random measure μ. The process z is called a *point process*, the 'points' being the set

$$\big\{\big(t, z_t(\omega)\big) : t \geq 0, z_t(\omega) \in \mathcal{Z}\big\} \subset [0,\infty] \times \mathcal{Z}.$$

Lemma 13.3.4. *Suppose μ is an integer valued random measure given by (13.5) for some process z. Then $\mu \in \tilde{\mathcal{V}}$ if and only if z is an optional process, in which case D is an optional thin set.*

Proof. Suppose $\mu \in \tilde{\mathcal{V}}$. Let $\{B_n\}_{n\in\mathbb{N}}$ be a $\tilde{\Sigma}_o$-measurable partition of $\tilde{\Omega}$ such that $M_\mu(B_n) < \infty$ for each n. Then, for every $C \subseteq \mathcal{Z}$,

$$\{(\omega, t) : z_t(\omega) \in C\} = \bigcup_n \{(\omega, t) : \Delta(I_C I_{B_n} * \mu)_t = 1\} \in \Sigma_o.$$

Consequently, D is an optional set and z is an optional process.

Conversely, suppose z is an optional process. We know

$$D = \big\{(\omega, t) : \mu(\omega, \{t\} \times \mathcal{Z}) = 1\big\} = \{(\omega, t) : z_t(\omega) \in \mathcal{Z}\}$$

and so D is an optional set. As μ is integer valued and σ-finite, we know that D is thin. Therefore, there exists a sequence $\{T_n\}_{n\in\mathbb{N}}$ of stopping times with disjoint graphs such that $D = \bigcup_n [\![T_n]\!]$. If W is a nonnegative $\tilde{\Sigma}_o$-measurable process, then

$$W * \mu = \sum_n W(T_n, z_{T_n}) I_{[\![T_n,\infty]\!]}$$

is optional, so μ is optional. Furthermore, writing $B_n = [\![T_n]\!] \times \mathcal{Z}$ and $B = \tilde{\Omega}\backslash(\bigcup_n B_n)$, the measures $I_B \cdot \mu$ and $I_{B_n} \cdot \mu$ are of finite total mass, and so $\mu \in \tilde{\mathcal{V}}$.

□

Example 13.3.5. Suppose $\mathcal{Z}$ is a normed space and X is a process with left-limits in $\mathcal{Z}$. Write

$$\mu(dt, d\zeta) = \sum_{s>0} I_{\{X_{s-} \neq X_s\}} \delta_{(s,X_s)}(dt, d\zeta).$$

Then μ is an integer valued random measure, with

$$D = \{(\omega, t) : X_{t-}(\omega) \neq X_t(\omega)\} \cap [\![0]\!]^c$$

and $z = X$ on D.

If X is optional, then, from Lemma 13.3.4, so is μ. In fact, we can define the following stopping times:

$$\begin{aligned} T(0,m) &= 0, \\ T(n+1,m) &= \inf\Big\{t > T(n,m) : \|X_{t-} - X_t\| \in \Big[\frac{1}{m}, \frac{1}{m-1}\Big]\Big\}. \end{aligned}$$

Then the measure M_μ is finite on each of the $\tilde{\Sigma}_p$-measurable sets

$$\Big\{(\omega, t, \zeta) : t \leq T(n,m)(\omega), \big\|X_{t-}(\omega) - \zeta\big\| \in \Big[\frac{1}{m}, \frac{1}{m-1}\Big[\Big\},$$

whose union is $\tilde{\Omega}$, so in fact $\mu \in \tilde{\mathcal{A}}^1_\sigma$.

Definition 13.3.6. *The compensator* $\mu_p = \Pi^*_p(\mu)$ *of the measure* μ *of this example is called the* Lévy system *of the process* X.

Theorem 13.3.7. *If* $\mu \in \tilde{\mathcal{A}}^1_\sigma$ *there is a version* μ_p *of* $\Pi^*_p(\mu)$ *such that, for all* ω*,*

$$\mu_p(\omega, \cdot) \geq 0 \quad \textit{and} \quad \mu_p(\omega, \{t\} \times \mathcal{Z}) \leq 1.$$

As in the single jump case, as $\mathcal{Z}$ *is a Blackwell space, there is a decomposition*

$$\mu_p(\omega, dt, d\zeta) = \lambda(\omega, t, d\zeta)\Lambda(\omega, dt),$$

where

- $\Lambda(\omega, \cdot)$ *is a measure on* $[0, \infty[$ *for all* ω*, and* $\Lambda(\omega, [0, t])$ *is a predictable process,*
- λ *is a* regular positive transition measure *from the predictable* σ*-algebra to* $(\mathcal{Z}, \mathfrak{Z})$*, that is,* $\lambda(\omega, t, \cdot)$ *is a (nonnegative) measure on* $(\mathcal{Z}, \mathfrak{Z})$ *for all* ω, t *and* $\lambda(\omega, t, A)$ *is predictable for all* $A \in \mathfrak{Z}$*.*

Proof. Recall that any integer valued random measure is nonnegative. Suppose ν is a nonnegative version of $\Pi^*_p(\mu)$. Write

$$a_t(\omega) = \nu(\omega, \{t\} \times \mathcal{Z}).$$

If $\{B_n\}_{n\in\mathbb{N}}$ is a $\tilde{\Sigma}_p$-measurable partition of $\tilde{\Omega}$ such that $I_{B_n} \cdot \mu \in \tilde{\mathcal{A}}^+$, then

$$a_t = \sum_n \Delta(I_{B_n} * \nu)_t,$$

so the process $\{a_t\}_{t\geq 0}$ is predictable. Therefore, we see that the random measure $\mu_p = I_{\{a\leq 1\}} \cdot \nu$ is a predictable positive random measure, which satisfies the conditions of the theorem and which is a version of $\Pi_p^*(\mu)$ if the set $\{a > 1\}$ is evanescent. However, from (13.4), for every predictable stopping time T, we know that $a_T \leq 1$ almost surely. Therefore, applying the section theorem (Theorem 7.3.17) to the predictable set $\{a > 1\}$ we see this set is evanescent.

Writing $\Lambda(\omega, [0,t]) = \mu_p(\omega, [0,t]\times\mathcal{Z})$ we see that $\mu_p(\omega, dt\times B)$ is absolutely continuous with respect to Λ for every $B \in \mathfrak{Z}$, so, as discussed in Section 2.6, we can construct a Radon–Nikodym derivative

$$\lambda(\omega, t, B) = \frac{d\mu_p(\omega, [0,t] \times B)}{d\Lambda(\omega, [0,t])}$$

which is a regular conditional probability measure on $(\mathcal{Z}, \mathfrak{Z})$. The decomposition follows directly. $\square$

Note that $\Lambda(\omega, \{t\}) = 0$ unless there is an accessible jump at (ω, t). As the set of all jumps D is thin, the sum $\sum_{s\in[0,t]} \Lambda(\omega, \{s\})$ is well defined.

Example 13.3.8. Suppose λ and Λ are deterministic measures over $\mathcal{Z}$ and $[0, \infty[$ respectively. For simplicity, assume $\Lambda(dt) = dt$. We seek to construct a random measure with compensator $\lambda(d\zeta)dt$.

As λ is σ-finite, there exists a partition $\{A_i\}_{i\in\mathbb{N}}$ of $\mathcal{Z}$ such that $\lambda(A_i)$ is finite. Without loss of generality, $\lambda(A_i) > 0$ for all i. For each i, we take a Poisson process with rate $\lambda(A_i)dt$, which defines a sequence of times $T_1^i, T_2^i, \ldots$. Independently, we also define a sequence of independent random variables $Z_1^i, Z_2^i, \ldots$, valued in A_i, with identical distribution given by $I_{\{\zeta\in A_i\}}\lambda(d\zeta)/\lambda(A_i)$. Using these random variables, we define the random measure μ^i by

$$\mu^i(d\zeta, dt) = \sum_k \delta_{\{\zeta=Z_k^i, t=T_k^i\}}.$$

By independence of the size of the jump and its timing, it is easy to check that the compensator of μ^i is given by $I_{\{\zeta\in A_i\}}\lambda(d\zeta)dt$. Finally $\mu = \sum_i \mu^i$ gives a random measure on $\mathcal{Z}$ with compensator $\lambda(d\zeta)dt$.

Note that the assumption $\Lambda(dt) = dt$ is not needed, provided one can construct a Poisson process with (deterministic) time varying rate (which can be done using a time-change argument, among other methods). Note also that, while the times $\{T_k^i\}_{k\in\mathbb{N}}$ are well ordered for each i, it is typically not the case that their union will also be well ordered.

13.3.1 Stochastic Integrals with Random Measures

So far, our approach to random measures has been fundamentally deterministic. We have defined the integral $W * \mu$ simply by fixing ω and calculating the integral with respect to the measure $\mu(\omega; \cdot)$. However, just as we generalized finite variation martingales to consider purely discontinuous martingales, so we will generalize our integral to incorporate the approach developed in Chapter 12. In the light of Example 13.3.5, we restrict our attention to those random measures in $\tilde{\mathcal{A}}^1_\sigma$ and their compensators.

Definition 13.3.9. *For $\mu \in \tilde{\mathcal{A}}^1_\sigma$, let $\tilde{\mu} = \mu - \mu_p$. We say W is stochastically integrable with respect to $\tilde{\mu}$ if*

$$\Big(\sum_{s \leq (\cdot)} \Big(\int_{\mathcal{Z}} W(\omega, s, \zeta) \tilde{\mu}(\omega, \{s\} \times d\zeta) \Big)^2 \Big)^{1/2} \in \mathcal{A}_{\mathrm{loc}}.$$

In this case, the stochastic integral of W with respect to $\tilde{\mu}$ is defined to be the purely discontinuous local martingale, denoted $W * \tilde{\mu}$, such that

$$\Delta(W * \tilde{\mu})_t = \int_{\mathcal{Z}} W(\omega, t, \zeta) \tilde{\mu}(\omega, \{t\} \times d\zeta)$$

up to indistinguishability.

Remark 13.3.10. Note that, as

(i) $W * \tilde{\mu}$ has at most countably many discontinuities,
(ii) each discontinuity occurs at a stopping time, and
(iii) all the discontinuities are integrable by the stochastic integrability assumption,

we can construct a process satisfying these requirements by application of Theorem 11.5.11, and it is unique by Exercise 11.7.12.

Remark 13.3.11. In the case where

$$\sum_{s \leq (\cdot)} \Big(\int_{\mathcal{Z}} W(\omega, s, \zeta) \tilde{\mu}(\omega, \{s\} \times d\zeta) \Big)^2 \in \mathcal{A}_{\mathrm{loc}},$$

we can also define the stochastic integral by decomposing $\tilde{\mu}$ into pieces in $\tilde{\mathcal{A}}^1$, taking the integral with respect to each of these (finite activity) pieces separately, and then recombining them by addition and a limit in $\mathcal{H}^2_{\mathrm{loc}}$.

We now seek to describe the quadratic variation of $W * \tilde{\mu}$. The approach we use is taken from Jacod and Yor [109], and gives an elegant way of working with these random measures. It depends on the following peculiar object, the 'conditional projection' under the Doléans measure. Recall that $\tilde{\Omega} = \Omega \times \mathcal{Z}$ and $\tilde{\Sigma}_p = \Sigma_p \otimes \mathfrak{Z}$, where Σ_p is the predictable σ-algebra. The key result is Theorem 13.3.16. The associated proofs may be omitted on a first reading.

Definition 13.3.12. *Let $\mu \in \tilde{\mathcal{A}}_\sigma$ have associated Doléans measure M_μ, and $X : \tilde{\Omega} \to \mathbb{R}$ be such that $|X| \cdot \mu \in \tilde{\mathcal{A}}_\sigma$. We define the conditional projection*

$$M_\mu[X|\tilde{\Sigma}_p] = \tilde{X}$$

where $\tilde{X} : \tilde{\Omega} \to \mathbb{R}$ is a predictable version of the Radon–Nikodym derivative $dM_{(X\cdot\mu)}/dM_\mu$, where both $M_{(X\cdot\mu)}$ and M_μ are restricted to $(\tilde{\Omega}, \tilde{\Sigma}_p)$.

Remark 13.3.13. For X an optional process, we can consider X to be a function of $\tilde{\Omega} = \Omega \times [0, \infty[\times \mathcal{Z}$ which is constant in $\zeta \in \mathcal{Z}$, so the projection is still well defined. Intuitively, for X an optional process, if $\tilde{X} := M_\mu[X|\tilde{\Sigma}_p]$, then $\tilde{X}$ is a $\tilde{\Sigma}_p = \Sigma_p \otimes \mathfrak{Z}$-measurable map satisfying

$$\tilde{X}(t, \zeta) = E\big[X_t \big| \mathcal{F}_{t-} \cap \{\mu(\{t\} \times \{\zeta\}) = 1\}\big].$$

That is, $\tilde{X}(t, \zeta)$ is the conditional expectation of the value of X_t, given $\mathcal{F}_{t-}$ and that $\mu(\{t\} \times \{\zeta\}) = 1$. In particular, if $X = W * (\mu - \mu_p)$, then we shall see that

$$M_\mu[X|\tilde{\Sigma}_p] = W - \int_{\mathcal{Z}} W(\zeta')\mu_p(\{t\} \times d\zeta').$$

Recall from Remark 13.3.3 that, for $\mu \in \tilde{\mathcal{A}}^1_\sigma$, we defined D to be the set of points (ω, t) where μ is nonzero, and z_T is the point such that $\mu\big(\{(T, z_T)\}\big) = 1$, if one exists.

Definition 13.3.14. *We write $\tilde{\mathcal{T}}$ for the set of stopping times T such that $[\![T]\!] \subset D$ and there exists a set $A \in \tilde{\Sigma}_p$ with $\mu\big((]\!]0, T[\![\times \mathcal{Z}) \cap A\big) = 0$ and $\mu\big(([\![T]\!] \times \mathcal{Z}) \cap A\big) = 1$ on $\{T < \infty\}$.*

The following result is taken from Jacod [106] and formalizes the intuition behind Remark 13.3.13.

Lemma 13.3.15. *Let μ and X be as in Definition 13.3.12.*

(i) If $T \in \tilde{\mathcal{T}}$, $T > 0$ and $X(T, z_T)$ is integrable, then

$$M_\mu[X|\tilde{\Sigma}_p](T, Z_T) = E[X(T, z_T)|\mathcal{F}_{T-} \vee \sigma(z_T)].$$

(ii) If $\{T_n\}_{n\in\mathbb{N}}$ is a sequence of elements in $\tilde{\mathcal{T}}$ such that $D = \cup_n [\![T_n]\!]$ up to evanescence and $X(T_n, z_{T_n})$ is integrable for each n, and if there is a $\tilde{\Sigma}_p$-measurable function V such that

$$V(T_n, z_{T_n}) = E[X(T_n, z_{T_n})|\mathcal{F}_{T_n-} \vee \sigma(z_{T_n})] \qquad \textit{for every } n,$$

then $V = M_\mu[X|\tilde{\Sigma}_p]$.

Proof. To show (i), let $V = M_\mu[X|\tilde{\Sigma}_p]$ and let A be the element of $\tilde{\Sigma}_p$ associated with T by the definition of $\tilde{\mathcal{T}}$. As $T > 0$, we know that $\mathcal{F}_{T-}$ is generated by the sets $\{B = B' \cap \{s < T\} : B' \in \mathcal{F}_s\}$. For any $C \in \mathfrak{Z}$, defining

$F = A \cap (B \times]s, \infty[\times C) \cap ([\![0, T]\!] \times \mathcal{Z})$, we have $F \in \tilde{\Sigma}_p$. On the other hand, from Theorem 7.6.5, it is clear that for any $\tilde{\Sigma}_p$-measurable W, we know that $W(T, z_T)$ is $\mathcal{F}_{T-} \vee \sigma(z_T)$ measurable. The result then follows from the equality

$$E[V(T, z_T) I_B I_C(z_T)] = M_\mu[V I_F] = M_\mu[X I_F] = E[X(T, z_T) I_B I_C(z_T)].$$

For (ii), it is enough to see that, for any nonnegative $\tilde{\Sigma}_p$-measurable function Y with $M_\mu\big[|XY|\big] < \infty$, we have

$$M_\mu[XY] = \sum_n E[(XY)(T_n, z_{T_n})] = \sum_n E[(VY)(T_n, z_{T_n})] = M_\mu[VY]. \qquad \square$$

Note that if $\{A_n\}_{n \in \mathbb{N}}$ is a partition of $\tilde{\Omega}$ with $M_\mu(A_n) < \infty$ for all n, then $T^{m,n} := \inf\big\{t : \mu\big(([0,t] \times \mathcal{Z}) \cap A_n\big) \geq m\big\}$ is a stopping time in $\tilde{\mathcal{T}}$ and $D = \cup_{n,m} [\![T^{m,n}]\!]$. This sequence (after reordering) will satisfy the conditions of (ii) above.

Theorem 13.3.16. *For $\mu \in \tilde{\mathcal{A}}^1_\sigma$, let $\tilde{\mu} = \mu - \mu_p$. Let W be stochastically integrable with respect to $\tilde{\mu}$ and $X = W * \tilde{\mu}$. Let N be a local martingale and $V = M_\mu[\Delta N | \tilde{\Sigma}_p]$. Then, if $[X, N] \in \mathcal{A}_{\mathrm{loc}}$, we have $\langle X, N\rangle = VW * \mu_p$.*

In particular, if W^2 is μ_p-integrable, then

$$\langle W * \tilde{\mu}\rangle_t = (W^2 * \mu_p)_t - \sum_{s \leq t} \Big(\int_{\mathcal{Z}} W(\zeta')\mu_p(\{s\} \times d\zeta')\Big)^2.$$

Proof. For notational simplicity, we write

$$\hat{W}_t = \int_{\mathcal{Z}} W(\zeta')\mu_p(\{t\} \times d\zeta').$$

Let $A = \langle X, N\rangle$. For any predictable stopping time T, we know that $T_D \in \tilde{\mathcal{T}}$, (where as usual, D is the support of μ, $T_D = T$ on D and $T_D = \infty$ on D^c). Therefore, if $\Delta N_T I_{\{T<\infty\}}$ is integrable, from Lemma 13.3.15 we have

$$V(T, z_T) = E[\Delta N_T | \mathcal{F}_T \vee \sigma(z_T)] \qquad \text{on } \{T < \infty\} \cap \{T \in D\}.$$

For any $\tilde{\Sigma}_p$-measurable function U such that $I_D(T)U(T, z_T)$ is integrable, the martingale property then yields

$$\begin{aligned} \hat{V}_T &= -E[\Delta N_T I_{D^c}(T) | \mathcal{F}_{T-}] \quad \text{on } \{T < \infty\}, \\ \hat{U}_T &= E[I_D(T) U(T, z_T) | \mathcal{F}_{T-}] \quad \text{on } \{T < \infty\}. \end{aligned}$$

If T is a predictable stopping time such that ΔN_T and $\Delta [X, N]_T$ are integrable on $\{T < \infty\}$, it follows that, P-a.s on $\{T < \infty\}$,

$$\begin{aligned}
\Delta A_T &= E\big[\Delta[X,N]_T\big|\mathcal{F}_{T-}\big] = E\big[\Delta M_T \Delta N_T\big|\mathcal{F}_{T-}\big] \\
&= E\big[(I_D(T)W(T,z_T) - \hat{W}_T)\Delta N_T\big|\mathcal{F}_{T-}\big] \\
&= E\big[(I_D(T)W(T,z_T) - \hat{W}_T)E[\Delta N_T|\mathcal{F}_{T-}\vee\sigma(z_T)]\big|\mathcal{F}_{T-}\big] \\
&\qquad - \hat{W}_T E[I_{D^c}(T)\Delta N_T|\mathcal{F}_{T-}] \\
&= E\big[(I_D(T)W(T,z_T) - \hat{W}_T)V(T,z_T)|\mathcal{F}_{T-}\big] + \hat{W}_T\hat{V}_T \\
&= (\widehat{VW})_T - \hat{V}_T\hat{W}_T + \hat{W}_T\hat{V}_T = (\widehat{VW})_T.
\end{aligned}$$

As A has only predictable jumps, the predicable section theorem (Theorem 7.3.17) implies that $\Delta A = (\widehat{VW})$ up to indistinguishability.

Now let $A = A^d + A^c$, where $A^d_t = \sum_{s\leq t}\Delta A$. As A is increasing, and $\widehat{VW} = 0$ on the set $\Delta A = 0$, we have just shown that

$$A^d = (VWI_{\{\Delta A>0\}}) * \nu. \tag{13.6}$$

If u is a predictable process such that $u\bullet A \in \mathcal{A}$, then

$$\begin{aligned}
E[(u\bullet A^c)_\infty] &= E\big[\big((uI_{\{\Delta A=0\}})\bullet A\big)_\infty\big] = E\Big[\sum_s I_{\{\Delta A_s=0\}}u_s\Delta M_s\Delta N_s\Big] \\
&= E\Big[\sum_s I_{\{\Delta A_s=0\}\cap D}u_s W(s,z_s)\Delta N_s\Big] \\
&= \int_{\tilde{\Omega}} uWI_{\{\Delta A=0\}}\Delta N dM_\mu = \int_{\tilde{\Omega}} uWI_{\{\Delta A_s=0\}}V dM_\mu \\
&= \int_{\tilde{\Omega}} uWVI_{\{\Delta A_s=0\}} dM_{\mu_p} = E\big[\big((uWVI_{\{\Delta A_s=0\}}) * \mu_p\big)_\infty\big],
\end{aligned}$$

which implies

$$A^c = (VWI_{\{\Delta A=0\}}) * \mu_p. \tag{13.7}$$

Combining (13.6) and (13.7), we have the desired result, in particular,

$$\langle W * \tilde{\mu}\rangle = W(W-\hat{W}) * \mu_p = W^2 * \mu_p - ((\hat{W}W) * \mu_p) = W^2 * \mu_p - \sum_{s\leq(\cdot)}(\hat{W}_s)^2.$$

□

The following corollary is of particular use when we have a quasi left-continuous filtration, so no martingale can jump at an accessible time (by Theorem 6.4.4).

Corollary 13.3.17. *For $\mu \in \tilde{\mathcal{A}}^1_\sigma$, suppose $\mu_p(\{t\}\times\mathcal{Z}) \equiv 0$ (so μ does not have any accessible jumps). Then, for any $\tilde{\mu}$-stochastically integrable W,*

$$\langle W * \tilde{\mu}\rangle = W^2 * \mu_p,$$

and, furthermore,

$$E\big[\sup_t |\Delta(W * \tilde{\mu})_t|^p\big] \leq E\big[(|W|^p * \mu_p)_\infty\big].$$

Proof. The first statement is simply a special case of Theorem 13.3.16. The second follows from the fact that

$$\begin{aligned} &E\big[\sup_t |\Delta(W * \tilde{\mu})_t|^p\big] \\ &= E\Big[\sup_t \Big|\int_{\mathcal{Z}} W\,\mu(\{t\}\times d\zeta)\Big|^p\Big] = E\Big[\sup_t \Big(\int_{\mathcal{Z}} |W|^{p/2}\mu(\{t\}\times d\zeta)\Big)^2\Big] \\ &\leq E\Big[\sum_t \Big(\int_{\mathcal{Z}} |W|^{p/2}\mu(\{t\}\times d\zeta)\Big)^2\Big] = E\big[[|W|^{p/2} * \tilde{\mu}]_\infty\big] \\ &= E\big[\langle |W|^{p/2} * \tilde{\mu}\rangle_\infty\big] = E\big[(|W|^p * \mu_p)_\infty\big]. \end{aligned}$$

□

13.4 Characteristics of Semimartingales

Using the theory of random measures, we now give a representation of a semimartingale in three predictable parts. These parts will describe the behaviour of the semimartingale, in particular the conditional distribution of its increments, in a useful way. To begin with, the following example gives a fundamental connection between random measures and càdlàg processes.

Example 13.4.1. Suppose X is a càdlàg process with values in $\mathcal{Z}$, but now suppose $\mathcal{Z}$ is a subset of an additive group, for example, $\mathcal{Z}$ might even be a vector space. Consider the following random measure:

$$\mu^X(dt, d\zeta) = \sum_{s>0} I_{\{\Delta X_s \neq 0\}} \delta_{(s,\Delta X_s)}(dt, d\zeta).$$

Here

$$D = \{(\omega, t) :\, X_{t-}(\omega) \neq X_t(\omega)\} \cap [\![0]\!]^c$$

and $z = \Delta X$ on D. As in Example 13.3.5, we can show that if X is adapted, then $\mu^X \in \tilde{\mathcal{A}}^1_\sigma$.

Remark 13.4.2. The random measure $\Pi_p^*(\mu^X)$ is also sometimes called the Lévy system of X (cf. Example 13.3.5), though care should be taken to indicate which measure is under discussion (and it is perhaps clearer to call it the Lévy system of the jumps of X).

Remark 13.4.3. If μ is as in Example 13.3.5, that is,

$$\mu(dt, d\zeta) = \sum_{s>0} I_{\{\Delta X_s \neq 0\}} \delta_{(s,X_s)}(dt, d\zeta),$$

then μ and μ^X are closely related. In particular, for $B \subset [0,\infty] \times \mathcal{Z}$ we have

$$\mu(B) = \int_{[0,\infty]\times\mathcal{Z}} I_B(s, X_{s-} + \zeta)\mu^X(ds, d\zeta)$$

and

$$\big(\Pi_p^*(\mu)\big)(B) = \int_{[0,\infty]\times\mathcal{Z}} I_B(s, X_{s-} + \zeta)\Pi_p^*(\mu^X)(ds, d\zeta).$$

We can now define the characteristics of a semimartingale (these are sometimes called the local characteristics or the predictable characteristics). Suppose $X = (X^j)_{1\leq j\leq m}$, is a process with values in $\mathbb{R}^m$. In particular, suppose X is a vector semimartingale and write

$$\tilde{X}_t = \sum_{s\leq t}(\varDelta X_s)I_{\{\|\varDelta X_s\|>1\}}I_{]\!]0,\infty[\![}\,.$$

Then the process $X - \tilde{X} - X_0$ is a semimartingale with bounded jumps and so, by Theorem 11.6.10, it is a special semimartingale. Therefore, X has a canonical decomposition

$$X - \tilde{X} - X_0 = M + B, \tag{13.8}$$

where M is a process with components in $\mathcal{M}_{0,\mathrm{loc}}$ and B is a predictable process with components in $\mathcal{V}_0$.

Definition 13.4.4. *The* characteristics *of the semimartingale X is the triplet (B, C, μ_p^X) where:*

(i) B is the process defined in (13.8),
(ii) $C = [C^{jk}]_{1\leq j,\,k\leq m}$ is the $m\times m$ matrix process with components

$$C_t^{jk} = \langle (X^j)^c, (X^k)^c\rangle_t\,,$$

that is, the predictable quadratic variation process of the continuous martingale parts of X^j and X^k,
(iii) μ_p^X is the Lévy system of the jumps of X, that is, $\mu_p^X = \Pi_p^*(\mu^X)$, the dual predictable projection of the measure μ^X associated with X as in Example 13.4.1.

Remark 13.4.5. We know that $\mu_p^X \in \tilde{\mathcal{A}}_\sigma^1$, but while μ_p^X and C are intrinsic to X, the process B depends on the size of the jumps considered in the process $\tilde{X}$. For example, if $0 < \alpha < \infty$ and we consider a process

$$\tilde{X}_t^\alpha = \sum_{0<s\leq t}(\varDelta X_s)I_{\{|\varDelta X_s|>\alpha\}},$$

then $X - \tilde{X}^\alpha - X_0$ is again special, but the process B in its decomposition would be different. However, the decomposition is easily seen to be unique if we fix the value $\alpha = 1$.

Remark 13.4.6. We shall also see (Remark 15.2.5) that the semimartingale characteristics B and μ_p^X may change if the probability measure P is allowed to vary, while C remains fixed.

Remark 13.4.7. Because X is a process in $\mathbb{R}^m$, we know that $\mu_p^X(\omega;\cdot,\cdot)$ is (almost surely) a measure on $[0,\infty[\times\mathbb{R}^m$. That is, $\mathbb{R}^m$ plays the role of the abstract Blackwell space $\mathcal{Z}$ in our analysis.

Remark 13.4.8. In some sense, the process B describes a predictable 'drift' of our semimartingale through time, C describes the 'volatility of the continuous martingale part' and μ_p^X describes the rate of jumps of various sizes.

Theorem 13.4.9. *A version of (B, C, μ_p^X) can be chosen such that, for each ω,*

(i) $B(\omega)$ is càdlàg and of finite variation on each compact set,
(ii) $C(\omega)$ is continuous and, if $s \leq t$, the matrix $C_t(\omega) - C_s(\omega)$ is symmetric and nonnegative definite,
(iii) For every $t \in \mathbb{R}^+$,
  (a) $\mu_p^X(\omega, \mathbb{R}^+ \times \{0\}) = 0$,
  (b) $\mu_p^X(\omega, \{t\} \times \mathbb{R}^m) \leq 1$,
  (c) $\int_{\mathbb{R}^m}(1 \wedge \|\zeta\|^2)\mu_p^X(\omega, [0,t] \times d\zeta) < \infty$, and
  (d) $\sum_{s\leq t}\Big(\int_{\mathbb{R}^m} I_{\{\|\zeta\|\leq 1\}}\mu_p^X(\omega, \{s\} \times d\zeta)\Big) < \infty$,
(iv) for every $t \in \mathbb{R}^+$

$$\Delta B_t(\omega) = \int_{\mathbb{R}^m} I_{\{\|\zeta\|\leq 1\}}\mu_p^X(\omega, \{t\} \times d\zeta).$$

Proof. It is enough to show that the above properties are satisfied, except on a set of measure zero, by any version (B, C, μ_p^X) of the local characteristics. By definition, this is the case for (i).

For (ii), if $\alpha \in \mathbb{R}^m$, then simple calculations show that

$$\sum_{j,k=1}^m \alpha^j\alpha^k C^{jk} = \Big\langle \Big(\sum_{j=1}^m \alpha^j X^j\Big)^c, \Big(\sum_{j=1}^m \alpha^j X^j\Big)^c\Big\rangle,$$

so the process $\sum_{j,k=1}^m \alpha^j\alpha^k C^{jk}$ is increasing. By symmetry of $\langle\cdot,\cdot\rangle$, the matrix $C_t - C_s$ is a.s. symmetric, and by the above observation, taking α in a dense set guarantees it is a.s. nonnegative if $s \leq t$ (see the proof of Lemma 12.5.3 for a more careful argument). Furthermore, C is a continuous process, and so (ii) is satisfied up to evanescence.

By definition, $\mu^X(\omega, \mathbb{R}^+ \times \{0\}) = 0$. The second statement of (iii) is the result of Theorem 13.3.7.

Write $H = \{(\omega, t) : \|\Delta X_t(\omega)\| > 1\}$. As X is a semimartingale it is càdlàg and

$$A_t(\omega) := \sum_{s\leq t} I_{\{(\omega,s)\in H\}} \in \mathcal{V},$$

while

$$A'_t(\omega) := \sum_{s\leq t}(\Delta X_s)^2 I_{\{(\omega,s)\in H^c\}} \leq [X,X]_t(\omega),$$

so $A' \in \mathcal{V}$.

In fact, because A and A' both have jumps bounded by 1, they are in $\mathcal{A}_{\mathrm{loc}}$. However,

$$A_t + A'_t = \int_{\mathbb{R}^m} (1 \wedge \|\zeta\|^2)\mu^X([0,t] \times d\zeta)$$

by the definition of μ^X. The dual predictable projection of this process is

$$\Pi_p^*(A + A')_t = \int_{\mathbb{R}^m} (1 \wedge \|\zeta\|^2)\mu_p^X([0,t] \times d\zeta),$$

which is, therefore, almost surely finite for every $t < \infty$.

Now, because B is predictable,

$$\Delta B = \Pi_p^*(\Delta B) = \Pi_p^*(\Delta X - \Delta\tilde{X} - \Delta M) = \Pi_p^*(I_{H^c}\Delta X),$$

while from Theorem 13.2.21, as μ_p^X is the Lévy system of the jumps of X,

$$\Pi_p^*(I_{H^c}\Delta X)_t = \int_{\mathbb{R}^m} I_{\{\|\zeta\|\leq 1\}}\mu_p^X(\{t\} \times d\zeta),$$

which implies the last part of (iii), and (iv). □

Theorem 13.4.10. *Let X be a semimartingale. Then X can be decomposed in terms of its characteristics:*

$$X_t = X_0 + B_t + X_t^c + \int_{[0,t]\times\{\|\zeta\|<1\}} \zeta\,\tilde{\mu}^X(ds, d\zeta) + \int_{[0,t]\times\{\|\zeta\|\geq 1\}} \zeta\,\mu^X(ds, d\zeta)$$

where

- *B is the predictable process of finite variation in the characteristics of X,*
- *X^c is a continuous local martingale in $\mathcal{M}_{0,\mathrm{loc}}$, with covariation matrix C given by the characteristics of X,*
- *$\tilde{\mu}^X = \mu^X - \mu_p^X$ is the (compensated) martingale random measure associated with the jumps of X, and the associated integral is stochastic,*
- *μ^X is the (uncompensated) random measure associated with the jumps of X, and the associated integral is pathwise.*

Proof. Taking the decomposition in (13.8), we have $X = X_0 + B + M + \tilde{X}$. From Theorem 10.3.4, M can be uniquely written $M = M^c + M^d$, where M^c is a continuous local martingale and M^d is orthogonal to every continuous local martingale. We also know that $M^c = X^c = (X - \tilde{X})^c$, as $\tilde{X}$ is a process of finite variation. By definition of $\tilde{X}$, we know that $\tilde{X}_t = \int_{[0,t]\times\{\|\zeta\|\geq 1\}} \zeta\,\mu^X(ds, d\zeta)$.

It remains to show that

$$M^d = \int_{[0,t]\times\{\|\zeta\|<1\}} \zeta\,\tilde{\mu}^X(ds, d\zeta).$$

We know both processes are local martingales orthogonal to every continuous local martingale. As μ_p^X is predictable, for any totally inaccessible time T, we know μ_p^X almost surely does not charge $[\![T]\!]$, so

$$\Delta M^d_T = I_{\{\|\Delta X_T\|<1\}}\Delta X_T = \int_{\{\|\zeta\|<1\}} \zeta\, \tilde{\mu}^X(\{T\}\times d\zeta).$$

At any predictable time T, on $\{T<\infty\}$,

$$\Delta M^d_T = I_{\{\|\Delta X_T\|<1\}}\Delta X_T - E[I_{\{\|\Delta X_T\|<1\}}\Delta X_T|\mathcal{F}_{T-}] = \int_{\{\|\zeta\|<1\}} \zeta\, \tilde{\mu}^X(\{T\}\times d\zeta).$$

Therefore, the jumps of M^d and the integral agree. Exercise 11.7.12 then states that they are indistinguishable local martingales. The desired representation follows. □

13.5 Example: Lévy Processes

A class of random measures which appear commonly in applications are those associated with Lévy processes. Many books have been written on these topics, and we will only attempt to make some superficial comments here, which we present without proof. Further details can be found in Sato [165], Applebaum [1] or Protter [152].

Definition 13.5.1. *An $\mathbb{R}^d$-valued stochastic process $\{X_t\}_{t\geq 0}$ is called a* Lévy process *if*

(i) for any $s<t$, the increment $X_t - X_s$ is independent of $\mathcal{F}_s$,
(ii) the distribution of $X_t - X_s$ depends only on $t-s$,
(iii) $X_0 = 0$ a.s.
(iv) X is continuous in probability *that is, $P(|X_t - X_s|>\epsilon)\to 0$ as $s\to t$ for any ϵ, and*
(v) X is almost surely càdlàg.

If X satisfies (i – iv) only, then we say that it is a Lévy process in law.

Remark 13.5.2. Comparing with the definitions in Section 5.5, we see that both Brownian motions and Poisson processes are special cases of Lévy processes.

Definition 13.5.3. *Let $\odot$ denote the convolution of measures on $\mathbb{R}^d$, that is if μ and ν are measures, then*

$$\mu\odot\nu(d\zeta) = \int_{\mathbb{R}^d}\mu(d\zeta - y)\nu(dy).$$

A probability measure μ on $\mathbb{R}^d$ is called infinitely divisible *if, for any $n\in\mathbb{N}$, there exists a measure μ_n such that*

$$\underbrace{\mu_n\odot\mu_n\odot\ldots\odot\mu_n}_{n \text{ times}} = \mu.$$

Intuitively, a measure is infinitely divisible if for any $n \in \mathbb{N}$ one can find a sequence $X_1, \dots, X_n$ of iid random variables (with law μ_n) such that $\sum_{i=1}^n X_i \sim \mu$. That Gaussian distributions are infinitely divisible is the main result of Lemma 5.5.3.

Lemma 13.5.4. *If X is a Lévy process on $\mathbb{R}^d$, then the distribution of X_t is infinitely divisible for every t.*

Proof. We have

$$X_t = \sum_{i=1}^n (X_{ti/n} - X_{t(t-1)/n}).$$

By properties (i, ii) of the definition, this gives the desired decomposition.

The following converse theorem is a consequence of Kolmogorov's extension theorem (Theorem A.2.7) in the same way as the construction of Brownian motion in Theorem 5.5.4 (with Lemma 5.5.3 replaced by the assumption of infinite divisibility).

Theorem 13.5.5. *If μ is an infinitely divisible distribution on $\mathbb{R}^d$, then there is a Lévy process in law X such that $X_1 \sim \mu$. Furthermore, if X and X' are Lévy processes in law such that X_1 and X'_1 have the same law, then X_t and X'_t have the same law for all t.*

An elegant representation for Lévy processes is then given by the Lévy–Khintchine formula, a proof of which can again be found in [165], [1] or [152].

Theorem 13.5.6 (Lévy–Khintchine Formula). *If μ is an infinitely divisible distribution on $\mathbb{R}^d$, then*

$$\int_{\mathbb{R}^d} e^{ix} d\mu = \exp\Big[-\frac{1}{2}\langle x, Ax\rangle + i\langle \gamma, x\rangle + \int_{\mathbb{R}^d} (e^{i\langle x,\zeta\rangle} - 1 - i\langle x, \zeta\rangle I_{\|\zeta\|\leq 1})\nu(d\zeta)\Big]$$

for some symmetric nonnegative definite $d \times d$ matrix A, some $\gamma \in \mathbb{R}^d$ and some measure ν on $\mathbb{R}^d$ satisfying

$$\nu(\{0\}) = 0 \quad \text{and} \quad \int_{\mathbb{R}^d} (\|\zeta\|^2 \wedge 1)\nu(d\zeta) < \infty.$$

This representation (A, γ, ν) is unique. Conversely, if (A, γ, ν) are as stated, then an infinitely divisible distribution with this representation exists.

Note that we do not have that $\int_{\mathbb{R}^d} (\|\zeta\| \wedge 1)\nu(d\zeta) < \infty$, and so ν may place a lot of weight near $0 \in \mathbb{R}^d$.

Definition 13.5.7. *Let X be a Lévy process. The* jump measure *of X is given by μ^X as in Example 13.4.1. We write $N = \mu^X$ for simplicity, and $\tilde{N} = N - \Pi_p^*(N)$ for the compensated random measure.*

Lemma 13.5.8. *We have $d\Pi_p^*(N) = \nu(d\zeta)dt$, where ν is from the Lévy–Khintchine representation of X and, therefore, for any $\tilde{N}$ stochastically integrable H,*

$$\langle H * \tilde{N}\rangle_t = \int_{[0,t]} \int_{\mathbb{R}^d} H_s^2(\zeta)\nu(d\zeta)dt.$$

In this case, μ is called a Poisson random measure.

(More generally, we call μ a Poisson random measure whenever we can write $\mu_p(d\zeta, dt) = \nu(d\zeta)dt$, for ν some deterministic measure over the jump space $\mathcal{Z}$.)

Theorem 13.5.9 (Lévy–Itô Decomposition). *For X a Lévy process, we can write*

$$X_t = \gamma t + A^{1/2}W_t + \int_{[0,t]\times\{\|\zeta\|\leq 1\}} \zeta\tilde{N}(ds, d\zeta) + \int_{[0,t]\times\{\|\zeta\|>1\}} \zeta N(ds, d\zeta),$$

where W is a Brownian motion, $\gamma \in \mathbb{R}^d$ and N is the jump measure of X, and these integrals are well defined.

Remark 13.5.10. From this decomposition, we see that the characteristics of X, as a semimartingale, are the triple $(\gamma t, At, \nu(d\zeta)dt)$. In this sense, we see that we have the simple interpretation of γ as the infinitesimal drift of X, A as its volatility and $\nu(d\zeta)$ as the rate of jumps of sizes in $[\zeta, \zeta + d\zeta[$.

Lemma 13.5.11. *Let $B \subseteq \mathbb{R}^d$ be such that $\nu(B) < \infty$. Then $J_t = \mu^X(\omega, [0,t] \times B)$ is a Poisson process with parameter $\lambda = \nu(B)$.*

Remark 13.5.12. From this representation and Lemma 13.5.11, if X is a Lévy process with $A = 0$, then we call X a *pure jump Lévy process.*

Note that if $\int_{\mathbb{R}^d}(\|\zeta\| \wedge 1)d\nu = \infty$, then, as ν is the compensator of the jump measure of X, we see that X has *infinitely many* jumps near 0, and is not of finite variation. In this case, we say that X (and hence N) has *infinite activity.* Nevertheless, as X is a càdlàg process, we know that, for any $\epsilon > 0$, X has at most finitely many jumps of size greater than ϵ on any interval $[0,t]$. From this it is easy to deduce that $N \in \tilde{\mathcal{A}}_\sigma^1$.

13.6 The Martingale Representation Theorem

For many applications, it is important to know whether all local martingales can, in general, be represented as stochastic integrals. We shall now consider this question in the setting of a filtration generated by a random measure. We begin with the case where we have an integer valued random measure with finitely many jumps, where the argument is due to Chou and Meyer [33].

13.6.1 Finite Jump Case

Suppose that $\mu \in \tilde{\mathcal{A}}^1$, so μ is an integrable integer valued random measure. Let $\{\mathcal{F}_t\}_{t\geq 0}$ be the filtration generated by μ, that is, the smallest filtration such that μ is optional. Write $\mu_p = \Pi_p^*(\mu)$ for the compensator of μ, and $\tilde{\mu} = \mu - \mu_p$ for the associated martingale random measure.

As μ is integrable and integer valued, this implies that $\mu(\omega, [0,\infty] \times \mathcal{Z})$ is almost surely finite. Hence we can define a sequence of stopping times, with $T_0 = 0$,

$$T_i = \inf\{t > T_{i-1} : \mu(\omega, \{t\} \times \mathcal{Z}) = 1\}.$$

This increasing sequence will almost surely exhaust the jumps of $\mu(\omega, [0,\cdot]\times\mathcal{Z})$ and satisfies $T_n \to \infty$ a.s. Let $z : \Omega \times [0,\infty] \to \mathcal{Z}$ be the process associated with μ, as in Lemma 13.3.4, and write $Z_i = z_{T_i}$. We can then write the filtration as being generated by $\{T_i, Z_i\}_{i\in\mathbb{N}}$, that is,

$$\mathcal{F}_t^0 = \sigma\Big(I_{\{s\geq T_i\}}, Z_{T_i} I_{\{s\geq T_i\}} : i \in \mathbb{N}, s \leq t\Big)$$

and $\mathcal{F}_t$ is the completion of $\mathcal{F}_t^0$. As the generating processes are right-constant, it follows easily that $\{\mathcal{F}_t\}_{t\geq 0}$ is a right-continuous filtration.

Theorem 13.6.1. *In the setting described, any $\{\mathcal{F}_t\}_{t\geq 0}$-local martingale M has a representation*

$$M_t = M_0 + \int_{]0,t]\times\mathcal{Z}} g(\omega, s, \zeta)\tilde{\mu}(ds, d\zeta)$$

for some predictable, locally μ_p-integrable function g. In this case we say that $\tilde{\mu}$ has the predictable representation property *in $(\{\mathcal{F}_t\}_{t\geq 0}, P)$.*

Proof. Recall from Theorem 13.1.15 that we have a martingale representation theorem up to the first jump T_1. That is, we can write

$$M_{t\wedge T_1} = M_0 + \int_{]0,t\wedge T_1]} g^0(s,\zeta)\tilde{\mu}(ds, d\zeta)$$

for some $\mathcal{F}_0$-measurable function g^0. Now, for each $n \in \mathbb{N}$, consider the filtration $\mathcal{F}_t^n = \mathcal{F}_{t+T_n}$, and associated measure $\tilde{\mu}^n(\omega, [0,t] \times d\zeta) = \tilde{\mu}(\omega,]T_n, t + T_n] \times d\zeta)$. Then applying Theorem 13.1.15 (modified slightly to allow $\mathcal{F}_0^n$ to be nontrivial), we can also write

$$\begin{aligned} M_{(t+T_n)\wedge T_{n+1}} &= M_{0+T_n} + \int_{]0,t\wedge(T_{n+1}-T_n)]} g^n(\omega, s, \zeta)\tilde{\mu}^n(ds, d\zeta) \\ &= M_{T_n} + \int_{]T_n,(t+T_n)\wedge T_{n+1}]} g^n(\omega, s - T_n, \zeta)\tilde{\mu}(ds, d\zeta) \end{aligned}$$

for some $\mathcal{F}_0^n = \mathcal{F}_{T_n}$-measurable function g^n. Writing

$$g(\omega,t,\zeta)=\sum_n I_{]\!]T_n,T_{n+1}]\!]}g^n(\omega,t-T_n,\zeta)$$

we obtain a predictable function g such that

$$M_{t\wedge T_n}=M_0+\int_{]0,t\wedge T_n]}g(\omega,s,\zeta)\tilde{\mu}(ds,d\zeta)$$

for all $n\in\mathbb{N}$. As $T_n\to\infty$ a.s., we see that $M_{t\wedge T_n}\to M_t$ a.s. and $\int_{]0,t\wedge T_n]}g\,d\tilde{\mu}\to\int_{]0,t]}g\,d\tilde{\mu}$ a.s., which yields the desired representation. $\square$

13.6.2 Well-Ordered Jumps

Now suppose that μ is an integer valued random measure. Let $\{\mathcal{F}_t\}_{t\geq 0}$ be the natural filtration of μ. Write $\tilde{\mu}=\mu-\mu_p$ for the martingale random measure generated by μ. Suppose that the jumps of μ are well ordered, that is, for any stopping time S there exists a stopping time $T>S$ such that

$$\mu(\omega,]\!]S,T[\![\times\mathcal{Z})=0,\qquad \mu(\omega,[\![T]\!]\times\mathcal{Z})=1.$$

In order to prove the martingale representation theorem in this setting, we use the following version of the principle of transfinite induction.

We denote by $\daleth$ the first uncountable ordinal (this is conventionally called I, Ω or ω_1; however this leads to confusion in our probabilistic setting). Considered as a set, $\daleth$ is well ordered and uncountable, and for every $\beta\in\daleth$ the set $\{\alpha:\alpha\leq\beta\}$ is countable. The smallest element of $\daleth$ is denoted 0, the successor to $\alpha\in\daleth$ is denoted $\alpha+1$, and α is called the predecessor of $\alpha+1$. If $\alpha\in\daleth$ does not have a predecessor it is called a limit ordinal, other elements are non-limit ordinals. The following theorem gives the properties of $\daleth$ we use and can be found in Dellacherie and Meyer [54, Vol I, Chapter 0].

Theorem 13.6.2 (Principle of Transfinite Induction).

- *If f is a monotonic increasing function from $\daleth$ into $\overline{\mathbb{R}}$, then there is an ordinal $\alpha\in\daleth$ such that $f(\beta)=f(\alpha)$ for every $\beta\geq\alpha$.*
- *Suppose $P(\alpha)$ is some property of the ordinal α, and suppose that*
  (i) if $P(\alpha)$ is true, then $P(\alpha+1)$ is true,
  (ii) if β is a limit ordinal and if $P(\alpha)$ is true for all $\alpha<\beta$, then $P(\beta)$ is true, and
  (iii) $P(0)$ is true.
  Then $P(\alpha)$ is true for all $\alpha\in\daleth$.

We first show the following lemma.

Lemma 13.6.3. *For a random measure in $\tilde{\mathcal{A}}^1_\sigma$ with well ordered jumps, we can write the jumps of μ as a family of stopping times $\{T_\alpha\}_{\alpha\in\daleth}$ such that μ is supported on $\cup_{\alpha\in\daleth}[\![T_\alpha]\!]$. These stopping times satisfy $T_\alpha<T_\beta$ on the set $\{T_\alpha<\infty\}$ for any $\alpha<\beta$ in $\daleth$, and for each ω, there exists an $\alpha\in\daleth$ such that $T_\alpha(\omega)=\infty$.*

Proof. Define $T_0 = 0$. Under the conditions of the theorem, if T_α is defined then we can find a stopping time $T_{\alpha+1}$ satisfying $T_{\alpha+1} > T_\alpha$ on $\{T_\alpha < \infty\}$ and μ does not charge the set $\rrbracket T_\alpha, T_{\alpha+1}\llbracket$. If α is a limit ordinal, then we define $T_\alpha = \sup_{\beta<\alpha} T_\beta$, which is a stopping time as the set $\beta < \alpha$ is countable. As $\alpha \mapsto T_\alpha(\omega)$ is a monotonic increasing function from $\gimel$ into $\overline{\mathbb{R}}$ for each ω, from Theorem 13.6.2 there is an ordinal $\alpha \in \gimel$, depending on ω, such that $T_\beta(\omega) = T_\alpha(\omega)$ for every $\beta \geq \alpha$. As $T_\beta > T_\alpha$ for all $\beta > \alpha$ unless $T_\alpha = \infty$, we see that $T_\alpha = \infty$. □

Theorem 13.6.4. *In the setting of a random measure with well-ordered jumps, any $\{\mathcal{F}_t\}_{t\geq 0}$-local martingale M has a representation*

$$M_t = M_0 + \int_{]0,t]\times\mathcal{Z}} H(\omega, s, \zeta)\tilde{\mu}(ds, d\zeta),$$

for some predictable, locally μ_p-integrable function H.

Proof. By localizing, we can assume that M is a uniformly integrable martingale. For notational simplicity, we write H_t for $H(\omega, t, z)$. We now show that there is a predictable function H such that for any $\alpha \in \gimel$,

$$M_{T_\alpha} = M_0 + \int_{]0,T_\alpha]\times\mathcal{Z}} H_t d\tilde{\mu}. \tag{13.9}$$

This is clear for $\alpha = 0$. We next apply transfinite induction.

Suppose that we can establish (13.9) for a given α. Then consider the filtration $\mathcal{F}^\alpha_t = \mathcal{F}_{t+T_\alpha}$. By Theorem 13.1.15 we can find $\bar{H}$ such that

$$M_{T_{\alpha+1}} = M_{T_\alpha} + \int_{]T_\alpha,T_{\alpha+1}]\times\mathcal{Z}} \bar{H}_t d\tilde{\mu}.$$

Defining $\tilde{H} = I_{[0,T_\alpha]}H + I_{]T_\alpha,T_{\alpha+1}]}\bar{H}$, rearrangement yields

$$M_{T_{\alpha+1}} = M_0 + \int_{]0,T_{\alpha+1}]\times\mathcal{Z}} \tilde{H}_t d\tilde{\mu},$$

so (13.9) holds for $\alpha + 1$.

Now suppose that we can establish (13.9) for all $\alpha < \beta$, where β is a limit ordinal. As M is assumed to be uniformly integrable, by martingale convergence we have (the convergence being almost sure),

$$M_{T_\beta-} = \lim_{\alpha\uparrow\beta} M_{T_\alpha} = \lim_{\alpha\uparrow\beta}\Big(M_0 + \int_{]0,T_\alpha]\times\mathcal{Z}} H_t d\tilde{\mu}\Big) = M_0 + \int_{]0,T_\beta[\times\mathcal{Z}} H_t d\tilde{\mu}$$

for some $H : \rrbracket 0, T_\beta\llbracket \times \mathcal{Z} \to \mathbb{R}$. Furthermore, by the Doob–Dynkin lemma, we know that $M_{T_\beta} = M_{T_\beta-} + g(\omega, Z_{T_\beta}(\omega))$, where $Z_{T_\beta}(\omega)$ is the unique value in $\mathcal{Z}$ such that $\mu(\omega, \{T_\beta\}\times\{Z_{T_\beta}(\omega)\}) = 1$, if such a value exists, and $Z_{T_\beta}(\omega) = \emptyset_{\mathcal{Z}}$

otherwise (where $\emptyset_{\mathcal{Z}}$ denotes a value not in $\mathcal{Z}$), and g is some $\mathcal{F}_{T_\beta-} \otimes (\mathfrak{Z} \vee \sigma(\{\emptyset_{\mathcal{Z}}\}))$-measurable function.

As T_β is predictable for β a limit ordinal, $E[M_{T_\beta}|\mathcal{F}_{T_\beta-}] = M_{T_\beta-}$, so $\int_{\mathcal{Z}} g(\omega,z)\mu_p(\omega,\{T_\beta\}\times dz) = 0$. Therefore,

$$M_{T_\beta} = M_{T_\beta-} + \int_{\mathcal{Z}} g(\omega,z)\mu(\omega,\{T_\beta\}\times dz) = M_{T_\beta-} + \int_{\mathcal{Z}} g(\omega,z)\tilde{\mu}(\omega,\{T_\beta\}\times dz),$$

which implies, writing $H_{T_\beta} = g$,

$$M_{T_\beta} = M_0 + \int_{]0,T_\beta]\times\mathcal{Z}} H_t d\tilde{\mu}.$$

By the principle of transfinite induction, we therefore know that (13.9) holds for all $\alpha \in \beth$. By Lemma 13.6.3, for each ω there exists an α such that $T_\beta = \infty$ for all $\beta > \alpha$, so we have

$$M_\infty = M_0 + \int_{]0,\infty]\times\mathcal{Z}} H_t d\tilde{\mu}.$$

Stopping M at time t and taking a conditional expectation, we obtain the desired representation. □

13.6.3 Deterministic Compensator

We now show that, if the random measure has a deterministic compensator (but the jumps are not necessarily well ordered), then the martingale representation theorem remains true. A counterexample to a more general statement (when the compensator is not deterministic) is given in Exercise 18.4.5. We do this using a monotone class argument. In order to use this, a preliminary approximation result is needed. For simplicity we shall assume that μ_p is also continuous, which implies $[H*\tilde{\mu}] = H^2*\mu$ and $\langle H*\tilde{\mu}\rangle = H^2*\mu_p$. Effectively, this is equivalent to assuming our filtration is quasi left-continuous.

Lemma 13.6.5. *Suppose that, for every $B \in \mathcal{F}_\infty$, we can write*

$$E[I_B|\mathcal{F}_t] = E[I_B|\mathcal{F}_0] + \int_{]0,t]\times\mathcal{E}} H_t d\tilde{\mu}$$

for some $\tilde{\mu}$-stochastically integrable function H. Then any local martingale M can be written $M = M_0 + H*\tilde{\mu}$ for some predictable, $\tilde{\mu}$-stochastically integrable process H.

Proof. First suppose M is a uniformly integrable martingale and $M_\infty \in L^{1+\epsilon}$ for some $\epsilon > 0$. Then there exists a sequence of simple random variables $M_\infty^{(n)}$ (i.e. of the form $M_\infty^{(n)} = \sum_{k=1}^n c_k I_{B_k}$) which converge to M_∞ in $L^{1+\epsilon}$. By linearity of the integral, every simple function $M^{(n)}_\infty$ has the desired representation in terms of an integrand H^n, and by the BDG inequality and Doob's L^p inequality, as $n \to \infty$

$$E\Big[\big(\,((H^n - H^m)^2 * \mu)_\infty\,\big)^{\frac{1+\epsilon}{2}}\Big]^{1/(1+\epsilon)} \leq K\|M^{(n)}_\infty - M^{(m)}_\infty\|_{L^{1+\epsilon}} \to 0,$$

for some constant K depending on ϵ. By completeness of L^p spaces, there exists a limiting process H, which is stochastically integrable with respect to $\tilde{\mu}$, such that

$$M_\infty = M_0 + \int_{]0,\infty]} H_t d\tilde{\mu}.$$

Now suppose $M^*_\infty = \sup_t |M_t| \in L^1$, so $M \in \mathcal{H}^1$. For any $\epsilon > 0$, as $\mathcal{H}^{1+\epsilon}$ is dense in $\mathcal{H}^1$ (Theorem 10.1.7), we can find a sequence $M^{(n)}$ such that $\|M^{(n)} - M\|_{\mathcal{H}^1} \to 0$ and $M^{(n)}_\infty \in L^{1+\epsilon}$. Therefore, from the BDG inequality, as $M^{(n)}$ has the desired representation for each n, there exists a constant K such that

$$E\Big[\big(\,((H^n - H^m)^2 * \mu)_\infty\,\big)^{\frac{1}{2}}\Big] \leq K\|M^{(n)} - M^{(m)}\|_{\mathcal{H}^1} \to 0,$$

and so an appropriate limit H exists. That H is stochastically integrable with respect to $\tilde{\mu}$ follows from the fact M is a semimartingale, so $[M]^{1/2} \in \mathcal{A}_{\mathrm{loc}}$.

Finally, if M is a local martingale, then the stopping times $T_n = \inf\{t : |M_t| \geq n\}$ form a localizing sequence with $\sup_t |M_{t\wedge T_n}| \leq n + |\Delta M_{T_n}| \in L^1$. We can therefore write

$$M_{t\wedge T_n} = M_0 + \int_{]0,t\wedge T_n]} H_s d\tilde{\mu}$$

for each n, and pasting yields the desired representation. □

Remark 13.6.6. From the BDG inequality, we also see that this representation is unique, in the sense that if $M = a + (g * \tilde{\mu}) = b + (h * \tilde{\mu})$, then $a = b = M_0$ and for $\{T_n\}_{n\in\mathbb{N}}$ a localizing sequence such that the stopped processes satisfy $M^{T_n} \in \mathcal{H}^1$, we have

$$E\Big[\Big(\int_{]0,T_n]\times\mathcal{E}} (g_t - h_t)^2 d\mu\Big)^{\frac{1}{2}}\Big] = 0$$

for all n, and hence $E\big[\big((g-h)^2 * \mu_p\big)_\infty\big] = 0$.

Theorem 13.6.7. *Let $\mu \in \tilde{\mathcal{A}}^1_\sigma$ be a random measure with a continuous deterministic compensator in its natural filtration $\{\mathcal{F}_t\}_{t\geq 0}$. Write $\tilde{\mu} = \mu - \mu_p$. Then any local martingale M can be written $M = M_0 + H * \tilde{\mu}$ for some predictable, $\tilde{\mu}$-stochastically integrable process H, that is, $\tilde{\mu}$ has the predictable representation property in $(\{\mathcal{F}_t\}_{t\geq 0}, P)$.*

Proof. By Lemma 13.6.5, it is enough to prove the theorem under the assumption that $M_\infty = I_B$ for some $B \in \mathcal{F}_\infty$. As μ_p is deterministic, there exists an increasing family of deterministic sets $\{A_n\}_{n\in\mathbb{N}}$ such that $\cup_n A_n = [0,\infty] \times \mathcal{Z}$ and, for every n, $I_{A_n} \cdot \tilde{\mu}$ has well ordered jumps. Write $\{\mathcal{F}^n_t\}_{t\geq 0}$ for the natural filtration of $I_{A_n} \cdot \mu$.

If $B \in \mathcal{F}^n_\infty$ for some n, we have a representation for the martingale $M = \{E[I_B|\mathcal{F}^n_t]\}_{t\geq 0}$. As A_n and μ_p are deterministic, $I_{A_n} \cdot \mu_p$ is the compensator of $I_{A_n} \cdot \mu$ in $\{\mathcal{F}^n_t\}_{t\geq 0}$, so this representation is in terms of the $\{\mathcal{F}_t\}_{t\geq 0}$-martingale random measure $\mu - \mu_p$, and we know that $M_t = E[I_B|\mathcal{F}_t]$.

Let $\mathcal{C}$ be the family of all sets $B \in \mathcal{F}_\infty$ such that $M = E[I_B|\mathcal{F}_{(\cdot)}]$ has a representation of the desired form. From our earlier comments, $\mathcal{C}$ contains the algebra of sets $\cup_{n\in\mathbb{N}}\mathcal{F}^n_\infty$ (this is an algebra as $\mathcal{F}^n_\infty$ is increasing in n). If $\{B_m\}_{m\in\mathbb{N}}$ is an increasing sequence in $\mathcal{C}$ with $B = \cup_m B_m$, then $I_{B_m} \to I_B$ in L^2, by monotone convergence. Considering the corresponding representations,

$$I_{B_m} = M_0 + \int_{]0,\infty]\times\mathcal{Z}} H^m_t d\tilde{\mu},$$

and from Itô's isometry (as μ_p is continuous), we have

$$\begin{aligned} E[(I_{B_m} - I_{B_{m'}})^2] &= E\Big[\Big(\int_{]0,\infty]\times\mathcal{Z}} (H^m_t - H^{m'}_t) d\tilde{\mu}\Big)^2\Big] \\ &= E\Big[\int_{]0,\infty]\times\mathcal{Z}} (H^m_t - H^{m'}_t)^2 d\mu_p\Big] \end{aligned}$$

so H^m converges in the Hilbert space $L^2(\mu_p \times P)$ to a process H. Hence there exists the limit

$$I_B = M_0 + \int_{]0,\infty]} H_t d\tilde{\mu},$$

and we see that $B \in \mathcal{C}$. By an identical argument, if $\{B_m\}_{m\in\mathbb{N}}$ is a decreasing sequence we have $\cap_m B_m \in \mathcal{C}$. Therefore $\mathcal{C}$ is a monotone class.

The monotone class theorem then states that $\mathcal{C}$ contains the σ-algebra generated by $\cup_n \mathcal{F}^n_\infty$, that is, $\mathcal{F}_\infty$. Therefore, for any $B \in \mathcal{F}_\infty$ we have the desired representation of $M = E[I_B|\mathcal{F}_{(\cdot)}]$. By Lemma 13.6.5, the result follows. □

Remark 13.6.8. The case of general Lévy processes shall be addressed in Section 14.5 and can be considered directly through the results of Nualart and Schoutens [140]. A significant general result in this area is the Jacod–Yor Theorem (see Section 18.3), which relates martingale representations with the non-existence of equivalent martingale measures. A review is given by Davis [43].

13.7 Exercises

Exercise 13.7.1. Consider a single jump process, and let F, Λ be as in Definition 13.1.4. Assume that F is absolutely continuous, so we can define

$$\frac{1}{F_{s-}}\frac{dF_s}{ds} = \frac{d(\log(F_s))}{ds} = -H(s).$$

Here H is the 'hazard rate'. Assuming $P(T < \infty) = 1$, show that $\int_{[0,\infty]} H(s)ds = \infty$. Conversely, given a nonnegative function H satisfying this property, show that there exists a well-defined jump time T with H as a hazard rate. Under what conditions is the jump time bounded a.s.?

Exercise 13.7.2. Give an example of a random measure in $\tilde{\mathcal{V}}$ which is not in $\tilde{\mathcal{A}}_\sigma$.

Exercise 13.7.3. Show that if $\mu \in \tilde{\mathcal{A}}^1$, then the section $D_\omega = \{t \geq 0 : (\omega, t) \in D\}$ (where D is the set of points charged by μ, as in Remark 13.3.3) is almost surely finite. Hence show that D is thin whenever $\mu \in \tilde{\mathcal{V}}^1$.

Exercise 13.7.4. Let $\mu \in \tilde{\mathcal{A}}^1_\sigma$ be such that $\mu_p(\omega, \{t\} \times \mathcal{Z}) = 0$ up to indistinguishability. Show that $W * (\mu - \mu_p) \in \mathcal{H}^2$ if and only if $W \in L^2(M_\mu)$, where M_μ is the Doléans measure associated with μ.

Exercise 13.7.5. Give an example of a random measure $\mu \in \tilde{\mathcal{A}}_\sigma$ with compensator μ_p such that D is an infinite set and μ_p is supported by D (where D is as in Remark 13.3.3), and an example where $\mu_p(\omega, D) = 0$ a.s.

Exercise 13.7.6. Give an example of a random measure $\mu \in \tilde{\mathcal{A}}^1_\sigma$ such that there exists no process X generating μ, in the sense of Example 13.3.5.

Exercise 13.7.7. (This exercise is a counterexample to extending Theorem 11.4.7 to a general local martingale setting, due to Ruf and Larsson [164]).

Let X be a compensated single jump process, defined by the deterministic functions ψ and ϕ, and the jump time T, where X is the local martingale

$$X_t = \psi_T I_{\{t \geq T\}} - \int_{[0,t\wedge T]} \psi_t \phi_t dt.$$

(i) Write down the Lévy system of X.
(ii) Show that $P(T = \infty) > 0$ whenever $\int_{[0,\infty]} \phi_t dt < \infty$.
(iii) Give an example of functions ψ and ϕ such that $P(T = \infty) > 0$ and $X_t \to -\infty$ whenever $T = \infty$.
(iv) Hence show that convergence of $[X]$ to a finite limit is not equivalent to convergence of a general local martingale X.
(v) Show directly that $E\big[[X]_\infty^{1/2}\big] = \infty$ for your example, and compare with the result of the BDG inequality.

Part IV

Stochastic Differential Equations

14

Itô's Differential Rule

In order to use the theory of stochastic integration, much like in classical integration, certain rules are of fundamental importance. The most famous of these, 'Itô's Differential Rule', generalizes the chain rule from classical calculus. Deriving this rule and exploring its consequences are the aims of this chapter.

As before, we assume that we have a filtration $\{\mathcal{F}_t\}_{t\geq 0}$ satisfying the usual conditions, and for simplicity $\mathcal{F}_{\infty-} = \mathcal{F}_\infty$. Martingales will be assumed to be càdlàg throughout, and all (in)equalities should be read as 'up to indistinguishability' unless otherwise indicated.

Remark 14.0.1. In this chapter, we shall often be dealing with partitions $\pi = \{0 = t_0 < t_1 < \dots\}$ of the interval $[0, \infty[$, where the t_i are stopping times. We always assume that the set $\{t_i \in \pi : t_i < t\}$ is almost surely finite for every $t < \infty$. For a sequence of partitions $\{\pi_n\}_{n\in\mathbb{N}} = \{\{t_i^n\}_{i\in\mathbb{N}}\}_{n\in\mathbb{N}}$, we write

$$|\pi_n| \to 0 \quad \text{if} \quad \sup_{i\in\mathbb{N}}\{|t_i^n \wedge s - t_{i-1}^n \wedge s|\} \to 0 \quad \text{a.s. for every } s > 0.$$

For a process X and partition π, we write X^π for the left-continuous process defined by

$$X_t^\pi = X_{t_i} \text{ for } t \in]t_i, t_{i+1}] \text{ for each } i.$$

Note that if X is càdlàg and π_n is a sequence of partitions with $|\pi_n| \to 0$, then $X_t^{\pi_n} \to X_{t-}$ almost surely for every t, as $n \to \infty$.

S.N. Cohen, R.J. Elliott, *Stochastic Calculus and Applications*,
Probability and Its Applications, DOI 10.1007/978-1-4939-2867-5_14

14.1 Integration by Parts

Before proceeding to the general Itô rule, we will first prove a generalization of the integration by parts formula which we obtained, for finite variation paths, in Theorem 1.3.43. While we state this result for local martingales, we shall see (Theorems 14.2.3 and 14.2.4) that it equally holds for X and Y semimartingales.

Theorem 14.1.1. *Let X and Y be two càdlàg local martingales. Then*

$$X_tY_t = X_0Y_0 + \int_{]0,t]} X_{u-}dY_u + \int_{]0,t]} Y_{u-}dX_u + [X,Y]_t.$$

Proof. By rearrangement and the fact $2XY = (X+Y)^2 - X^2 - Y^2$, it is enough to consider the case $X = Y$.

For some $n \in \mathbb{N}$, let $T = \inf\{t \geq 0 : |X_t| \geq n\}$. Let π be a partition of $[0,\infty[$ as in Remark 14.0.1. As X is càdlàg, we see that $|X^\pi I_{[\![0,T]\!]}| \leq n$ and $X_t^\pi I_{[\![0,T]\!]} \to X_{t-} I_{[\![0,T]\!]}$ a.s. as $|\pi_n| \to 0$.

We now write $(X^T)^2$ as a telescoping sum,

$$\begin{aligned}(X^T)_t^2 &= (X_0^T)^2 + 2\sum_{i\in\mathbb{N}} X_{t_i}^T(X_{t_{i+1}\wedge t}^T - X_{t_i\wedge t}^T) + \sum_i (X_{t_{i+1}\wedge t}^T - X_{t_i\wedge t}^T)^2 \\ &= (X_0^T)^2 + 2((X^\pi I_{[\![0,T]\!]})\bullet X)_t + \sum_{i\in\mathbb{N}}(X_{t_{i+1}\wedge t\wedge T} - X_{t_i\wedge t\wedge T})^2.\end{aligned}$$

By the assumptions on π, we note that only finitely many terms of this sum are nonzero. Write $Q(\pi,t) := \sum_i (X_{t_{i+1}\wedge t} - X_{t_i\wedge t})^2$. By the stochastic dominated convergence theorem (Theorem 12.4.10) we know that

$$(X^\pi I_{[\![0,T]\!]})\bullet X \to \int_{]0,\cdot\wedge T]} X_{u-}dX_u$$

as $|\pi| \to 0$, in the semimartingale topology. Hence, as $|\pi| \to 0$,

$$\{Q(\pi, T\wedge t)\}_{t\geq 0} \to \{Q_t^{(T)}\}_{t\geq 0},$$

the convergence being ucp[1], for some càdlàg process $Q^{(T)}$. As ucp convergence implies convergence of the jumps of a process (Lemma 12.4.2), we see that $\Delta Q^{(T)} = (\Delta X^T)^2$.

As n was arbitrary, we now have a family of processes $\{Q^{(T_n)}\}_{n\in\mathbb{N}}$, where $T_n = \inf\{t \geq 0 : |X_t| \geq n\}$. It is easy to verify that if $m > n$ (and so $T_m \geq T_n$), then $Q^{(T_n)} = Q^{(T_m)}$ on the interval $[\![0, T_n]\!]$. Hence, by pasting these processes together, we can define a single process Q such that $Q = Q^{(T_n)}$ on $[\![0, T_n]\!]$

[1]Given the result of the theorem, we can see that X^2 is a semimartingale, so the convergence $Q(\pi, T\wedge\cdot) \to Q^{(T)}$ is also in the semimartingale topology.

for all n. As X is real valued, we know $T_n \to \infty$ a.s. as $n \to \infty$, so by Lemma 12.4.8, we know $Q(\pi, \cdot) \to Q$ ucp.

As $Q(\pi, t)$ is nondecreasing in t, we see that Q_t is nondecreasing in t, and also that $X^2 - Q = 2\int_{]0,\cdot]} X_{u-} dX_u$ is a local martingale. We also know that $\Delta Q = (\Delta X)^2$. By Exercise 11.7.2, $[X]$ is the only process satisfying these properties, so $Q = [X]$, as desired. □

In the course of this proof, we have also established this fundamental corollary, which justifies the name 'quadratic variation' for $[X]$.

Corollary 14.1.2. *For any sequence of partitions $\{\pi_j\}_{j\in\mathbb{N}} = \{\{t_i^j\}_{i\in\mathbb{N}}\}_{j\in\mathbb{N}}$ of $[0,\infty[$ by stopping times, as $|\pi_j| \to 0$ the squared difference process satisfies*

$$Q(\pi_j, t) = \sum_i (X_{t_{i+1}^j \wedge t} - X_{t_i^j \wedge t})^2 \to [X]_t$$

in probability, and furthermore, $Q(\pi_j, \cdot) \to [X]$ in $\mathcal{S}$ and hence ucp.

14.2 Itô's Rule

The main result of this chapter, Itô's differentiation rule (also known as Itô's lemma), will now be proven. It is a generalization of the differentiation rule established by Itô in 1942 ([101], see also [102]), for stochastic integrals with respect to Brownian motion.

Interestingly, a version of this result was developed by Wolfgang Döblin in 1940 (see [27]) and was submitted to the French Académie des Sciences in a sealed envelope, shortly before Döblin's death on the western front. This envelope was only opened 60 years later, well after the development of stochastic calculus as a major discipline, and so the rule is conventionally attributed to Itô.

The rule is first established in the simplest case, when X is a continuous, bounded, real semimartingale.

Theorem 14.2.1. *Let X be a continuous real semimartingale with $|X| < K$, for some $K \in \mathbb{R}$. Let f be a real valued function on $\mathbb{R}$ which is twice continuously differentiable. Then $f(X)$ is a semimartingale and, writing f' and f'' for the first and second derivatives of f,*

$$f(X_t) = f(X_0) + \int_{]0,t]} f'(X_u) dX_u + \frac{1}{2}\int_{]0,t]} f''(X_u) d\langle X\rangle_u .$$

up to indistinguishability.

Proof. By Taylor's approximation theorem, we can write

$$f(b) - f(a) = (b-a)f'(a) + \frac{1}{2}(b-a)^2 f''(a) + r(a,b),$$

where, as f'' is continuous,

$$|r(a,b)| \leq \rho(|b-a|)(b-a)^2 \qquad \text{for } a, b \in [-K, K],$$

for some increasing function ρ with $\lim_{s\to 0} \rho(s) = 0$.

For a partition π as in Remark 14.0.1 we have

$$\begin{aligned} f(X_t) &= f(X_0) + \sum_{i=0}^{\infty} \big(f(X_{t_{i+1}\wedge t}) - f(X_{t_i\wedge t})\big) \\ &= f(X_0) + \sum_{i=0}^{\infty} \Big(f'(X_{t_i})(X_{t_{i+1}\wedge t} - X_{t_i\wedge t}) \\ &\qquad + \frac{1}{2} f''(X_{t_i})(X_{t_{i+1}\wedge t} - X_{t_i\wedge t})^2 + r(X_{t_{i+1}\wedge t}, X_{t_i\wedge t})\Big). \end{aligned} \tag{14.1}$$

Again, note that by assumption on π, only finitely many terms in these sums are nonzero. As one might expect, we now take the limit $|\pi| \to 0$ and show this converges to the desired expression.

The fact $|X| < K$ and f is twice continuously differentiable implies $f'(X)$ and $f''(X)$ are both bounded. As X is a.s. continuous, writing $X_t^\pi = X_{t_i}$ for $t \in]t_i, t_{i+1}]$, we know that $f'(X_t^\pi) \to f'(X_t)$ a.s. as $|\pi| \to 0$, and similarly for f''. In particular, note that $f''(X^\pi) \in \Lambda$ (where Λ is defined in Definition 12.1.1). Therefore, Lemma 12.4.6, Theorems 12.4.10 and 12.4.13 and Corollary 14.1.2 imply that, as $|\pi| \to 0$,

$$\begin{aligned} \sum_{i=0}^{\infty} \big(f'(X_{t_i})(X_{t_{i+1}\wedge t} - X_{t_i\wedge t})\big) &= \int_{]0,t]} f'(X^\pi) dX \\ &\to \int_{]0,t]} f'(X_u) dX_u \end{aligned} \tag{14.2}$$

in probability. Similarly, as $Q(\pi, \cdot) \to [X]$ in $\mathcal{S}$ and $f''(X^\pi) \in \Lambda$ is uniformly bounded, using Lemma 12.4.6 and Corollary 14.1.2 we know

$$\int_{]0,\cdot]} f''(X_u^\pi) d(Q(\pi, u) - [X]_u) \to 0 \qquad \text{in ucp},$$

so Theorem 12.4.10 implies

$$\begin{aligned} &\sum_{i=0}^{\infty} \big(f''(X_{t_i})(X_{t_{i+1}\wedge t} - X_{t_i\wedge t})^2\big) \\ &= \int_{]0,t]} f''(X_u^\pi) dQ(\pi, u) \\ &= \int_{]0,t]} f''(X_u^\pi) d(Q(\pi, u) - [X]_u) + \int_{]0,t]} f''(X_u^\pi) d[X]_u \\ &\to 0 + \int_{]0,t]} f''(X_u) d[X]_u. \end{aligned} \tag{14.3}$$

It now only remains to deal with the remainder term. We know that

$$\Big|\sum_{i\in\mathbb{N}} r(X_{t_{i+1}\wedge t}, X_{t_i\wedge t})\Big| \leq \sum_{i=0}^{\infty} \rho(|X_{t_{i+1}\wedge t} - X_{t_i\wedge t}|)(X_{t_{i+1}\wedge t} - X_{t_i\wedge t})^2.$$

Suppose $\{\pi^n\}_{n\in\mathbb{N}}$ is a sequence of sufficiently fine partitions that

$$|X_{t^n_{i+1}\wedge s} - X_{t^n_i\wedge s}| \leq n^{-1} \text{ for all } s < t.$$

For such a sequence of partitions[2], we know

$$\rho(|X_{t^n_{i+1}\wedge s} - X_{t^n_i\wedge s}|) \leq \rho(n^{-1}) \to 0.$$

Therefore, by Corollary 14.1.2 we have

$$\sum_{i=0}^{\infty} \rho(|X_{t_{i+1}\wedge t} - X_{t_i\wedge t}|)(X_{t_{i+1}\wedge t} - X_{t_i\wedge t})^2 \leq \rho(n^{-1})Q(\pi^n, t) \to 0, \quad (14.4)$$

the convergence being in probability.

Substituting (14.2), (14.3) and (14.4) into (14.1) and taking the limit, as X is continuous $[X] = \langle X\rangle$, so we see that

$$f(X_t) = f(X_0) + \int_{]0,t]} f'(X_u)dX_u + \frac{1}{2}\int_{]0,t]} f''(X_u)d\langle X\rangle_u \qquad \text{a.s.}$$

for every t. As $f(X)$ is continuous, the equality holds up to indistinguishability. As $f(X_t)$ can be represented in this way, it follows that $f(X)$ is a semimartingale. □

In order to extend this result to discontinuous X, we first show the following lemma.

Lemma 14.2.2. *Let X be a semimartingale and $f : \mathbb{R} \to \mathbb{R}$ a twice continuously differentiable function. Then, for each t, the sum*

$$\sum_{0<s\leq t} \big(f(X_s) - f(X_{s-}) - f'(X_{s-})\Delta X_s\big)$$

is almost surely absolutely convergent.

---

[2] An example of such a sequence is given by

$$t^n_i = \inf\{t > t^n_{i-1} : |X_t - X_{t^n_{i-1}}| \geq n^{-1}\} \wedge (t^n_{i-1} + n^{-1}),$$

where, as X is continuous, we know that $t^n_i \to \infty$ a.s. as $i \to \infty$.

Proof. For almost any $\omega \in \Omega$ the path $\{X_s(\omega)\}_{s\in[0,t]}$ remains in a compact interval $[-C(t,\omega), C(t,\omega)]$. On such an interval, the second derivative of f is bounded by some constant $K(t,\omega)$. Therefore, for $s \leq t$,

$$\big|f(X_s) - f(X_{s-}) - f'(X_{s-})\Delta X_s\big| \leq \frac{1}{2}K(t,\omega)(\Delta X)_s^2.$$

As in the remarks following Definition 11.6.6, we know that $\sum_{s\leq t}(\Delta X)_s^2$ is almost surely finite. Therefore, for any t, the sum is almost surely absolutely convergent. □

The differentiation rule will now be extended to the situation when X is a general semimartingale and f is a twice continuously differentiable function. Taking $f(x) = x^2$ gives the result of Theorem 14.1.1 for $X = Y$ a semimartingale.

Theorem 14.2.3 (Itô's Rule). *Suppose X is a semimartingale and $f : \mathbb{R} \to \mathbb{R}$ a twice continuously differentiable function. Then $f(X)$ is a semimartingale, and*

$$\begin{aligned}
f(X_t) &= f(X_0) + \int_{]0,t]} f'(X_{s-})dX_s + \frac{1}{2}\int_{]0,t]} f''(X_{s-})d\langle X^c\rangle_s \\
&\quad + \sum_{0<s\leq t}\big(f(X_s) - f(X_{s-}) - f'(X_{s-})\Delta X_s\big) \\
&= f(X_0) + \int_{]0,t]} f'(X_{s-})dX_s + \frac{1}{2}\int_{]0,t]} f''(X_{s-})d[X]_s \\
&\quad + \sum_{0<s\leq t}\Big(f(X_s) - f(X_{s-}) - f'(X_{s-})\Delta X_s - \frac{1}{2}f''(X_{s-})(\Delta X)_s^2\Big).
\end{aligned}$$

Proof. Philosophically, this proof works in the same way as that of Theorem 14.2.1; we use Taylor's theorem to approximate the function f on a partition π, then show that the limit can be taken. The only issues are that we no longer have boundedness of our integrands and we need to take care at the jumps of X. Note that the two stated forms of the equality differ only in the decomposition of $[X]_t = \langle X^c\rangle_t + \sum_{0<s\leq t}(\Delta X)^2$, and by Lemma 14.2.2 the sums are absolutely convergent. Therefore, if we can show the stated equality holds, we also establish that $f(X)$ is a semimartingale. As $f(X)$ is clearly càdlàg, it is enough to show that the equality holds almost surely for each t.

By considering stopping times of the form

$$\inf\{t : |f'(X_{t-})| + |f''(X_{t-})| \geq n\}$$

we see that $\{f'(X_{t-})\}_{t\geq 0}$ and $\{f''(X_{t-})\}_{t\geq 0}$ are locally bounded processes. As the stochastic dominated convergence theorem (in particular Corollary 12.4.11) applies to locally bounded integrands, this will help resolve the first issue.

Fix $t > 0$. Let $\{\pi^n\}_{n\in\mathbb{N}}$ be a sequence of partitions of $[0,t]$ by stopping times, $\pi^n = \{0 = t_1^n \le t_2^n \le \ldots\}$ such that

$$\text{(i)}\quad \sup_i \left\{|t_{i+1} - t_i| + |X_{(t_{i+1}^n-)} - X_{t_i^n}|\right\} \to 0 \text{ as } n \to \infty,$$

$$\text{(ii)}\quad \{(t,\omega) : |\Delta X_t(\omega)| > n^{-1}\} \subset [\![\pi^n]\!] =: \bigcup_i [\![t_i^n]\!]$$

$$\text{(iii)}\quad \text{for each } n, t_i^n < t \text{ for a.s. finitely many } i.$$

The key idea is that the approximations X^{π^n} will converge, and π^n will specifically capture all jumps of X above a certain size (by property (ii)). As X is càdlàg and has a.s. at most finitely many jumps of size $\ge n^{-1}$ on any compact, and X and $[X]$ are finite valued, such a partition exists. An example is given recursively by

$$t_{i+1}^n = \inf\left\{t \ge t_i^n : |X_t - X_{t_i^n}| + |t - t_i^n| + |\Delta X_t(\omega)| > n^{-1}\right\}.$$

For notational simplicity, we suppress the n when writing t_i^n.

We express $f(X_t)$ using the telescoping sum

$$\begin{aligned} &f(X_t) - f(X_0) \\ &= \sum_{\{i:t_i<t\}} \Big(\big(f(X_{t_{i+1}-}) - f(X_{t_i})\big) + \big(f(X_{t_{i+1}}) - f(X_{t_{i+1}-})\big)\Big) \\ &= \sum_{\{i:t_i<t\}} (K_i^n + J_i^n), \end{aligned} \tag{14.5}$$

where

$$\begin{aligned} K_i^n &= f(X_{t_{i+1}-}) - f(X_{t_i}) + f'(X_{t_{i+1}-})\Delta X_{t_{i+1}} + \frac{1}{2} f''(X_{t_{i+1}-})(\Delta X_{t_{i+1}})^2, \\ J_i^n &= f(X_{t_{i+1}}) - f(X_{t_{i+1}-}) - f'(X_{t_{i+1}-})\Delta X_{t_{i+1}} - \frac{1}{2} f''(X_{t_{i+1}-})(\Delta X_{t_{i+1}})^2. \end{aligned}$$

We shall consider the remainder function $r : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x) - f(y) = f'(y)(x-y) + \frac{1}{2} f''(y)(x-y)^2 + r(x,y).$$

We treat each of the J and K terms separately.

For the J terms, consider the stopping times $T_k = \inf\{t : |X_t| + |X_{t-}| > k\}$ where $k \in \mathbb{N}$. As f is twice differentiable, on the interval $[\![0, T_k[\![$ we know

$$|J_i^n| = r(X_{t_{i+1}}, X_{t_{i+1}-}) \le \rho_k (\Delta X_{t_{i+1}})^2$$

for some constant ρ_k. Therefore, $\sum_{\{i:t_i<t\}} |J_i^n| \le \rho_k [X]$ is a.s. finite on $[\![0, T_k[\![$, and is uniformly bounded with respect to n. By dominated convergence, as π^n contains all jumps of size at least n^{-1}, it follows that, for $(t,\omega) \in [\![0, T_k[\![$,

$$\lim_{n\to\infty}\sum_{\{i:t_i<t\}} J_i^n = \sum_{s\in]0,t]} \Big(f(X_s)-f(X_{s-})-f'(X_{s-})\Delta X_s-\frac{1}{2}f''(X_{s-})(\Delta X_s)^2\Big). \tag{14.6}$$

Taking $k \to \infty$, we know $T_k \to \infty$, and so the above limit must hold for all t and almost all ω.

For the K terms, we can write

$$\begin{aligned} K_i^n &= f'(X_{t_i})(X_{t_{i+1}-} - X_{t_i}) + f'(X_{t_{i+1}-})\Delta X_{t_{i+1}} \\ &\quad + \frac{1}{2}\Big(f''(X_{t_{i+1}-})(\Delta X_{t_{i+1}})^2 + f''(X_{t_i})\big(X_{t_{i+1}-} - X_{t_i}\big)^2\Big) + r(X_{t_{i+1}-}, X_{t_i}). \end{aligned} \tag{14.7}$$

Taking the sum, we see

$$\begin{aligned} \sum_{\{i:t_i<t\}} K_i^n = &\sum_{\{i:t_i<t\}} f'(X_{t_i})(X_{t_{i+1}-} - X_{t_i}) + f'(X_{t_{i+1}-})\Delta X_{t_{i+1}} \\ &+ \frac{1}{2}\int_{]0,t]} f''(X_{s-}^{\pi^n})d[X]_s \\ &+ \frac{1}{2}\sum_{i:t_i<t} \Big(f''(X_{t_{i+1}-})(\Delta X_{t_{i+1}})^2 + f''(X_{t_i})\big(X_{t_{i+1}-} - X_{t_i}\big)^2\Big) \\ &- \frac{1}{2}\int_{]0,t]} f''(X_{s-}^{\pi^n})d[X]_s + \sum_{\{i:t_i<t\}} r(X_{t_{i+1}-}, X_{t_i}) \end{aligned}$$

Because $\{X_{s-}\}$ and $\{X_{s-}^{\pi^n}\}$ are locally bounded (uniformly in π^n), and $f'(X_{t_{i+1}-}) - f'(X_{t_i}) \to 0$ locally uniformly in n, in the same way as in the continuous case, stochastic and classical dominated convergence yields

$$\begin{aligned} &\sum_{\{i:t_i<t\}} \Big(f'(X_{t_i})(X_{t_{i+1}-} - X_{t_i}) + f'(X_{t_{i+1}-})\Delta X_{t_{i+1}}\Big) + \frac{1}{2}\int_{]0,t]} f''(X_{s-}^{\pi^n})d[X]_s \\ &\to \int_{]0,t]} f'(X_{s-})dX_s + \frac{1}{2}\int_{]0,t]} f''(X_{s-})d[X]_s \end{aligned} \tag{14.8}$$

and

$$\sum_{\{i:t_i<t\}} r(X_{t_{i+1}-}, X_{t_i}) \to 0.$$

By construction of π^n, we know $f''(X_{t_{i+1}-}) - f''(X_{t_i}) \to 0$ locally uniformly in n. Furthermore,

$$[X]_{t_{i+1}-} - [X]_{t_i} - (X_{t_{i+1}-} - X_{t_i})^2 \to 0$$

and, as X is càdlàg,

$$\sum_i \Big([X]_{t_{i+1}-} - [X]_{t_i} - (X_{t_{i+1}-} - X_{t_i})^2\Big) \le [X]_t + 2Q(\pi^n, t) + 2\sum_{s\le t}(\Delta X_s)^2.$$

We know $Q(\pi^n,\cdot) \to [X]$ in $\mathcal{S}$, so this sum is bounded, uniformly with respect to n. Dominated convergence then yields

$$
\begin{aligned}
&\frac{1}{2}\sum_{i:t_i<t}\Big(f''(X_{t_{i+1}-})(\Delta X_{t_{i+1}})^2 + f''(X_{t_i})\big(X_{t_{i+1}-}-X_{t_i}\big)^2\Big) - \frac{1}{2}\int_{]0,t]} f''(X^\pi)d[X]\\
&= \frac{1}{2}\sum_{\{i:t_i<t\}}\big(f''(X_{t_{i+1}-}) - f''(X_{t_i})\big)(\Delta X_{t_{i+1}})^2\\
&\quad + \frac{1}{2}\sum_{\{i:t_i<t\}} f''(X_{t_i})\big([X]_{t_{i+1}-} - [X]_{t_i} - (X_{t_{i+1}-}-X_{t_i})^2\big)\\
&\to 0. \qquad (14.9)
\end{aligned}
$$

Substituting (14.6), (14.7), (14.8) and (14.9) into (14.5), we know that, as $n\to\infty$, we have the limit in probability

$$
\begin{aligned}
&f(X_t) - f(X_0)\\
&= \sum_{\{i:t_i<t\}} K_i^n + \sum_{\{i:t_i<t\}} J_i^n\\
&\to \int_{]0,t]} f'(X_{s-})dX_s + \frac{1}{2}\int_{]0,t]} f''(X_{s-})d[X]_s\\
&\qquad + \sum_{s\in]0,t]}\Big(f(X_s) - f(X_{s-}) - f'(X_{s-})\Delta X_s - \frac{1}{2}f''(X_{s-})(\Delta X_s)^2\Big).
\end{aligned}
$$

As the right- and left-hand sides are independent of n, we have our result. □

The differentiation rule for a vector $\mathbb{R}^n$-valued semimartingale is established by the same method. However, the notation becomes very involved so the vector form of the theorem is now stated without proof.

Theorem 14.2.4. *Suppose X is a process with values in $\mathbb{R}^n$, each of whose components X^i is a semimartingale. Suppose $f:\mathbb{R}^n\to\mathbb{R}$ is a real valued twice continuously differentiable function. Then $f(X)$ is a semimartingale and*

$$
\begin{aligned}
f(X_t) &= f(X_0) + \sum_{i=1}^n \int_{]0,t]} \frac{\partial f}{\partial x_i}(X_{s-})dX_s^i\\
&\quad + \frac{1}{2}\sum_{i,j=1}^n \int_0^t \frac{\partial^2 f}{\partial x_i\partial x_j}(X_{s-})d\langle X^{ic}, X^{jc}\rangle_s\\
&\quad + \sum_{0<s\le t}\Big(f(X_s) - f(X_{s-}) - \sum_{i=1}^n \frac{\partial f}{\partial x_i}(X_{s-})\Delta X_s^i\Big).
\end{aligned}
$$

Remark 14.2.5. So far, we have considered only real-valued processes. However, it is natural to allow complex processes in many situations, for example in Theorem 14.4.1. In these cases, by using the map $a+bi\to(a,b)$, we can use Theorem 14.2.4 to show that Itô's rule applies also to complex-valued processes.

Remark 14.2.6. When X is vector valued, it is often convenient to write Itô's lemma in the differential form

$$df(X_t) = (\nabla f|_{X_t})dX + \frac{1}{2}\,\mathrm{Tr}((\nabla^2 f|_{X_t})d\langle X^c\rangle) \\ + f(X_s) - f(X_{s-}) - (\nabla f|_{X_t})\Delta X$$

where Tr is the trace operator (which commutes with the integral by linearity), $\langle X\rangle$ is the quadratic covariation matrix, and $\nabla f|_{X_t}$ and $\nabla^2 f|_{X_t}$ are the gradient and Hessian of f respectively, evaluated at X_t. With a further abuse of notation, this can also be expressed as

$$df(X_t) = (\nabla f|_{X_t})dX + \frac{1}{2}(dX^c)^\top(\nabla^2 f|_{X_t})(dX^c) \\ + f(X_s) - f(X_{s-}) - (\nabla f|_{X_t})\Delta X$$

where the product of differential terms $d(X^c)^i d(X^c)^j$ is formally defined to mean $d\langle (X^c)^i, (X^c)^j\rangle$. This notation is often easier to work with, and is natural by considering the second order Taylor expansion of f in a vector setting.

Remark 14.2.7. Expanding the notation in the previous remark, we obtain the *Box calculus*, which provides a set of formal rules for manipulating stochastic equations in a differential form, by treating the object 'dX' simply as a formal symbol. The basic rule of this calculus is that $dXdY := d[X, Y]$, whenever X and Y are semimartingales. As a consequence of this, if one of the processes is continuous and of finite variation, then $dXdY \equiv 0$, which enables quick simplification of many equations. One can easily verify that the basic rules of associativity and commutativity are satisfied by this calculus, that is

$$(a\,dX + b\,dY)^\top dZ = dZ^\top(a\,dX + b\,dY) = a\,d[X, Z] + b\,d[Y, Z],$$

for any $\mathbb{R}^{N\times N}$ valued (appropriately integrable) predictable processes a, b and $\mathbb{R}^N$-valued semimartingales X, Y, Z of constant dimension. The other rule of this calculus is the Itô rule as stated above. Together, these rules provide a direct method for manipulating stochastic equations in their differential form, which is formally equivalent to the integral definition of the stochastic calculus.

Remark 14.2.8. As noted above, for any semimartingales X and Y, we have

$$X_tY_t = X_0Y_0 + \int_{]0,t]} X_{s-}dY_s + \int_{]0,t]} Y_{s-}dX_s + [X, Y]_t.$$

If Y is predictable and of finite variation, then we also have the identity $[X, Y]_t = \sum_{0<s\leq t} \Delta X_s \Delta Y_s = \int_{]0,t]} \Delta Y_s dX_s$, so we can write

$$X_tY_t = X_0Y_0 + \int_{]0,t]} X_{s-}dY + \int_{]0,t]} Y_s dX_s.$$

Corollary 14.2.9. *Let X be a semimartingale and $f : [0,\infty[\times \mathbb{R} \to \mathbb{R}$ a $C^{1,2}$-function (i.e. f is once continuously differentiable in its first argument and twice continuously differentiable in its second). Then, $f(t, X_t)$ is a semimartingale, and writing f_t for the derivative in the first argument and f_x and f_{xx} for the first and second derivatives in the second argument, we have*

$$\begin{aligned} f(t, X_t) = f(0, X_0) &+ \int_{]0,t]} f_t(s, X_{s-})ds + \int_{]0,t]} f_x(s, X_{s-})dX_s \\ &+ \frac{1}{2}\int_{]0,t]} f_{xx}(s, X_{s-})d\langle X^c\rangle_s \\ &+ \sum_{0<s\leq t} \big(f(s, X_s) - f(s, X_{s-}) - f_x(s, X_{s-})\Delta X_s\big). \end{aligned}$$

Proof. If f is twice differentiable in its first argument, then this is the result of Theorem 14.2.4 applied to the vector semimartingale (t, X_t), as the process 't' is of finite variation. Otherwise, for each $\epsilon > 0$ we define the mollification

$$f^\epsilon(t, x) = \frac{1}{\epsilon}\int_{]t-\epsilon,t]} f(s, x)ds.$$

As f is continuously differentiable in t, we know $f^\epsilon \to f$ and

$$\frac{\partial}{\partial t} f^\epsilon(t, x) = \frac{f(t, x) - f(t-\epsilon, x)}{\epsilon} \to f_t(t, x)$$

as $\epsilon \to 0$, and both convergences are locally uniform. For each $\epsilon > 0$ we know f^ϵ is twice differentiable in both arguments, so applying Theorem 14.2.4 to f^ϵ gives the desired representation for f^ϵ. Localizing to ensure boundedness of integrands, taking $\epsilon \to 0$ and using dominated convergence (both classical and stochastic), we obtain the representation for f. □

14.3 The Tanaka–Meyer–Itô Rule

Itô's rule is a powerful tool provided the function f is twice differentiable. When this is not the case, the theorem does not hold. However, if f is continuous and convex, or more generally the difference of convex functions, then it is possible to derive a related result involving the local time process of X. The simple, yet fundamental, case when $f(x) = |x|$ was considered by Tanaka, while the general result is due to Meyer [133].

Remark 14.3.1. Suppose $f : \mathbb{R} \to \mathbb{R}$ is a convex function (or the difference of convex functions). Then it is classical (see for example Rockafellar [157, p.213]) that f is continuous and its left and right derivatives

$$\frac{df}{dx-} = \lim_{y\uparrow x}\frac{f(x) - f(y)}{x - y}, \qquad \frac{df}{dx+} = \lim_{y\downarrow x}\frac{f(x) - f(y)}{x - y}$$

are both well defined, but may differ. For convenience, we define

$$f' := \frac{1}{2}\Big(\frac{df}{dx-} + \frac{df}{dx+}\Big).$$

Furthermore, $f' = \frac{df}{dx-} = \frac{df}{dx+}$ except for countably many x. For further details on this theory, see Rockafellar [157, Part V].

We now give a version of Itô's rule, where the second order term is replaced by a process A.

Theorem 14.3.2. *Let f be a convex function and X a real-valued semimartingale. Then $f(X)$ is a semimartingale and there exists a continuous process $A \in \mathcal{A}^+_{0,\mathrm{loc}}$ such that*

$$\begin{aligned} f(X_t) = f(X_0) &+ \int_{]0,t]} f'(X_{u-})dX_u + A_t \\ &+ \sum_{0<s\leq t} \big(f(X_u) - f(X_{u-}) - f'(X_{u-})\Delta X_u\big) \end{aligned}$$

up to indistinguishability. Similarly for f the difference of two convex functions, in which case $A \in \mathcal{A}_{0,\mathrm{loc}}$.

Proof. The proof is similar to that of Corollary 14.2.9: we mollify f to give a twice differentiable function, apply Itô's rule, and take a limit.

Let ϕ be a nonnegative, twice differentiable, even function with support in $[-1,1]$ such that $\int_{[-1,1]} \phi(x)dx = 1$, for example, $\phi(x) \propto \exp(-1/(1-x^2))I_{\{|x|<1\}}$. For $\epsilon > 0$, let

$$f^\epsilon(x) = \frac{1}{\epsilon}\int_{\mathbb{R}} f(x+y)\phi(y/\epsilon)dy = \frac{1}{\epsilon}\int_{\mathbb{R}} f(y)\phi((y-x)/\epsilon)dy.$$

Integrating by parts, we see that

$$\frac{df^\epsilon}{dx} = -\frac{1}{\epsilon^2}\int_{\mathbb{R}} f(y)\phi'((y-x)/\epsilon)dy$$

and

$$\frac{d^2 f^\epsilon}{dx^2} = \frac{1}{\epsilon^3}\int_{\mathbb{R}} f(y)\phi''((y-x)/\epsilon)dy,$$

so f^ϵ is twice differentiable. As f is convex, we can see that f^ϵ is also convex, and so $\frac{d^2 f^\epsilon}{dx^2} \geq 0$. We can, therefore, apply Itô's rule to obtain

$$\begin{aligned} f^\epsilon(X_t) = f^\epsilon(X_0) &+ \int_{]0,t]} (f^\epsilon)'(X_{s-})dX_s + \frac{1}{2}\int_{]0,t]} (f^\epsilon)''(X_{s-})d\langle X^c\rangle_s \\ &+ \sum_{0<s\leq t} \big(f^\epsilon(X_s) - f^\epsilon(X_{s-}) - (f^\epsilon)'(X_{s-})\Delta X_s\big). \end{aligned} \tag{14.10}$$

Taking $\epsilon \to 0$, as f is continuous, we see that $f^\epsilon(x) \to f(x)$ for all x, and as f is convex and ϕ is even, $(f^\epsilon)'(x) \to f'(x)$ for all x. As $\{X_{s-}\}_{s\geq 0}$ is locally bounded, so are $\{f^\epsilon(X_{s-})\}_{s\geq 0}$ and $\{(f^\epsilon)'(X_{s-})\}_{s\geq 0}$ (uniformly in ϵ) and $\{f'(X_{s-})\}_{s\geq 0}$. Therefore, stochastic dominated convergence implies

$$\int_{]0,t]} (f^\epsilon)'(X_{s-})dX_s \to \int_{]0,t]} (f)'(X_{s-})dX_s$$

in ucp, and, as $\sum_{0<s\leq t}(\Delta X_s)^2 < \infty$, after localizing to obtain a bound we have

$$\begin{aligned}
&\sum_{0<s\leq t} \big(f^\epsilon(X_s) - f^\epsilon(X_{s-}) - (f^\epsilon)'(X_{s-})\Delta X_s\big) \\
&\quad \to \sum_{0<s\leq t} \big(f(X_s) - f(X_{s-}) - f'(X_{s-})\Delta X_s\big)
\end{aligned}$$

almost surely uniformly in t, which implies global ucp convergence of the sums.

Finally, taking the limit in ucp in (14.10) and rearranging allows us to define

$$\begin{aligned}
A_t &:= \lim_{\epsilon\to 0} \frac{1}{2} \int_{]0,t]} (f^\epsilon)''(X_{s-})d\langle X^c\rangle_s \\
&= f(X_t) - f(X_0) - \int_{]0,t]} f'(X_{s-})dX_s \\
&\qquad - \sum_{0<s\leq t} \big(f(X_s) - f(X_{s-}) - f'(X_{s-})\Delta X_s\big).
\end{aligned}$$

As the convergence is in ucp, we know that A is continuous (so locally integrable), and as f^ϵ is convex we see that A is an increasing real valued adapted function with $A_0 = 0$, that is, $A \in \mathcal{A}^+_{0,\mathrm{loc}}$.

The case of the difference of two convex functions follows immediately. □

Theorem 14.3.2 is not particularly useful unless we can obtain a better understanding of the process A. To do this we define the 'local time' process of a semimartingale, which is a generalization of Definition 9.3.1. We here use the convention that $\mathrm{sign}(x) := x/|x|$, where $0/0 := 0$.

Definition 14.3.3. *For X a semimartingale and $a \in \mathbb{R}$, we define L^a, the* local time of X at a, *to be the unique process in $\mathcal{A}^+_{0,\mathrm{loc}}$ such that*

$$\begin{aligned}
|X_t - a| = |X_0 - a| &+ \int_{]0,t]} \mathrm{sign}(X_{s-} - a)dX_s + L^a_t \\
&+ \sum_{0<s\leq t} \big(|X_s - a| - |X_{s-} - a| - \mathrm{sign}(X_{s-} - a)\Delta X_s\big).
\end{aligned}$$

Remark 14.3.4. Note that the process

$$\sum_{0<s\leq t} (|X_s - a| - |X_{s-} - a| - \operatorname{sign}(X_{s-} - a)\Delta X_s)$$

is a nondecreasing process which tracks the jumps of X which pass over a. Also, L^a is a continuous process for each a.

Lemma 14.3.5. *For any $a \in \mathbb{R}$, we have the ucp limit*

$$L_t^a = \lim_{\epsilon \to 0} \int_{]0,t]} \frac{1}{\epsilon} I_{\{|X_{s-}-a|<\epsilon\}} d\langle X^c \rangle_t.$$

Proof. For simplicity, suppose $a = 0$ without loss of generality. For fixed $\epsilon > 0$, consider the function

$$f^\epsilon(x) = \begin{cases} \frac{x^2}{2\epsilon} + \frac{\epsilon}{2} & \text{if } |x| < \epsilon, \\ |x| & \text{otherwise.} \end{cases}$$

We first show that Itô's rule remains valid for f^ϵ, even though it is not C^2. Define $g^\epsilon(x) = I_{\{|x|\neq\epsilon\}} \frac{d^2}{dx^2} f^\epsilon(x) = \epsilon^{-1} I_{\{|x|<\epsilon\}}$. We can easily check that f^ϵ is C^1, and by mollifying f^ϵ there exist approximations $f^{\epsilon,\eta} \in C^2$ such that, as $\eta \to 0$, $f^{\epsilon,\eta} \to f^\epsilon$ uniformly, $\frac{d}{dx} f^{\epsilon,\eta} \to \frac{d}{dx} f^\epsilon$ uniformly, $\frac{d^2}{dx^2} f^{\epsilon,\eta} \to g^\epsilon$ pointwise and $\frac{d^2}{dx^2} f^{\epsilon,\eta}$ is uniformly bounded in both x and η.

Clearly, Itô's rule holds for the C^2 functions $f^{\epsilon,\eta}$. By stochastic dominated convergence, taking $\eta \to 0$ we observe that Itô's rule remains valid for f^ϵ, with g^ϵ playing the role of the second derivative, for any $\epsilon > 0$. We can therefore repeat the argument of the proof of Theorem 14.3.2 (starting at (14.10)), with f^ϵ as an approximation of $f(x) = |x|$. The result follows. □

Remark 14.3.6. A consequence of this result is that L^a is adapted to the sub-filtration generated by $|X - a|$ and $\langle X^c \rangle$. (See Example 18.0.1 to see why this is significant.)

Lemma 14.3.7. *Let X be a semimartingale and L^a its local time at a. Then*

$$\begin{aligned}
&(X_t - a)^+ \\
&= (X_0 - a)^+ + \int_{]0,t]} \Big(I_{\{X_{s-}>a\}} + \frac{1}{2} I_{\{X_{s-}=a\}} \Big) dX_s + \frac{1}{2} L_t^a \\
&\quad + \sum_{0<s\leq t} \Big((X_s - a)^+ - (X_{s-} - a)^+ - \big(I_{\{X_{s-}>a\}} + \frac{1}{2} I_{\{X_{s-}=a\}} \big) \Delta X_s \Big)
\end{aligned}$$

and

$$\begin{aligned}
&(X_t - a)^- \\
&= (X_0 - a)^- - \int_{]0,t]} \Big(I_{\{X_{s-}<a\}} + \frac{1}{2} I_{\{X_{s-}=a\}} \Big) dX_s + \frac{1}{2} L_t^a \\
&\quad + \sum_{0<s\leq t} \Big((X_s - a)^- - (X_{s-} - a)^- + \big(I_{\{X_{s-}<a\}} + \frac{1}{2} I_{\{X_{s-}=a\}} \big) \Delta X_s \Big).
\end{aligned}$$

Proof. Simply apply Theorem 14.3.2 with $f(x) = (x-a)^+$ and $f(x) = (x-a)^-$ to obtain the stated equations with $\frac{1}{2}L^a$ replaced for the moment by some processes $A^{(+)}$ and $A^{(-)}$ in $\mathcal{A}^+_{0,\mathrm{loc}}$. As $(x-a)^+ + (x-a)^- = |x-a|$ we see that $A^{(+)} + A^{(-)} = L^a$. As $(x-a)^+ - (x-a)^- = x-a$, we see that $A^{(+)} - A^{(-)} = 0$. The result follows. $\square$

The following result generalizes and strengthens Theorem 9.3.2.

Theorem 14.3.8. *Let X be a semimartingale and $a \in \mathbb{R}$. Then the measure on the predictable σ-algebra (Σ_p) induced by the finite variation process L^a is supported on the boundary of the set $\{(t,\omega) : X_{t-}(\omega) = a\}$.*

Proof. Let S, T be two stopping times such that $S \leq T$ and $[\![S,T[\![\subset \{X_{s-} < a\}$. Then, by Lemma 14.3.7, we see

$$0 = (X_T - a)^+ - (X_S - a)^+ = \frac{1}{2}(L^a_T - L^a_S),$$

so L^a is constant on any interval $[\![S,T[\![\subset \{X_{s-} < a\}$. For $r \in \mathbb{Q}$, taking $S^r = rI_{\{X_{r-}<a\}} + \infty I_{\{X_{r-}\geq a\}}$ and $T^r = \inf\{t > S^r : X_{t-} \geq a\}$, we know L^a does not charge the set $\cup_{r\in\mathbb{Q}}[\![S^r, T^r[\![$. As the set $\{X_{s-} < a\}$ is open on the left, we see that $\cup_{r\in\mathbb{Q}}[\![S^r, T^r[\![$ differs from $\{X_{s-} < a\}$ on at most countably many points. As L^a is continuous, it cannot charge any countable set, so L^a does not charge $\{X_{s-} < a\}$. In the same way, by considering $(X-a)^-$ we see that L^a cannot charge the set $\{X_{s-} > a\}$.

Now suppose $[\![S,T]\!] \subset \{X_{s-} = a\}$, which implies $[\![S,T[\![\subset \{X_s = a\}$. By Theorem 14.3.2 we see that

$$|X_T - a| = |X_S - a| + L^a_T - L^a_S + |X_T - a| - |X_{T-} - a| - \mathrm{sign}(X_{T-} - a)\Delta X_T,$$

and so $L^a_T - L^a_S = 0$. Taking $S^r = rI_{\{X_{r-}=a\}} + \infty I_{\{X_{r-}\geq a\}}$ and $T^r = \inf\{t > S^r : X_{t-} \neq a\}$, we see that L^a does not charge $\cup_{r\in\mathbb{Q}}[\![S^r, T^r]\!]$, which equals the interior of the set $\{X_{s-} = a\}$. Therefore, L^a is supported on the boundary of the set $\{X_{s-} = a\}$, as desired. $\square$

As L^a has been constructed for each a separately, we now show that there is a version of it with useful continuity (and hence measurability) properties in a and t jointly. To do this, we refer to the previously established stochastic Fubini theorem.

Theorem 14.3.9. *For X a semimartingale, we can find a $\mathcal{B}(\mathbb{R}) \otimes \Sigma_o$-measurable function $L : \mathbb{R} \times [0,\infty[\times \Omega \to \mathbb{R}$ such that $L(a,\cdot,\cdot)$ is indistinguishable from the local time of X at a for every a.*

Proof. Apply the stochastic Fubini theorem (Theorem 12.4.18) to the map $(a, t, \omega) \mapsto \int_{]0,t]} \mathrm{sign}(X_s - a) dX_s$, to obtain a version which is measurable in a. With this we can define

$$L(a, t, \cdot) = |X_t - a| - |X_0 - a| - \int_{]0,t]} \mathrm{sign}(X_{s-} - a) dX_s \\ - \sum_{0<s\le t} \big(|X_s - a| - |X_{s-} - a| - \mathrm{sign}(X_{s-} - a)\Delta X_s\big),$$

and so see that L is $\mathcal{B}(\mathbb{R}) \otimes \Sigma_o$-measurable. □

Remark 14.3.10. Our construction shows that $L(a, t, \omega)$ is almost surely continuous in t for each value of a. However, one can also find a version which is a.s. continuous in t for every a simultaneously, and is $\mathcal{B}(\mathbb{R}) \otimes \Sigma_p$-measurable.

We now note one further property of convex functions. For a function f of bounded variation, taking an arbitrary bounded, twice differentiable function with compact support ϕ, by twice applying integration by parts we can define the 'second derivative' measure ρ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ by the property

$$\int \phi d\rho = -\int \phi' d(f') = \int \phi'' df.$$

If f is convex and finite, f' is a finite increasing function, so we see that ρ is a σ-finite positive measure. If f is twice differentiable, we see that $\rho\big(]a, b]\big) = f''(b) - f''(a)$. This clearly extends to when f is the difference of convex functions, in which case ρ is a σ-finite signed measure.

Using this, we can give an explicit expression for the process A in Theorem 14.3.2, in terms of the local time of X.

Theorem 14.3.11 (Tanaka–Meyer–Itô Rule). *Suppose X is a semimartingale. Let f be the difference of two convex functions, f' as in Remark 14.3.1 and ρ its second derivative measure. Then*

$$f(X_t) = f(X_0) + \int_{]0,t]} f'(X_{s-}) dX_s + \frac{1}{2}\int_{\mathbb{R}} L_t^a \rho(da) \\ + \sum_{0<s\le t} \big(f(X_s) - f(X_{s-}) - f'(X_{s-})\Delta X_s\big).$$

Proof. First note that, if f is linear, then by the simple Itô rule, the result holds with $\rho \equiv 0$.

For a general function f, let $g(x) = \frac{1}{2}\int |x - y|\rho(dy)$. Then g is a convex function, and its second derivative measure is also ρ. Therefore $f - g$ has second derivative zero, hence is linear, and so the result holds for the function $f - g$. By linearity, it remains to show that the result holds for the function g.

By integrating $\frac{1}{2}|X_t - a|$ with respect to $\rho(da)$, we see that

$$
\begin{aligned}
g(X_t) &= \frac{1}{2}\int_{\mathbb{R}} |X_t - a|\rho(da) \\
&= \frac{1}{2}\int_{\mathbb{R}} |X_0 - a|\rho(da) + \frac{1}{2}\int_{\mathbb{R}}\int_{]0,t]} \Big(\operatorname{sign}(X_{s-} - a)dX_s + L_t^a\Big)\rho(da) \\
&\quad + \frac{1}{2}\int_{\mathbb{R}} \sum_{0<s\le t} \big(|X_s - a| - |X_{s-} - a| - \operatorname{sign}(X_{s-} - a)\Delta X_s\big)\rho(da) \\
&= g(X_0) + \frac{1}{2}\int_{\mathbb{R}}\int_{]0,t]} \Big(\operatorname{sign}(X_{s-} - a)dX_s + L_t^a\Big)\rho(da) \\
&\quad + \sum_{0<s\le t} \Big(g(X_s) - g(X_{s-}) - \Big(\int_{\mathbb{R}} \operatorname{sign}(X_{s-} - a)\rho(da)\Big)\Delta X_s\Big).
\end{aligned}
$$

By monotone convergence, we have

$$g'(x) = \int_{\mathbb{R}} \operatorname{sign}(x - a)\rho(da)$$

and hence, by Theorem 12.4.18,

$$\frac{1}{2}\int_{\mathbb{R}}\int_{]0,t]} \Big(\operatorname{sign}(X_{s-} - a)dX_s + L_t^a\Big)\rho(da) = \int_{]0,t]} g'(X_{s-})dX_s + \frac{1}{2}\int_{\mathbb{R}} L_t^a \rho(da).$$

Combining these yields the result. □

Corollary 14.3.12. *For all Borel measurable functions $g : \mathbb{R} \to \mathbb{R}$,*

$$\int_{\mathbb{R}} L_t^a g(a)da = \int_{]0,t]} g(X_{u-})d\langle X^c\rangle_u$$

up to indistinguishability.

Proof. If $g = f''$ with f a twice continuously differentiable function, then the result follows from comparing Theorem 14.3.11 with Theorem 14.2.3. Now consider a sequence of twice differentiable functions which converge uniformly on compacts. By stopping at times $\inf\{t : |X_{t-}| \ge n\}$ we can apply the dominated convergence theorem to obtain the desired equality. □

14.4 Lévy's Characterization of Brownian Motion

We now turn our attention to some of the basic consequences of Itô's rule. We first use it to prove Lévy's characterization of Brownian motion, and a similar result for a Poisson process.

Theorem 14.4.1. *Suppose B is a continuous local martingale on the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\ge 0}, P)$ such that $\{B_t^2 - t\}_{t\ge 0}$ is a local martingale (equivalently, that $\langle B\rangle_t = t$). Then B is a Brownian motion.*

Proof. We need to show that $B_t - B_s \sim N(0, t-s)$ and is independent of $\mathcal{F}_s$. A simple way to do this, using the characteristic function of the normal distribution, is to show that for any $u \in \mathbb{R}$,

$$E[e^{iu(B_t - B_s)}|\mathcal{F}_s] = e^{-u^2(t-s)/2},$$

where $i = \sqrt{-1}$. Apply Itô's rule to the complex-valued process e^{iuB_t}. As we have assumed that $B_t^2 - t$ is a local martingale, we know $\langle B \rangle_t = t$, hence B is a martingale, and is locally square integrable. As B is continuous, this yields

$$e^{iuB_t} = 1 + iu \int_{]0,t]} e^{iuB_s} dB_s - \frac{u^2}{2} \int_{]0,t]} e^{iuB_s} ds.$$

As $|e^{iuB_s}| < 1$ for all real u and B_s, we know that $\int_{]0,t]} e^{iuB_s} dB_s$ is a true (complex) martingale. Rearranging, we have

$$e^{iuB_t} = e^{iuB_s} - \frac{u^2}{2} \int_{]s,t]} e^{iuB_r} dr + iu \int_{]s,t]} e^{iuB_r} dB_r. \tag{14.11}$$

Consider the random function $g(\omega, s) = E[e^{iuB_t}|\mathcal{F}_s]$. As $g(\omega, \cdot)$ is a martingale, we see that g admits a modification which is right continuous in s. Using this modification, the conditional expectation of (14.11) gives

$$g(\omega, t) = g(\omega, s) - \frac{u^2}{2} \int_{]s,t]} g(\omega, r) dr$$

up to indistinguishability. For almost all ω, this is an ordinary differential equation with solution, for $t \geq s$,

$$g(\omega, t) = g(\omega, s) e^{-u^2(t-s)/2}.$$

Therefore,

$$E[e^{iu(B_t - B_s)}|\mathcal{F}_s] = \frac{g(\omega, t)}{g(\omega, s)} = e^{-u^2(t-s)/2},$$

as desired. □

The following extension can be proven the same way.

Corollary 14.4.2. *Suppose $B = (B^1, \ldots, B^n)$ is a continuous vector local martingale with values in $\mathbb{R}^n$ such that for all i, j,*

$$\langle B^i, B^j \rangle_t = \delta^{ij} t.$$

Then B is a standard n-dimensional Brownian motion (that is, a vector of independent Brownian motions).

As a consequence, we can derive the following representation of all continuous martingales, which complements Exercise 5.7.11.

Definition 14.4.3. *A family* $\{C_s\}_{s\geq 0}$ *is called a* time-change *if* C_s *is a stopping time for every* s*, and* $s \to C_s$ *is an a.s. increasing càdlàg map.*

Theorem 14.4.4 (Dambis–Dubins–Schwarz Theorem). *Let* M *be a continuous* $\{\mathcal{F}_t\}_{t\geq 0}$*-local martingale such that* $\langle M\rangle_\infty = \infty$ *and* $M_0 = 0$*. If we write* $C_t = \inf\{s : \langle M\rangle_s \geq t\}$*, we know that* C *is a time-change, and*

$$B_t := M_{C_t}$$

defines an $\{\mathcal{F}_{C_t}\}_{t\geq 0}$*-Brownian motion.*

Proof. As M is continuous, we know that $\langle M\rangle$ is continuous increasing. Therefore it is easy to verify that C is a time-change and that B is an $\{\mathcal{F}_{C_t}\}_{t\geq 0}$-local martingale. As M is constant on any interval where $\langle M\rangle$ is constant (Exercise 11.7.11), we see that B is continuous, and that $\langle M\rangle_{C_t} = t$. Therefore, as $M^2 - \langle M\rangle$ is an $\{\mathcal{F}_t\}_{t\geq 0}$-local martingale, we see that $\left\{M_{C_t}^2 - \langle M\rangle_{C_t}\right\}_{t\geq 0} = \left\{B_t^2 - \langle M\rangle_{C_t}\right\}_{t\geq 0}$ is an $\{\mathcal{F}_{C_t}\}_{t\geq 0}$-local martingale, and so $\langle B\rangle_t = \langle M\rangle_{C_t} = t$. By Lévy's characterization (Theorem 14.4.1), it follows that B is a Brownian motion. □

14.4.1 Poisson Processes

A similar result can be obtained for the Poisson process. Recall that a purely discontinuous local martingale Q is one which is orthogonal to every continuous local martingale, or equivalently where $\langle Q^c\rangle \equiv 0$.

Theorem 14.4.5. *Suppose* $\{Q_t\}_{t\geq 0}$ *is a purely discontinuous local martingale on a filtered probability space, all of whose jumps equal* $+1$*. If* $\{Q_t^2 - \lambda t\}_{t\geq 0}$ *is a local martingale for some* $\lambda > 0$*, then* $N_t = Q_t - Q_0 + \lambda t$ *defines a Poisson process with parameter* λ*.*

Proof. We can suppose $Q_0 = N_0 = 0$. By assumption, we know that $\langle Q\rangle = \lambda t$, and hence Q is a martingale, and is locally square integrable. Clearly $E\big[[Q]_t\big] = E[Q_t^2] = \lambda t < \infty$, but because $\Delta Q_s = +1$, and Q is purely discontinuous

$$[Q]_t = \sum_{0<s\leq t} \Delta Q_s^2 = \sum_{0<s\leq t} \Delta Q_s.$$

and we see that Q is a local martingale of locally integrable variation, and that Q has a.s. finitely many jumps on any finite interval. Writing $N_t = \sum_{s\leq t} \Delta Q_s$, as $[Q]_t$ is integrable we see that N_t is integrable. Furthermore, Q is a compensated sum of jumps, by Theorem 10.2.6, so $Q = N - (\Pi_p^* N)$. However,

$$N_t - \lambda t = [Q]_t - \lambda t = (Q_t^2 - \lambda t) + ([Q]_t - Q_t^2),$$

so $\{N_t - \lambda t\}_{t\geq 0}$ is a local martingale. Consequently, $\Pi_p^* N_t = \lambda t$, that is, $Q_t = N_t - \lambda t$, as stated in the theorem.

Applying the differentiation rule to the process e^{iuQ_t}, if S and T denote consecutive jumps of Q we have

$$e^{iuQ_T} = e^{iuQ_S} + iu \int_{]S,T]} e^{iuQ_{r-}} dQ_r + (e^{iuQ_T} - e^{iuQ_{T-}} - iue^{iuQ_{T-}} \Delta Q_T).$$

As $|e^{iuQ_{v-}}| < 1$ we know the stochastic integral is a true martingale. As $Q_{T-} = Q_S - \lambda(T-S)$ and $\Delta Q_T = 1$, taking the conditional expectation with respect to $\mathcal{F}_S$ and rearranging we have

$$E[e^{-iu\lambda(T-S)}|\mathcal{F}_S] = (1 + iut)^{-1}.$$

Therefore, $T - S$ is independent of $\mathcal{F}_S$ and is exponentially distributed with parameter λ. By Theorem 5.5.22, we see that N is a Poisson process. □

As before, we can obtain a representation of all counting processes, similar to the Dambis–Dubins–Schwarz theorem

Theorem 14.4.6. *Let M be a purely discontinuous $\{\mathcal{F}_t\}_{t\geq 0}$-local martingale such that $\langle M\rangle_\infty = \infty$, $\langle M\rangle$ is continuous, $M_0 = 0$ and $\Delta M \in \{0, 1\}$. For any $\lambda > 0$, if we write $C_t = \inf\{s : \langle M\rangle_s \geq \lambda t\}$, we know that C is a time-change, and*

$$N_t = M_{C_t} + C_t$$

defines an $\{\mathcal{F}_{C_t}\}_{t\geq 0}$-Poisson process with parameter λ.

Proof. This follows almost exactly as in Theorem 14.4.4, the only difference being that continuity is replaced by pure discontinuity. □

14.5 The Martingale Representation Theorem

A key result in the theory of martingales is the "Martingale Representation Theorem". This allows us to say that all martingales, in a certain space, can be represented in terms of stochastic integrals. We have already seen special cases of this result for Poisson Processes (Exercise 8.4.9) and more generally for random measures with deterministic compensators (Section 13.6). In this section, we will give a version of this result for the Brownian motion, based on an elegant argument due to Dellacherie (our presentation follows Davis [45]), and then an extension for a sequence of Brownian motions and a random measure simultaneously. A more abstract and general result, the Jacod–Yor Theorem, is given in Section 18.3. Counterexamples are given in Exercises 18.4.4 and 18.4.5.

This result does not directly use Itô's rule; however, the proof we use requires[3] us to know $\{B_t^2 - t\}_{t\geq 0} = 2B \bullet B$. As this result is most easily seen by applying Itô's rule, we have left the argument until now.

[3] Alternatively, if we knew that the optional quadratic variation of a semimartingale is the same under all equivalent probability measures (Lemma 15.2.4), this can also be used to construct a similar argument.

Theorem 14.5.1. *Suppose* $\{B_t\}_{t\geq 0}$ *is a Brownian motion on the filtered probability space* $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$. *Write*

$$\mathcal{G}_t^0 = \sigma\{B_s : s \leq t\}$$

and $\{\mathcal{G}_t\}_{t\geq 0}$ *for the right-continuous completion of* $\{\mathcal{G}_t^0\}_{t\geq 0}$.

Then every random variable $X \in L^2(\Omega, \mathcal{G}_\infty)$ *can be represented as a stochastic integral*

$$X = E[X|\mathcal{G}_0] + \int_{]0,\infty[} H_s dB_s \,,$$

where $\{H_t\}_{t\geq 0}$ *is a* $\{\mathcal{G}_t\}_{t\geq 0}$*-predictable process with* $E[\int_{]0,\infty[} H_s^2 ds] < \infty$ *(that is,* $H \in L^2(B)$*). The process* H *is unique up to equality in* $L^2(B)$ *(equivalently,* $dt \times dP$*-a.e.). Furthermore,*

$$E[X|\mathcal{G}_t] = E[X|\mathcal{F}_t] = E[X|\mathcal{G}_0] + \int_{]0,t]} H_s dB_s \,.$$

Before proving this theorem, we will prove the following lemma.

Lemma 14.5.2. *In the setting of Theorem 14.5.1, every* $\{\mathcal{G}_t\}_{t\geq 0}$*-stopping time is* $\{\mathcal{G}_t\}_{t\geq 0}$*-predictable, and hence* $\{\mathcal{G}_t\}_{t\geq 0}$ *is quasi-left continuous, that is,* $\mathcal{G}_{T-} = \mathcal{G}_T$ *for every (predictable) stopping time. Furthermore, every* $\{\mathcal{G}_t\}_{t\geq 0}$ *local martingale has continuous paths.*

Proof. Suppose T is any $\{\mathcal{G}_t\}_{g\geq 0}$-stopping time and write

$$p_t = I_{\{t\geq T\}} \,.$$

Then $\{p_t\}_{t\geq 0}$ is a submartingale, and it has a Doob–Meyer decomposition $p_t = r_t + q_t$, where q is a martingale and r is a predictable increasing process with jumps bounded by 1. In particular, $q \leq 1 - r \leq 1$. Write

$$\tau_n = \inf\{t : r_t \geq n\},$$

and

$$L = 1 - \frac{1}{2}\, q_\infty^{\tau_n} \,.$$

Then $L \geq \frac{1}{2}$ a.s. and $E[L] = 1$, so a new probability measure Q can be defined on $(\Omega, \mathcal{G}_\infty)$ by putting $\frac{dQ}{dP} = L$, so that P and Q are equivalent measures. Now q^{τ_n} is a square integrable martingale, and, as a compensated jump process, it is orthogonal to every continuous martingale. Therefore, in particular, it is orthogonal to B and $\{B_t^2 - t\}_{t\geq 0}$. In other words, $\{q_t^{\tau_n} B_t\}_{t\geq 0}$ and $\{q_t^{\tau_n}(B_t^2 - t)\}_{t\geq 0}$ are martingales, so B and $\{B_t^2 - t\}_{t\geq 0}$ are martingales under the measure Q. Therefore, B is a Brownian motion under measure Q as well as measure P.

By the Doob–Dynkin lemma (Lemma 1.3.12), any $\mathcal{G}_\infty$-measurable random variable X can be written

$$X = f(\{B_s : s \leq \infty\}) \qquad \text{a.s.}$$

for some measurable[4] function $f : \mathbb{R}^\mathbb{R} \to \mathbb{R}$. However, B is a Brownian motion under both P and Q, so P and Q must place the same measure on any set of paths of B, and hence X has the same law under P and Q. It follows that $E^P[X] = E^Q[X]$, that is, P and Q coincide on $\mathcal{G}_\infty$. Consequently, $q_\infty^{\tau_n} = 0$ a.s., so p and r are indistinguishable. Therefore, p is a predictable process and we see that T is a predictable stopping time.

As we have shown that every accessible stopping time is predictable, the quasi-left continuity of $\{\mathcal{G}_t\}_{t\geq 0}$ follows by Theorem 6.4.2. In particular, we see that $\mathcal{G}_T = \mathcal{G}_{T-}$ for any stopping time T.

By Theorem 5.6.13, for any uniformly integrable $\{\mathcal{G}_t\}_{t\geq 0}$ martingale M,

$$M_{T-} = E[M_T|\mathcal{G}_{T-}] = E[M_T|\mathcal{G}_T] = M_T\,.$$

As all local martingales are locally uniformly integrable, and localization cannot alter the almost sure continuity of a process, this implies that every $\{\mathcal{G}_t\}_{t\geq 0}$ local martingale has continuous paths. □

Proof of Theorem 14.5.1. We can restrict our attention to the filtration $\{\mathcal{G}_t\}_{t\geq 0}$, and by subtraction we can suppose that $X_0 := E[X|\mathcal{G}_0] = 0$.

Define $X_t = E[X|\mathcal{G}_t]$. For any $n \in \mathbb{N}$, we write Y_t^n for the projection of the square integrable martingale X onto the stable subspace of $\mathcal{H}^2$ generated by $\{B_{t\wedge n}\}_{t\geq 0}$. From Theorem 12.2.7, we know that $Y_t^n = H^n \bullet B$ for some predictable process $H^n \in L^2(B)$.

From the tower property of conditional expectation, for $s \leq n$ we have that $H_s^n = H_s^{n+1}$, so by pasting we obtain a $\{\mathcal{G}_t\}_{t\geq 0}$-predictable integrand $H \in L^2(B)$ and a process Y defined by $Y_t = \int_{[0,t]} H_s dB_s$. By construction, we have that $\langle X - Y, B\rangle = 0$.

It remains to be shown that this implies that $X - Y \equiv X_\infty - Y_\infty = 0$ a.s. We know our martingale $M := X - Y$ is continuous. Write

$$\sigma_n = \inf\{t : |M_t| \geq n\},$$

and

$$M_t^n = \frac{1}{2n} M_{t\wedge\sigma_n}\,.$$

Then $|M^n| \leq \frac{1}{2}$ and M is orthogonal to B and hence to $\{B_t^2 - t\}_{t\geq 0} = 2B\bullet B$. Writing

$$L = 1 + M_\infty^n,$$

the same argument as in the proof of Lemma 14.5.2 shows that $M_\infty^n = 0$ a.s. Therefore, letting $n \to \infty$, $M \equiv M_\infty = 0$ a.s.

[4] Where, for example, $\mathbb{R}^\mathbb{R}$ has the Borel cylinder σ-algebra.

To show uniqueness, we simply note that if

$$X = E[X|\mathcal{G}_0] + \int_{]0,\infty[} H_s dB_s = E[X|\mathcal{G}_0] + \int_{]0,\infty[} \tilde{H}_s dB_s$$

for two processes $H, \tilde{H}$, then

$$E\Big[\Big(\int_{]0,\infty[} (H_s - \tilde{H}_s) dB_s\Big)^2\Big] = E\Big[\int_{]0,\infty[} (H_s - \tilde{H}_s)^2 ds\Big] = 0,$$

that is, $H = \tilde{H}$ up to equality in $L^2(B)$.

Finally, suppose $\{Z_t\}_{t\geq 0}$ is a right continuous version of the martingale $\big\{E[X|\mathcal{G}_t]\big\}_{t\geq 0}$. Then, by the above result, Z has a representation as a stochastic integral with respect to B. However, B is an $\{\mathcal{F}_t\}_{t\geq 0}$-martingale, so Z is also a square integrable $\{\mathcal{F}_t\}_{t\geq 0}$-martingale. Therefore, the limit $Z_\infty = \lim_{t\to\infty} Z_t$ exists both almost surely and in L^2, and $Z_t = E[Z_\infty|\mathcal{F}_t]$. However, X is $\mathcal{G}_\infty = \mathcal{G}_{\infty-}$-measurable, so, applying the martingale convergence theorem with respect to the filtration $\{\mathcal{G}_t\}_{t\geq 0}$, we see that $\lim_{t\to\infty} Z_t = X = Z_\infty$. Therefore,

$$E[X|\mathcal{G}_t] = E[X|\mathcal{F}_t] \qquad \text{a.s.}$$

for all $t \geq 0$. □

Remark 14.5.3. Suppose $B = \{B^1, \dots, B^n\}$ is an n-dimensional Brownian motion and that

$$\mathcal{G}_t = \sigma\{B_s : s \leq t\}.$$

The same kind of argument shows that if X is a square integrable $\{\mathcal{G}_t\}_{t\geq 0}$-martingale, then there are $\{\mathcal{G}_t\}_{t\geq 0}$-predictable processes $\{H^i\}_{i=1,\dots,n}$ with $E[\int (H^i_s)^2 ds] < \infty$ such that

$$X_t = E[X_\infty|\mathcal{G}_0] + \sum_{i=1}^n \int_{]0,t]} H^i_s dB^i_s.$$

(This can also be seen as a special case of Theorem 14.5.7.)

Corollary 14.5.4. *Let B and $\{\mathcal{G}_t\}_{t\geq 0}$ be as in Theorem 14.5.1. Then every $\{\mathcal{G}_t\}_{t\geq 0}$-local martingale X can be expressed in the form*

$$X = X_0 + H \bullet B$$

for some $H \in L^2_{\mathrm{loc}}(B)$.

Proof. From Lemma 14.5.2, we know that every local martingale is continuous. Therefore, every local martingale is in $\mathcal{H}^2_{\mathrm{loc}}$. From Theorem 14.5.1, it follows that if $\{T_n\}_{n\in\mathbb{N}}$ is a sequence of stopping times with $T_n \to \infty$ such that $X^{T_n} \in \mathcal{H}^2$, then

$$X^{T_n} = X_0 + H^n \bullet B$$

for some process $H^n \in L^2(B)$. By uniqueness of this representation, we see $H^n = H^{n+1}$ on $[\![0, T^n]\!]$, and so pasting yields the result. □

Definition 14.5.5. *As when we considered a random measure (Theorem 13.6.1), we say B has the* predictable representation property *(in $(\{\mathcal{G}_t\}_{t\geq 0}, P)$) when a result of the type of Corollary 14.5.4 holds, that is, when every $\{\mathcal{G}_t\}_{t\geq 0}$-local martingale can be written as a stochastic integral with respect to B.*

Given the Dambis–Dubins–Schwarz theorem (Theorem 14.4.4), the following extension of this result also holds.

Corollary 14.5.6. *Let Y be a continuous local martingale with deterministic quadratic variation. Let $\{\mathcal{G}_t\}_{t\geq 0}$ be the filtration generated by Y. Then every $\{\mathcal{G}_t\}_{t\geq 0}$-local martingale X can be expressed in the form*

$$X = X_0 + H \bullet Y$$

for some $H \in L^2_{\mathrm{loc}}(Y)$.

Proof. This follows from Theorem 14.4.4, as we can deterministically change time to make Y a Brownian motion, and then apply Theorem 14.5.1. □

14.5.1 General Lévy-Type Setting

As we have a martingale representation result for Brownian motions and random measures with deterministic compensators, it is natural to ask how we might extend this to the setting of a general Lévy process. This can be achieved through a similar monotone class argument to that in Section 13.6. In fact, we here present a more general argument for the filtration generated by an integer valued random measure with deterministic continuous compensator and a sequence of independent Brownian motions.

Theorem 14.5.7. *Let $W = \{W^i\}_{i\in\mathbb{N}}$ be a sequence of independent one-dimensional Brownian motions, and $\mu \in \tilde{\mathcal{A}}^1_\sigma$ an integer valued random measure, with deterministic continuous compensator μ_p. Then any local martingale M in the (right-continuous, completed) filtration $\{\mathcal{F}_t\}_{t\geq 0}$ generated by $\{\mu, W^1, W^2, \dots\}$ can be written*

$$M = Z^\top \bullet W + H * \tilde{\mu},$$

for some stochastically integrable processes $H, Z = \{Z^i\}_{i\in\mathbb{N}}$ of appropriate dimension. As before, we say that (W, μ) has the predictable representation property.

Proof. By Lemma 13.6.5, it is enough to consider the case where $M_\infty = I_B$ for some set B. By Theorem 14.5.1, we know that the result holds whenever $B \in \mathcal{F}^{(i)}_\infty$ for some $i \geq 0$, where $\{\mathcal{F}^{(i)}_t\}_{t\geq 0}$ is the filtration generated by W^i for $i \geq 1$ and $\{\mathcal{F}^{(0)}_t\}_{t\geq 0}$ is the filtration generated by μ.

Let $\mathcal{C}$ be the family of sets $B \in \mathcal{F}_\infty$ for which we have the desired representation. Suppose $B = B^i \cap B^j$, for some $B^i \in \mathcal{F}_\infty^{(i)}$ and $B^j \in \mathcal{F}_\infty^{(j)}$. Then, as we can represent $E[I_{B^i}|\mathcal{F}_{(\cdot)}^{(i)}]$ and $E[I_{B^j}|\mathcal{F}_{(\cdot)}^{(i)}]$ we know there exists $\alpha^i, \beta^i, \alpha^j, \beta^j$ of appropriate integrability and dimensions that

$$I_{B^i \cap B^j} = I_{B^i} I_{B^j} = (\alpha^i + \beta^i \bullet W^i)(\alpha^j + \beta^j \bullet W^j),$$

(and similarly with $\beta^i \bullet W^i$ replaced with $\beta^i * \tilde{\mu}$ if $i = 0$). By Itô's product rule (Theorem 14.1.1), as our Brownian motions are independent (so $\langle W^i, W^j \rangle = 0$) and $\tilde{\mu}$ is purely discontinuous,

$$\begin{aligned} I_{B^i \cap B^j} = \alpha^i \alpha^j &+ ((\alpha^j + \beta^j \bullet W^j)\beta^i) \bullet W^i \\ &+ ((\alpha^i + \beta^i \bullet W^i)\beta^j) \bullet W^j \end{aligned}$$

and similarly with $\beta^i \bullet W^i$ replaced with $\beta^i * \tilde{\mu}$ if $i = 0$, which gives our desired representation. By linearity of the integral, we also have a representation for $I_{B^i \cup B^j} = I_{B^i} + I_{B^j} - I_{B^i \cap B^j}$, so $\mathcal{C}$ contains the algebra of sets given by finite intersections and unions of sets in $\bigcup_{i \in \mathbb{N} \cup \{0\}} \mathcal{F}_\infty^{(i)}$. As in the proof of Theorem 13.6.7, the Itô isometry then shows $\mathcal{C}$ forms a monotone class, so it follows that $\mathcal{C}$ contains all of $\bigvee_{i \in \mathbb{N} \cup \{0\}} \mathcal{F}_\infty^{(i)} = \mathcal{F}_\infty$, as desired. $\square$

Corollary 14.5.8. *Let X be a finite dimensional Lévy process, with decomposition as in Theorem 13.5.9. Then any local martingale M in the filtration generated by X has a representation*

$$M = Z^\top \bullet W + H * \tilde{N}.$$

14.6 The Stratonovich Integral

We now define and state, without proof, some of the properties of the *Stratonovich* integral. This type of integral was first described by Stratonovich [170]. It is further investigated by Meyer [133], and is particularly suitable for geometric and filtering applications (see [48, 136]).

Definition 14.6.1. *Suppose H and X are two (real) continuous semimartingales. The* Stratonovich *integral H with respect to X is the process*

$$\int_{[0,t]} H_s \circ dX_s = \int_{[0,t]} H_s dX_s + \frac{1}{2} \langle H^c, X^c \rangle_t \,,$$

where the integral on the right is the Itô stochastic integral defined in Chapter 12, and H^c, X^c denote the continuous martingale parts of H and X, respectively.

The Itô stochastic integral can be viewed as the limit in probability of sums

$$\sum_i H_{t_i}(X_{t_{i+1}} - X_{t_i}),$$

where $0 = t_0 < t_1 < \cdots < t_n = t$ is a subdivision of $[0, t]$. In a similar way, the Stratonovich integral is the limit in probability of sums of the form

$$\sum_i H_{s_i}(X_{t_{i+1}} - X_{t_i}), \text{ where } \quad s_i = \frac{1}{2}(t_i + t_{i+1}).$$

Alternatively, the Stratonovich integral is the limit of Stieltjes integrals of the form

$$\int_{[0,t]} H_s^n dX_s^n ,$$

where H^n, X^n are the polygonal functions obtained by linear interpolation of H and X between the times $\{t_i^n\}$ of the n'th dyadic subdivision of $[0, t]$.

Lemma 14.6.2. *The Stratonovich integral satisfies the usual differentiation rule. That is, suppose X is a continuous semimartingale with canonical decomposition*

$$X = X_0 + M + A$$

(so $M = X^c$), and let f be a three times continuously differentiable function. Then $f'(X_t)$ is a continuous semimartingale and

$$\int_{[0,t]} f'(X_s) \circ dX_s = f(X_t) - f(X_0).$$

Proof. Write $H_t = f'(X_t)$ so that, by Itô's rule,

$$H_t = H_0 + \int_{[0,t]} f''(X_s)dM_s + \int_{[0,t]} f''(X_s)dA_s + \frac{1}{2}\int_{[0,t]} f'''(X_s)d\langle M\rangle_s .$$

Then the continuous martingale part of H is

$$H_t^c = \int_{[0,t]} f''(X_s)dM_s ,$$

and so

$$\langle H^c, X^c\rangle_t = \int_{[0,t]} f''(X_s)d\langle M\rangle_s .$$

By definition,

$$\int_{[0,t]} H_s \circ dX_s = \int_{[0,t]} f'(X_s)dX_s + \frac{1}{2}\int_{[0,t]} f''(X_s)d\langle M\rangle_s$$

and by Itô's rule this is equal to $f(X_t) - f(X_0)$. □

Remark 14.6.3. Suppose X is a real semimartingale and f a continuously differentiable function. Write Y for the process

$$Y_t = f(X_t),$$

and $\mathcal{Y}$ for the set of all processes which can be expressed in this form.

The Stratonovich integral can be defined for integrands $H \in \mathcal{Y}$ and general (discontinuous) semimartingale integrators X by

$$\int_{[0,t]} H_{s-} \circ dX_s = \int_{[0,t]} H_{u-} dX_u + \frac{1}{2}\left[H, X^c\right]_t.$$

If f is a twice continuously differentiable function it can then be shown, as above, combined with a mollification argument, that

$$\int_{[0,t]} f'(X_{s-}) \circ dX_s + \sum_{s \leq t} \big(f(X_s) - f(X_{s-}) - f'(X_{s-})\Delta X_s\big) = f(X_t) - f(X_0).$$

That is, the Stratonovich integral satisfies the same formula as when X is a process of finite variation. Further details can be found in Meyer [133, p.360]. Another natural definition is given by Marcus [129].

14.7 Exercises

Exercise 14.7.1. Let X be a continuous semimartingale. Using Itô's rule, describe the dynamics of $\cos(X)$, in particular when $X_t \approx n\pi/2$ for $n \in \mathbb{N}$.

Exercise 14.7.2. For $a > 0$, consider the 'sawtooth' function $\phi : \mathbb{R} \to [0, a]$ defined by the properties $\phi(x + 2a) = x$ for all x and $\phi(x) = |x|$ for $|x| \leq a$. Let X be a continuous semimartingale. Find the dynamics of $\phi(X)$, which is sometimes called the 'reflection' of X in $[0, a]$.

Exercise 14.7.3. Suppose X is a continuous martingale such that $P(|X_t| \in [a, b]$ for some $t) > 0$ for every $a < b$, and f is a twice differentiable function that is not convex. Show that $f(X)$ is not a submartingale.

Exercise 14.7.4. For f a C^2 or convex real valued function, X and Y semimartingales, show that XY and $f(X)$ are semimartingales. If $Y \neq 0$ show X/Y is a semimartingale.

Exercise 14.7.5. Let X and H be semimartingales and π a partition of $[0, \infty[$. Assuming H is locally bounded, show that the sequence

$$\sum_{t_i \in \pi} H_{t_i \wedge t}(X_{t_i \wedge t} - X_{t_{i-1} \wedge t})$$

converges ucp as $|\pi| \to 0$, and give its limit in terms of the Itô integral. (This limit is sometimes called the *backwards* Itô integral.)

Exercise 14.7.6. Let X be a Brownian motion and T a stopping time. Show that the process defined by

$$Y_t = \begin{cases} X_t & \text{for } t < T \\ 2X_T - X_t & \text{for } t \geq T \end{cases}$$

is also a Brownian motion. (This is called the 'Reflection principle' for Brownian motion.)

Exercise 14.7.7. Let X be a Brownian motion and T a stopping time. Show that the process defined by

$$Y_t = X_t - X_T$$

is a Brownian motion in the filtration $\{\tilde{\mathcal{F}}_t = \mathcal{F}_{T+t}\}$, that is, B is a strong Markov process.

Exercise 14.7.8. Let $\{h_n\}_{n\in\mathbb{N}}$ be the sequence of Hermite polynomials, that is, the solution to the recurrence relation

$$h_{n+1}(x) = xh_n(x) - \frac{d}{dx}h_n(x)$$

with $h_0 = 1$. Let X be a continuous local martingale and $Y = \sqrt{\langle X\rangle}$. Show that $Yh_n(X/Y)$ is a local martingale for all n. (Hint: First show that $\frac{d}{dx}h_n(x) = nh_{n-1}(x)$, and hence $nh_n - x\frac{d}{dx}h_n(x) + \frac{d^2}{dx^2}h_n(x) \equiv 0$.)

Exercise 14.7.9. For X a continuous semimartingale, by considering the integral of $H_t = I_{\{X_t=0\}}$, show that the approximation of the local time of X in Lemma 14.3.5 does not generally converge in the semimartingale topology.

Exercise 14.7.10. Show that, if X is a Brownian motion and L is its local time at zero, then LX is a local martingale, and describe its quadratic variation. Does this hold for the local time at any other point?

Exercise 14.7.11. The following example is based on Johnson and Helms [111].

Let $B = (B^1, B^2, B^3)$ be a three dimensional Brownian motion in a filtered probability space, with $B_0 = 0 \in \mathbb{R}^3$.

(i) Assuming $B_t \neq 0$ up to indistinguishability for $t > 0$, show that, $X_t = \|B_{t+1}\|$ has dynamics

$$dX_t = \frac{1}{X_t}dt + dW_t$$

for W some Brownian motion, in the filtration given by $\tilde{\mathcal{F}}_t = \mathcal{F}_{t+1}$. ($X$ is called the Bessel process of dimension 3, started at $\|B_1\|$.)

(ii) Show that $1/X$ is a local martingale and a supermartingale, and that $\lim_{t\to\infty} 1/X_t = 0$ a.s.

(iii) Using direct calculation from the density of B_t, show that for all $t > 0$, $E[\|B_t\|^{-2}] = 1/t$, and hence that $\{1/X_t\}_{t\geq 0}$ is uniformly integrable.

(iv) Hence show that $1/X$ is not a true martingale.

15

The Exponential Formula and Girsanov's Theorem

In this chapter, we consider a particularly important example of a stochastic differential equation, the 'stochastic exponential'. The solutions to this type of equation have many applications, as they form the basis of understanding how to use stochastic processes to change from one probability measure to another.

It will be convenient to write equations in a differential form (cf. Remark 12.3.16), that is, we shall write $dX_t = f(t)dY_t$ whenever $X_t = X_0 + \int_{]0,t]} f(s)dY_s$. Itô's rule can then be expressed (in the one-dimensional case) as

$$d(f(X_t)) = f'(X_{t-})dX_t + \frac{1}{2}f''(X_{t-})d\langle X^c\rangle_t + \Big(f(X_t) - f(X_{t-}) - f'(X_{t-})\Delta X_t\Big).$$

Note that we are not giving a precise meaning to the quantity 'dX', but merely using this as useful shorthand, in the sense of the Box calculus. Note also that the 'integral' form of this equation corresponds to taking a sum of the final term, that is, considering ΔX as generating a discrete measure on $[0,\infty]$. For w, v two processes of finite variation, we shall write $dw \geq dv$ if $\int_A dw \geq \int_A dv$ a.s. for any set $A \in \mathcal{B}([0,\infty])$.

15.1 Stochastic Exponentials

Definition 15.1.1. *For X a semimartingale, we define the stochastic exponential (also known as the Doléans-Dade exponential) to be the process $\mathcal{E}(X)$ defined by*

$$\mathcal{E}(X)_t = \exp\Big(X_t - \frac{1}{2}\langle X^c\rangle_t\Big)\prod_{s\leq t}(1+\Delta X_s)e^{-\Delta X_s}.$$

S.N. Cohen, R.J. Elliott, *Stochastic Calculus and Applications*,
Probability and Its Applications, DOI 10.1007/978-1-4939-2867-5_15

Lemma 15.1.2. *The stochastic exponential is well defined (as the infinite product converges absolutely almost surely). It satisfies the stochastic differential equation (SDE)*

$$d\mathcal{E}(X)_t = \mathcal{E}(X)_{t-}dX_t; \qquad \mathcal{E}(X)_0 = 1$$

or equivalently

$$\mathcal{E}(X)_t = 1 + \int_{]0,t]} \mathcal{E}(X)_{s-}dX_s.$$

If $\Delta X \geq -1$, then $\mathcal{E}(X)$ also satisfies the inequality

$$0 \leq \mathcal{E}(X)_t \leq \exp(X_t).$$

Proof. As X is a semimartingale, its quadratic variation is well defined. Hence, for every $n \in \mathbb{N}$, it must have almost surely finitely many jumps with absolute size greater than $1/n$ on any finite interval. Therefore

$$\exp(X_t) \prod_{\substack{\{s\leq t:\\ |\Delta X_s|>1/n\}}} \left|1 + \Delta X_s\right| e^{-\Delta X_s}$$

is well defined. As $0 < (1+x)e^{-x} < 1$ for all $x > -1$, we also know that the sequence

$$\exp(X_t) \prod_{\substack{\{s\leq t:\\ |\Delta X_s|>1/n\}}} (1 + \Delta X_s) e^{-\Delta X_s}$$

is of constant sign and is decreasing in absolute value. Therefore, it converges absolutely to a limit. Hence $\mathcal{E}(X)$ is well defined. That $\mathcal{E}(X)$ satisfies the stated SDE can be seen by applying Itô's rule. Finally, the stated inequality follows from the observation that $0 \leq (1+x)e^{-x} \leq 1$ for all $x \geq -1$, and nonnegativity of the exponential. □

Remark 15.1.3. Note that this definition works equally well when X is a complex-valued semimartingale.

Remark 15.1.4. For X a process of locally finite variation, this definition simplifies to

$$\mathcal{E}(X)_t = \exp(X_t) \prod_{s\leq t} (1 + \Delta X_s) e^{-\Delta X_s}.$$

Lemma 15.1.5. *For X a semimartingale with $\Delta X \neq -1$ and continuous local martingale part X^c, we have the identity $1/\mathcal{E}(X)_t = \mathcal{E}(-\bar{X})_t$, where*

$$\bar{X}_t = X_t - \langle X^c \rangle_t - \sum_{0\leq s\leq t} \frac{(\Delta X_s)^2}{1 + \Delta X_s},$$

and this sum is well defined.

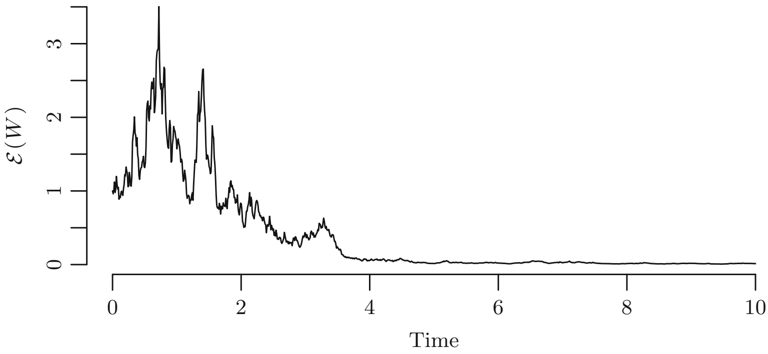

Fig. 15.1. A typical path of $\mathcal{E}(W)$, for W a Brownian motion. Note that $\mathcal{E}(W)_t \to 0$ as $t \to \infty$, as explained by Remark 15.1.10.

Proof. As X is a semimartingale, it has at most finitely many jumps of size $\Delta X_s < -1/2$ on $[0,t]$, so

$$\sum_{0\leq s\leq t} \frac{(\Delta X_s)^2}{1+\Delta X_s} \leq \sum_{\{\Delta X_s<-1/2\}} \frac{(\Delta X_s)^2}{1+\Delta X_s} + \sum_{\{\Delta X_s\geq -1/2\}} 2(\Delta X_s)^2 < \infty$$

and the sum is well defined.

We expand $\mathcal{E}(X)_t\mathcal{E}(-\bar{X})_t$ using the product rule, and notice that

$$(1-\Delta\bar{X}_s)^{-1} = \left(1-\Delta X_s + \frac{(\Delta X_s)^2}{1+\Delta X_s}\right)^{-1} = 1+\Delta X_s.$$

The result follows. (See also Corollary 15.1.9.) □

We can now state a useful variant of Grönwall's inequality.

Lemma 15.1.6 (Grönwall's Inequality). *Let X be a càdlàg process, Y an increasing real process and α a progressive almost-surely $Y(\omega)$-integrable process. Suppose*

$$X_t \leq \alpha_t + \int_{[0,t]} X_{s-}dY_s,$$

then, with $\bar{Y}$ as in Lemma 15.1.5,

$$X_t \leq \alpha_t + \mathcal{E}(Y)_t \int_{]0,t]} \mathcal{E}(-\bar{Y})_{s-}\alpha_s d\bar{Y}_s.$$

If $\alpha_t = \alpha$ is constant, this simplifies to

$$X_t \leq \alpha\mathcal{E}(Y)_t.$$

Proof. Note that $d\bar{Y}_t = \frac{dY_t}{1+\Delta Y_t}$ and that, considered as measures[1] on $[0,\infty]$, we have $\Delta\bar{Y}_t\Delta Y_t = \Delta Y_t d\bar{Y}_t$. Let

$$w_t := \mathcal{E}(-\bar{Y})_t \int_{[0,t]} X_{s-} dY_s.$$

From the product rule for Stieltjes integrals, as Y and $\bar{Y}$ are of finite variation and increasing,

$$\begin{aligned}
\frac{dw_t}{\mathcal{E}(-\bar{Y})_{t-}} &= -\Big(\int_{[0,t]} X_{s-} dY_s\Big) d\bar{Y}_t + X_{t-} dY - X_{t-}\Delta Y_t \Delta\bar{Y}_t \\
&= -\Big(\int_{[0,t]} X_{s-} dY_s\Big) d\bar{Y}_t + X_{t-} d\bar{Y}_t \\
&= \Big(X_t - \int_{[0,t]} X_{s-} dY_s\Big) d\bar{Y}_t \\
&\leq \alpha_t d\bar{Y}_t.
\end{aligned}$$

As $\mathcal{E}(\bar{Y})_{t-}$ is nonnegative, integration yields

$$\mathcal{E}(-\bar{Y})_t \int_{[0,t]} X_{s-} dY_s = w_t \leq \int_{]0,t]} \mathcal{E}(-\bar{Y})_{s-}\alpha_s d\bar{Y}_s,$$

hence

$$X_t \leq \alpha_t + \mathcal{E}(-\bar{Y}_t)^{-1} \int_{[0,t]} \mathcal{E}(-\bar{Y})_{t-}\alpha_s d\bar{Y}_s,$$

and the desired inequalities follow from the identity $\mathcal{E}(-\bar{Y})_t^{-1} = \mathcal{E}(Y)_t$. If $\alpha_t = \alpha$, then this simplifies to

$$\begin{aligned}
X_t &\leq \alpha\Big[1 + \mathcal{E}(Y)_t \int_{[0,t]} \mathcal{E}(-\bar{Y})_{t-} d\bar{Y}_s\Big] \\
&= \alpha\Big[1 + \mathcal{E}(Y)_t\Big(1 - \mathcal{E}(-\bar{Y})_t\Big)\Big] \\
&= \alpha\mathcal{E}(Y)_t.
\end{aligned}$$

□

As a corollary (taking $Y_t = \int_{[0,t]} \beta_s ds$), we obtain a classical version of Grönwall's inequality.

Corollary 15.1.7 (Classical Grönwall Inequality). *Suppose X is a càdlàg real process such that, for α and β measurable processes with $\beta \geq 0$ and both $\alpha\beta$ and β dt-integrable, we know*

$$X_t \leq \alpha_t + \int_{[0,t]} X_s \beta_s ds.$$

[1] That is, we have $\sum_{t\in[0,\infty]} I_A \Delta\bar{Y}_t \Delta Y_t = \int_{[0,\infty]} I_A \Delta Y_t d\bar{Y}_t$ for any $A \in \mathcal{B}$.

Then

$$X_t \le \alpha_t + \int_{[0,t]} \alpha_s \beta_s \exp\Big(\int_{[s,t]} \beta_r dr\Big) ds.$$

If β is constant and α is increasing or constant, then this reduces to

$$X_t \le \alpha_t \exp(\beta t).$$

Theorem 15.1.8. *The process $\mathcal{E}(X)$ is the unique (up to indistinguishability) solution to the SDE*

$$dZ_t = Z_{t-} dX_t; \qquad Z_0 = 1.$$

Proof. We have already seen that $Z = \mathcal{E}(X)$ solves the stated SDE. We need to verify that this solution is unique.

Suppose we have Z a solution to the SDE. Then, recalling that $\langle X^c \rangle$ is a continuous process, we can apply Itô's rule to obtain,

$$\begin{aligned}
&d \exp\Big(-X_t + \frac{1}{2}\langle X^c\rangle_t\Big) \\
&= e^{-X_t + \frac{1}{2}\langle X^c\rangle_t}\Big(-dX_t + \frac{1}{2} d\langle X^c\rangle_t\Big) + \frac{1}{2} e^{-X_t + \frac{1}{2}\langle X^c\rangle_t} d\langle X^c\rangle_t \\
&\qquad + \Big(e^{-X_t + \frac{1}{2}\langle X^c\rangle_t} - e^{-X_{t-} + \frac{1}{2}\langle X^c\rangle_{t-}} - e^{-X_{t-} + \frac{1}{2}\langle X^c\rangle_t} \Delta X_t\Big) \\
&= e^{-X_{t-} + \frac{1}{2}\langle X^c\rangle_{t-}}\Big(-dX_t + d\langle X^c\rangle_t + e^{-\Delta X_t} - 1 - \Delta X_t\Big).
\end{aligned}$$

Considering the process Y defined by

$$Y_t := \exp\Big(-X_t + \frac{1}{2}\langle X^c\rangle_t\Big) Z_t,$$

the integration by parts formula yields

$$\begin{aligned}
dY_t &= Y_{t-}\Big(-dX_t + d\langle X^c\rangle_t + e^{-\Delta X_t} - 1 - \Delta X_t\Big) + Y_{t-} dX_t \\
&\qquad + Y_{t-}\Big(-d\langle X^c\rangle_t + \Big(-\Delta X + e^{-\Delta X_t} - 1 - \Delta X_t\Big)\Delta X_t\Big) \\
&= Y_{t-}\Big((e^{-\Delta X_t} - 1 - \Delta X_t)(1 + \Delta X_t) - (\Delta X_t)^2\Big).
\end{aligned}$$

As $|(e^x - 1 - x)(1 + x) - x^2| < x^2$ for $|x| < 1$, and X is a semimartingale, we know that

$$J_t = \sum_{u \le t} \Big((e^{-\Delta X_u} - 1 - \Delta X_u)(1 + \Delta X_u) - (\Delta X_u)^2\Big)$$

is a process of finite variation, and Y is a solution to the SDE

$$dY_t = Y_{t-} dJ_t; \qquad Y_0 = 1.$$

It is clear that the uniqueness of Z is equivalent to the uniqueness of Y. Suppose we had two processes, Y and Y', both satisfying $dY_t = Y_{t-}dJ_t$ and $Y_0 = 1$. Then

$$|Y_t - Y'_t| = \Big| \int_{]0,t]} (Y_{s-} - Y'_{s-})dJ_s \Big| \le \int_{]0,t]} |Y_{s-} - Y'_{s-}||dJ_s|,$$

where $|dJ_s|$ denotes the total variation measure of J, which is generated by an increasing process. Grönwall's inequality (Lemma 15.1.6) implies that $|Y_t - Y'_t| \le 0$ a.s., that is, the solutions agree almost surely for each t. By right-continuity, the solution is unique up to indistinguishability. □

Corollary 15.1.9. *If X and Y are semimartingales, then*

$$\mathcal{E}(X)\mathcal{E}(Y) = \mathcal{E}(X + Y + [X,Y]\,).$$

Proof. This follows from the absolute convergence of the infinite product in Definition 15.1.1, because $(1+\Delta X_s)(1+\Delta Y_s) = (1+\Delta X_s+\Delta Y_s+\Delta X_s\Delta Y_s)$. Alternatively, writing $U = \mathcal{E}(X)$ and $V = \mathcal{E}(Y)$, we know

$$(UV)_t = 1 + \int_{]0,t]} U_{s-}dV_s + \int_{[0,t]} V_{s-}dU_s + [U,V]_t\,.$$

However,

$$U_t = 1 + \int_{]0,t]} U_{s-}dX_s \qquad \text{and} \qquad V_t = 1 + \int_{]0,t]} V_{s-}dY_s\,,$$

so

$$(UV)_t = 1 + \int_{]0,t]} (UV)_{s-}d(X_s + Y_s + [X,Y]_s).$$

That is,

$$UV = \mathcal{E}(X + Y + [X,Y]\,).$$

□

Remark 15.1.10. Consider $\mathcal{E}(W)$, the stochastic exponential of a Brownian motion (Fig. 15.1). From the law of the iterated logarithm (Theorem 5.5.14), $W_t - t/2 \to -\infty$ a.s. as $t \to \infty$, so $\mathcal{E}(W)_t \to 0$ a.s. as $t \to \infty$. On the other hand, it is easy to check that $E[\mathcal{E}(W)_t] = 1$ for all t, so $\mathcal{E}(W)$ converges almost surely, but not in L^1.

Furthermore, direct calculation shows that, for any $\epsilon > 0$,

$$P(\mathcal{E}(W)_t > \epsilon) = \Phi\Big(- \frac{\log(\epsilon)}{\sqrt{t}} - \frac{1}{2}\sqrt{t}\Big),$$

where Φ is the standard normal cumulative distribution function. Using the bound $\Phi(x) \le \frac{1}{\sqrt{2\pi}x}e^{-x^2/2}$, it follows that $P(W_t > \epsilon) \to 0$ exponentially fast as $t \to \infty$, for any $\epsilon > 0$.

15.2 Changes of Measure

Suppose $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$ is a filtered probability space, with the filtration satisfying the usual conditions (completeness and right-continuity). If Q is a second probability measure on $(\Omega, \mathcal{F}, P)$ which is equivalent to P (that is, $Q \ll P$ and $P \ll Q$), then the filtration also satisfies the usual conditions under Q. Furthermore, the evanescent sets are the same for Q and P.

Write

$$M_\infty = \frac{dQ}{dP},$$

and M for the càdlàg version of the uniformly integrable P-martingale $\{E[M_\infty|\mathcal{F}_t]\}_{t\geq 0}$. By the optional stopping theorem, for every stopping time T, the random variable M_T is a density for the restriction to $\mathcal{F}_T$ of Q with respect to P (Exercise 5.7.1). Note that, because Q and P are equivalent, M_∞ is a.s. strictly positive. Therefore, from Theorem 5.4.4, the stopping time

$$T = \inf\{t : M_t = 0, \text{ or } (t > 0 \text{ and } M_{t-} = 0)\}$$

must be ∞ a.s. Consequently, for almost every ω, the function $t \mapsto M_t(\omega)$ is bounded below on $[0, \infty]$ by a strictly positive number, and the process $\{M_{t-}\}_{t\geq 0}$ is locally bounded away from zero.

Lemma 15.2.1. *A process X is a (local) martingale under measure Q if and only if the process XM is a (local) martingale under measure P.*

Proof. Let $A \in \mathcal{F}_t$, and T be a stopping time. Then we have

$$E^Q[I_A X_t^T] = \int_A X_t^T dQ = \int_A X_t^T M_t dP = E^P[I_A X_t^T M_t]$$

If X^T is a Q-martingale, then, for any $s \geq t$,

$$E^Q[I_A X_t^T] = E^Q[I_A X_s^T],$$

and so $X^T M$ is a P-martingale. By optional stopping, so is $(XM)^T$. Similarly in the converse direction. Taking $T = \infty$ (in the martingale case) or a localizing sequence of stopping times $\{T_n\}_{n\in\mathbb{N}}$ (in the local martingale case) establishes the result. □

Corollary 15.2.2. *The process $\{M_t^{-1}\}_{t\geq 0}$ is a martingale under measure Q.*

Theorem 15.2.3. *A process $\{X_t\}_{t\geq 0}$ is a semimartingale under measure Q if and only if it is a semimartingale under measure P.*

Proof. By definition, a semimartingale is the sum of a local martingale and a process of finite variation. We need only prove the theorem in one direction and we can suppose $X_0 = 0$.

If X is a semimartingale under P, then, by the product rule, XM is also a semimartingale under P, which has a decomposition

$$XM = N + A, \qquad N \in \mathcal{M}_{0,\mathrm{loc}}, \quad A \in \mathcal{V}_0.$$

Therefore,

$$X = N/M + A/M.$$

By Lemma 15.2.1, N/M is a local martingale under Q, and A/M is the product of the finite variation process A and the Q-martingale M^{-1}, and hence is a semimartingale under Q. □

This in turn yields a fundamental property of the optional quadratic variation.

Lemma 15.2.4. *Let P and Q be equivalent measures on a complete filtered probability space and let X be a semimartingale. Then the optional variation $[X]$ under P is the same as the optional variation under Q. In particular $\langle X^c\rangle$ is the same under all measures equivalent to P.*

Proof. Let $[X]^P$ and $[X]^Q$ be the optional variations under the two measures, which exist, as X is a semimartingale under both measures. For any sequence of partitions π^n with $|\pi^n| \to 0$, we know from Corollary 14.1.2 that $Q(\pi^n, t) = \sum_i (X_{t_{i+1}\wedge t} - X_{t_i\wedge t})^2 \to [X]^P_t$ in P-probability. Choosing a subsequence, we have $Q(\pi^n, t) \to [X]^P_t$ almost surely (under both P and Q). However, by the same argument this subsequence must converge to $[X]^Q_t$ in Q-probability, and so we see that $[X]^P_t = [X]^Q_t$ a.s. As the two processes are both càdlàg, they are indistinguishable, and so the quadratic variations agree. As $[X]_t = \langle X^c\rangle_t + \sum_{0<s\leq t}(\Delta X)^2_s$ and this sum is defined pathwise, we see that $\langle X^c\rangle$ is independent of the choice of P. □

Remark 15.2.5. From this result, we see that if X is a semimartingale with characteristics (B, C, μ^X_p), as in Section 13.4, then the 'volatility' term C (which is based on the predictable variation of the continuous martingale part of X) does not depend on the choice of measure, while B and μ^X_p may change.

Suppose X is a local martingale under measure P. Then the above result states that X is certainly a semimartingale under measure Q, and the quadratic variation can be defined independently of the measure. The following result, which is a generalization due to Yoeurp [187] of the work of Girsanov [89], gives an explicit decomposition of X as a Q-semimartingale.

Theorem 15.2.6 (Girsanov's Theorem). *Suppose X is a P-local martingale which is zero at $t = 0$, Q is a probability measure equivalent to P, and M is the càdlàg version of the martingale $\{E[dQ/dP|\mathcal{F}_t]\}_{t\geq 0}$.*

(i) X is a special semimartingale under Q if the predictable covariation $\langle X, M\rangle$ exists (under P), and then the canonical decomposition of X under Q is

$$X_t = \Big(X_t - \int_{[0,t]} M_{s-}^{-1} d\langle X, M\rangle_s\Big) + \int_{[0,t]} M_{s-}^{-1} d\langle X, M\rangle_s.$$

Here the first term is a local martingale under Q, and the second is a predictable process of finite variation.

(ii) In general, the process

$$\Big\{X_t - \int_{[0,t]} M_s^{-1} d[X, M]_s\Big\}_{t\geq 0}$$

is a local martingale under Q.

Note that in statement (i) of the theorem, the predictable quadratic covariation $\langle X, M\rangle$ is defined under P, and we make no statement regarding the existence of the predictable quadratic covariation of X and M under Q (and this is not guaranteed when X and M can have unbounded jumps).

Proof. (i) Suppose $\langle X, M\rangle$ exists under P. Then

$$(XM) - \langle X, M\rangle \tag{15.1}$$

is a P-local martingale. For any predictable process of finite variation A with $A_0 = 0$, by integration by parts (Corollary 14.1.1),

$$(AM)_t - \int_{[0,t]} M_{s-} dA_s = \int_{[0,t]} A_{s-} dM_s + [M, A]_t\,. \tag{15.2}$$

From Theorem 11.6.8, we have $[M, A] \in \mathcal{M}_{0,\mathrm{loc}}(P)$, and $\{A_{s-}\}_{s\geq 0}$ is locally bounded, so the right-hand side is a P-local martingale.

As $\{M_{s-}\}_{s\geq 0}$ is locally bounded away from zero, we can define

$$A_t = \int_{[0,t]} M_{s-}^{-1} d\langle X, M\rangle_s,$$

so A is a càdlàg predictable process in $\mathcal{A}_{\mathrm{loc}}(Q)$, by Lemma 8.3.5, and $\int_{[0,t]} M_{s-} dA_s = \langle X, M\rangle_t$. Combining (15.1) and (15.2), we see that $M(X-A)$ is a P-local martingale. By Lemma 15.2.1 this is equivalent to stating that $X - A$ is a Q-local martingale, and hence that

$$X = (X - A) + A$$

is a special semimartingale under Q.

Conversely, if X is a special semimartingale under Q, then there is a predictable process of finite variation A such that $X - A$ is a local martingale under Q. Then $M(X - A)$ is a local martingale under P and, from (15.2),

$$\Big\{M_t X_t - \int_{[0,t]} M_{s-} dA_s\Big\}_{t\geq 0}$$

is a local martingale under P. However, $\{\int_{[0,t]} M_{s-} dA_s\}_{t\geq 0}$ is a predictable process of finite variation, so $\langle M, X\rangle_t = \int_{[0,t]} M_{s-} dA_s$.

(ii) Because M is strictly positive, we can define a process in $\mathcal{V}_0$ by

$$B_t := \int_{[0,t]} M_s^{-1} d[X, M]_s\,.$$

Because B is of finite variation, we have

$$M_t B_t = \int_{[0,t]} M_s dB_s + \int_{[0,t]} B_{s-} dM_s\,,$$

so

$$\int_{[0,t]} B_{s-} dM_s = M_t B_t - \int_{[0,t]} M_s dB_s = M_t B_t - [X, M]_t.$$

As $\{B_{s-}\}_{s\geq 0}$ is predictable and locally bounded, we see $MB - [X, M]$ is a local martingale under P. From the definition of $[X, M]$ we know $XM - [X, M]$ is a local martingale under P. Therefore, $M(X - B)$ is a local martingale under P, and so $X - B$ is a local martingale under Q. □

Theorem 15.2.7. *Let P and Q be equivalent probability measures. Suppose X is a semimartingale (under both P and Q, by the above result) and H is an X-integrable process under P. Then the following hold:*

(i) H is X-integrable under Q,
(ii) The stochastic integrals $H\bullet X$ defined under measure P and under measure Q are indistinguishable.

Proof. Clearly integrals with respect to processes of finite variation are the same under either measure, so by linearity it is sufficient to consider the case when X is a local martingale under P. Write $Y = H\bullet_P X$ for the integral under P. By Theorem 15.2.6(ii),

$$\widetilde{Y}_t := Y_t - \int_{[0,t]} M_t^{-1} d[Y, M]_s$$

defines a local martingale under measure Q. As $[X, M]$ is of finite variation, by Lemma 12.3.7,

$$[Y, M] = H\bullet_P [X, M] = H\bullet_Q [X, M]$$

and, for any P-semimartingale U,

$$
\begin{aligned}
[\widetilde{Y}, U]_t &= \Big[\Big(Y - \int_{[0,\cdot]} H_s M_s^{-1} d[X,M]_s\Big), \quad U\Big]_t \\
&= [Y,U]_t - \sum_{0\le s\le t} \frac{H_s}{M_s} \Delta X_s \Delta M_s \Delta U_s \\
&= \Big(H \bullet \Big[\Big(X - \int_{[0,\cdot]} M_s^{-1} d[X,M]_s\Big), U\Big]\Big)_t = \big(H \bullet [\widetilde{X}, U]\big)_t ,
\end{aligned}
$$

where $\widetilde{X}$ is the Q local martingale defined by

$$\widetilde{X}_t := X_t - \int_{[0,t]} M_s^{-1} d[X,M]_s.$$

If U is a Q-local martingale this identity is the relation which characterizes the stochastic integral $H \bullet_Q \widetilde{X}$ (see Corollary 12.3.6).

Therefore, if H is X-integrable under Q, we know $\widetilde{Y} = H \bullet_Q \widetilde{X}$. However, this implies

$$
\begin{aligned}
(H \bullet_Q X)_t &= (H \bullet_Q \widetilde{X})_t + \int_{]0,t]} H_s M_s^{-1} d[X,M]_s \\
&= \widetilde{Y}_t + \int_{[0,t]} H_s M_s^{-1} d[X,M]_s \\
&= (H \bullet_P X)_t \,.
\end{aligned}
$$

We now note that this could be used to *define* the stochastic integral $H \bullet_Q X$ by reference to $H \bullet_P X$. As this would define a bilinear map satisfying the requirements of Theorem 12.3.22, it follows that H must be X-integrable independently of the choice of measure. □

We can also see how the predictable representation property changes under a change of measure. Note that the following result does not assume that either Y or $\tilde{Y}$ generates the filtration.

Theorem 15.2.8. *Let P and Q be equivalent probability measures on a filtered probability space, and M be the càdlàg martingale defined by $M_t = E[dQ/dP|\mathcal{F}_t]$.*

Suppose Y is a P-local martingale with the predictable representation property, that is, any P-local martingale N can be written $N = H \bullet Y$ for some Y-integrable H. Suppose also that $\langle Y, M\rangle$ and $\langle Y \rangle$ exist under P. Then

$$\tilde{Y}_t = Y_t - \int_{[0,t]} M_{s-}^{-1} d\langle Y, M\rangle_s$$

defines a Q-local martingale $\tilde{Y}$ with the predictable representation property under Q.

Proof. Let X be a Q-local martingale. Suppose X is locally bounded, and so X is a P-special semimartingale. We write $X = A + N$, for N a P-local martingale and A a predictable finite variation process. As Y has the predictable representation property under P, there exists a predictable process H such that $X = A + H \bullet Y$. Under Q, the canonical decomposition of Y has martingale part $\tilde{Y}$ and $X - A$ is a special semimartingale, so by Theorem 12.3.18 we can replace Y with $\tilde{Y}$,

$$X_t = A_t + \int_{[0,t]} H_s M_{s-}^{-1} d\langle Y, M\rangle_s + (H \bullet \tilde{Y})_t$$

and $H \bullet \tilde{Y}$ is a Q-local martingale. Therefore, the predictable finite variation terms on the right must cancel (by Theorem 8.2.14), and it follows that $X = H \bullet \tilde{Y}$.

If X is a general Q-local martingale, then by localization we can assume that X is in $\mathcal{H}^1(Q)$. Theorem 10.1.7 implies that X can be approximated in $\mathcal{H}^1(Q)$ by a sequence $\{X^n\}_{n\in\mathbb{N}}$ of bounded Q-martingales, each of which has a representation $X^n = H^n \bullet \tilde{Y}$. Therefore, for $m \geq n$, using the BDG inequality, as $n \to \infty$ we have

$$\|X^n - X^m\|_{\mathcal{H}^1(Q)} \leq C E^Q\Big[\Big(\int_{[0,\infty[} (H^n - H^m)^2 d\langle \tilde{Y}\rangle\Big)^{1/2}\Big] \to 0.$$

It follows that, at least for a subsequence, H^n converges pointwise $d\langle\tilde{Y}\rangle \times dP$-a.e. to a process H, and that $X^n = H^n \bullet \tilde{Y} \to H \bullet \tilde{Y}$ in $\mathcal{H}^1(Q)$. We then observe $X = H \bullet \tilde{Y}$, as desired. □

Remark 15.2.9. From the above proof, we also note that the corresponding process H in the representation of the martingale part of X does not depend on the choice of measure.

Remark 15.2.10. If $\{Y^n\}_{n\in\mathbb{N}}$ is a sequence of martingales with the predictable representation property (jointly), that is, for any local martingale N there exists $\{H^n\}_{n\in\mathbb{N}}$ such that $N = \sum_{n\in\mathbb{N}} H^n \bullet Y^n$, the sum converging in $\mathcal{H}^1_{\mathrm{loc}}(P)$, then the same argument holds with only notational changes. By approximation, a similar extension can be made to include random measures with the predictable representation property.

15.3 Stochastic Exponentials as Measure Changes

We now discuss the relationship between the stochastic exponential, the change of probability measure and Girsanov's theorem.

Remark 15.3.1. From Theorem 15.1.8, if X is a semimartingale, then the unique solution of the equation

$$Z_t = 1 + \int_{[0,t]} Z_{s-} dX_s$$

is

$$Z_t = \exp\left(X_t - \frac{1}{2}\langle X^c\rangle_t\right) \prod_{0\leq s\leq t} (1+\Delta X_s)e^{-\Delta X_s} = \mathcal{E}(X)_t.$$

If X is a local martingale, then (as $t \mapsto Z_{t-}$ is left continuous and hence locally bounded) $Z = \mathcal{E}(X)$ is a local martingale. Furthermore, $\mathcal{E}(X)$ is strictly positive if and only if $\Delta X > -1$ up to indistinguishability (and $1/\mathcal{E}(X)$ is as in Lemma 15.1.5).

Lemma 15.3.2. *If X is a local martingale with $\Delta X \geq -1$ then $\mathcal{E}(X)$ is a nonnegative local martingale. Furthermore $\mathcal{E}(X)$ is a martingale if and only if $E[\mathcal{E}(X)_t] = 1$ for each $t \geq 0$.*

Proof. Because

$$\mathcal{E}(X)_t = Z_t = 1 + \int_{[0,t]} Z_{s-} dX_s$$

we know Z is a local martingale. As $Z \geq 0$ (by Lemma 15.1.2) by Exercise 9.4.5 we see that Z is a supermartingale. By the supermartingale property, $E[Z_t|\mathcal{F}_s] \leq Z_s$ a.s. for all $t > s$, so we have $E[Z_t] = E[Z_s]$ for all $t > s$ if and only if $E[Z_t|\mathcal{F}_s] = Z_s$ a.s. for all $t > s$. As $E[Z_0] = 1$ the second statement follows. □

Remark 15.3.3. By Theorem 5.2.1 and its corollaries, if $Z = \mathcal{E}(X)$ is a uniformly integrable positive martingale, then $Z_\infty = \lim_{t\to\infty} Z_t$ exists in L^1, and

$$E[Z_\infty|\mathcal{F}_t] = Z_t \qquad \text{a.s.}$$

Consequently,

$$E[Z_\infty] = E[Z_0] = 1.$$

Furthermore, in this situation a new probability measure Q can be defined on $(\Omega, \mathcal{F})$ by

$$\frac{dQ}{dP} = Z_\infty.$$

We see that Q is equivalent to P if and only if $Z_\infty > 0$ a.s.

By exploiting Lévy's characterization of Brownian motion, we can obtain the following useful result, which shows how a measure change can 'introduce a drift' to a Brownian motion. Note, however, that the assumption of uniform integrability is quite strong and often restricts us to working only over finite horizons (cf. Exercise 15.6.1).

Corollary 15.3.4. *Suppose $\{B_t\}_{t\geq 0}$ is a Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$. Let $f : \Omega \times [0, \infty[\to \mathbb{R}$ be a predictable process such that $\int_{[0,\infty[} f_t^2 dt < \infty$. Write*

$$\mathcal{E}(f \bullet B)_t = \exp\Big(\int_{[0,t]} f_s dB_s - \frac{1}{2}\int_{[0,t]} f_s^2 ds\Big)$$

and suppose $\mathcal{E}(f \bullet B)$ is a uniformly integrable martingale. If Q is the probability measure on $(\Omega, \mathcal{F})$ defined by $dQ/dP = \mathcal{E}(f \bullet B)_\infty$, then $\widetilde{B}_t$ is a Brownian motion on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, Q)$, where

$$\widetilde{B}_t := B_t - \int_{[0,t]} f_s ds.$$

Proof. Apply Theorem 15.2.6 with $M = \mathcal{E}(f \bullet B)$ and $X = B$. Then

$$\langle X, M\rangle_t = \int_{[0,t]} M_{s-} f_s ds$$

and by Girsanov's theorem (Theorem 15.2.6) we see that

$$X_t - \int_{[0,t]} M_{s-}^{-1} d\langle X, M\rangle_s = B_t - \int_{[0,t]} f_s ds = \widetilde{B}_t$$

defines a local martingale under Q. By Lemma 15.2.4, $\langle B\rangle_t = t$ under Q as well as P, and as B and $\widetilde{B}$ differ by a continuous finite variation process, we have $\langle \widetilde{B}\rangle_t = t$ also. Theorem 14.4.1 then implies that $\widetilde{B}$ is a Brownian motion. □

Remark 15.3.5. Suppose $\int_{[0,T]} f_t^2 dt < \infty$ for all deterministic $T > 0$. Then applying the above result to functions of the form $I_{[0,T]} f$, we can define a sequence of measures $Q^T \sim P$ such that $\tilde{B}_t^T := B_t^T - \int_{[0,t\wedge T]} f_s ds$ defines a stopped Q^T Brownian motion for all $T < \infty$.

If our space $(\Omega, \mathcal{F})$ is $((\mathbb{R}^n)^{[0,\infty]}, \mathcal{B}((\mathbb{R}^n)^{[0,\infty]}))$, and we take the filtration generated by the canonical process, then we can use Kolmogorov's extension theorem[2] (Theorem A.2.6) to find a measure Q such that $Q = Q^S$ on $\mathcal{F}_S$ for all $S < \infty$. However, it is not generally true that $Q \sim P$, which causes technical difficulties if we have assumed our filtration to be P-complete (that is, if we assume that $\mathcal{F}_0$ contains all P-null sets of $\mathcal{F}$, which will then include all subsets of some Q-non-null sets). A careful treatment of these issues can be found in the first chapter of Jacod and Shiryaev [110].

[2] In more general settings, but not all, Parthasarathy's extension theorem can be used in the place of Kolmogorov's extension theorem in constructing this measure. See [148].

In a similar way, we can modify the rate of a Poisson process through a change of measure.

Corollary 15.3.6. *Suppose N is a Poisson process with parameter λ on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$. Write $\tilde{N}_t = N_t - \lambda t$ for the martingale associated with N. Let $\mu > 0$ be a constant and, for any deterministic $T > 0$, write*

$$\frac{dQ^T}{dP} = \mathcal{E}((\mu/\lambda - 1)\bullet\tilde{N})_T.$$

Then Q^T is a probability measure equivalent to P and N^T is a Q^T-Poisson process, stopped at T, with parameter μ.

Proof. As N is a Poisson process and $\tilde{N}_t = N_t - \lambda t$, we know that $\langle \tilde{N}^c\rangle = 0$, so

$$\begin{aligned} Z_t &:= \mathcal{E}((\mu/\lambda - 1)\bullet\tilde{N})_t \\ &= \exp\big((\mu/\lambda - 1)(N_t - t)\big)\prod_{0\leq s\leq t}(1 + (\mu/\lambda - 1)\Delta N_s)e^{(\mu/\lambda-1)\Delta N_s} \\ &= \exp\big(-t(\mu/\lambda - 1)\big)(\mu/\lambda)^{N_t} \end{aligned}$$

which is uniformly integrable, as N_t has a Poisson distribution. Clearly $Z_t > 0$, so Q^T is a probability measure equivalent to P.

As $\langle\tilde{N}\rangle_t = \lambda t$ under P, we know that, for $t < T$,

$$\tilde{N}_t - \int_{[0,t]} Z_{t-}^{-1} d\langle\tilde{N}, Z\rangle_t = N_t - \lambda t - \int_{[0,t^T]}(\mu/\lambda - 1)\lambda dt = N_t - \mu t.$$

By Girsanov's theorem (Theorem 15.2.6), writing $\hat{N}_t := N_t - \mu t$ we see that $\hat{N}^T$ is a local martingale under Q^T.

Finally, as $[\hat{N}]_t = \sum_{0\leq u\leq t}(\Delta\hat{N}_u)^2 = N_t$, we see that

$$\hat{N}_t^2 - \mu t = \big(\hat{N}_t^2 - [\hat{N}]_t\big) + \big(N_t - \mu t\big)$$

and so $\{\hat{N}_t^2 - \mu t\}_{t\geq 0}$ is a Q^T-local martingale. By Theorem 14.4.5, it follows that N_t^T is a stopped Poisson process with parameter μ under Q^T. □

We can also obtain a result for random measures.

Corollary 15.3.7. *Suppose $\mu \in \tilde{\mathcal{A}}^1_\sigma$ (that is, μ is a predictably locally integrable integer valued random measure) on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$. Let μ_p be its compensator under P, and write $\tilde{\mu} = \mu - \mu_p$. Assume that $\mu_p(\{t\}\times\mathcal{Z}) = 0$ up to indistinguishability (so μ has no predictable jumps).*

Let α_p be a locally integrable nonnegative predictable random measure absolutely continuous with respect to μ_p for P-almost all ω. Let A be a predictable version of their Radon–Nikodym density $d\alpha_p/d\mu_p$.

Suppose $(A-1)$ *is stochastically* $\tilde{\mu}$*-integrable and* $\mathcal{E}((A-1)*\tilde{\mu})$ *is uniformly integrable. Then, if we define the measure* Q *by*

$$\frac{dQ}{dP} = \mathcal{E}((A-1) * \tilde{\mu})_\infty,$$

it follows that α_p *is the compensator of* μ *under* Q*.*

Proof. Write $(\mathcal{Z}, \mathfrak{Z})$ for the Blackwell space associated with μ. As $\mu \in \tilde{\mathcal{A}}^1_\sigma$, by localization (in time and in $\mathcal{Z}$) we can assume $\mu \in \tilde{\mathcal{A}}^1$. For an arbitrary set $B \in \mathfrak{Z}$, write $X_t = \tilde{\mu}([0,t] \times B)$. As $\tilde{\mu}$ is a martingale random measure, X is a local martingale with predictable quadratic covariations (Corollary 13.3.17)

$$\langle X, A * \tilde{\mu} \rangle = (I_B A) * \mu_p.$$

Theorem 15.2.6 states that

$$\begin{aligned} X_t - \langle X, A * \tilde{\mu} \rangle_t &= \tilde{\mu}([0,t] \times B) - \left(\left(I_B\left(\frac{d\alpha_p}{d\mu_p} - 1\right)\right) * \mu_p\right)_t \\ &= \mu([0,t] \times B) - \alpha_p([0,t] \times B) \end{aligned}$$

defines a Q-local martingale. As the compensator of μ under Q is unique, the result is proven. □

Remark 15.3.8. This gives a straightforward technique for constructing random measures with various stochastic compensators. One simply defines a Poisson random measure as in Example 13.3.8, then uses a stochastic exponential to change the measure to one where the compensator is a desired path-dependent function (for example, the rate at which jumps occur can depend on the time since the last jump). The only difficulty is verifying that the stochastic exponential is a uniformly integrable martingale.

Remark 15.3.9. In the setting where μ_p may have predictable jumps, higher order terms appear. See Section 20.1 for a special case where this occurs.

Combining these results gives a general view of how a change of measure affects the characteristics of a semimartingale.

Theorem 15.3.10. *Let* X *be a scalar semimartingale with characteristics* $(B, \langle X^c \rangle, \mu_p^X)$*, where* μ^X *is the random measure associated with the jumps of* X*, and has compensator* μ_p^X*. Let* μ_p^X*,* α_p *and* A *be as in Corollary 15.3.7,* $\tilde{\mu}^X = \mu^X - \mu_p^X$ *and* f *a predictable process. Suppose the stochastic exponential*

$$Z = \mathcal{E}\Big(f \bullet X^c + (A-1) * \tilde{\mu}^X\Big)$$

is a uniformly integrable martingale. Then under the measure Q *defined by* $dQ/dP = Z_\infty$*,* X *is a semimartingale with characteristics* $(\hat{B}, \langle X^c \rangle, \alpha_p)$*, where*

$$\hat{B}_t = B_t + (f \bullet \langle X^c \rangle)_t + \int_{\{|\zeta|<1\}} \zeta \left(\mu_p^X - \alpha_p\right)([0,t] \times d\zeta).$$

Proof. As $f\bullet X^c$ and $(A-1)*\tilde{\mu}^X$ are orthogonal local martingales, it is easy to see, as in the previous corollaries, that $\hat{X}^c := X^c - f\bullet\langle X^c\rangle$ is a continuous Q-martingale (with quadratic variation $\langle\hat{X}^c\rangle = \langle X^c\rangle$) and the compensator of μ^X under Q is α_p. Using Theorem 13.4.10, we write

$$\begin{aligned}
X_t &= X_0 + B_t + X_t^c + \int_{\{|\zeta|<1\}} \zeta\, \tilde{\mu}^X(dt, d\zeta) + \int_{\{|\zeta|\geq 1\}} \zeta\, \mu^X(dt, d\zeta)\\
&= X_0 + \hat{B}_t + \hat{X}_t^c + \int_{\{|\zeta|<1\}} \zeta\, \big(\mu^X - \alpha_p\big)(dt, d\zeta) + \int_{\{|\zeta|\geq 1\}} \zeta\, \mu^X(dt, d\zeta).
\end{aligned}$$

The uniqueness of the canonical decomposition of $X - \int_{\{|\zeta|\geq 1\}} \zeta\, \mu^X(dt, d\zeta)$ shows that $\hat{B}$ is the first term of the Q-characteristics of X. □

Remark 15.3.11. The case where X is a vector semimartingale is almost identical. The only distinction is that $\{|\zeta|\geq 1\}$ becomes $\{\|\zeta\|\geq 1\}$, $f\bullet X^c$ becomes $f^\top\bullet X^c$ and $(f\bullet\langle X^c\rangle)_t$ becomes

$$\int_{[0,t]} (d\langle X^c\rangle f) = (f^\top\bullet\langle X^c\rangle)_t^\top$$

where the integral of a vector with respect to the matrix valued finite variation process $\langle X^c\rangle$ is defined component by component, using the stochastic vector integral.

15.4 The Novikov and Kazamaki Criteria

From the examples above, it is clearly of importance to know when the stochastic exponential $\mathcal{E}(X)$ is a uniformly integrable martingale, rather than simply a local martingale. We here give two basic criteria which give sufficient conditions under which this is the case, in the setting where X is a continuous local martingale. Results when X is not continuous follow.

Theorem 15.4.1 (Kazamaki's Criterion). *Let $\{X_t\}_{t\in[0,\infty[}$ be a continuous local martingale. Suppose*

$$\sup_T E[\exp(X_T/2)] < \infty,$$

where the supremum is taken over bounded stopping times T. Then $\mathcal{E}(X)$ is a uniformly integrable martingale and $\mathcal{E}(X)_\infty := \lim_t \mathcal{E}(X)_t > 0$ a.s.

Proof. Write $\mathcal{T}_b$ for the family of bounded stopping times. Let $\lambda\in\,]0,1[$. We can write

$$\begin{aligned}
\mathcal{E}(\lambda X) &= \exp\Big(\lambda X - \frac{\lambda^2}{2}\langle X\rangle\Big)\\
&= \exp\Big(\lambda^2 X - \frac{\lambda^2}{2}\langle X\rangle\Big)\exp\big((\lambda-\lambda^2)X\big),\\
&= \big(\mathcal{E}(X)\big)^{\lambda^2}\exp\big(\lambda(1-\lambda)X\big).
\end{aligned}$$

Write $Z^{(\lambda)} = \exp\left(\frac{\lambda}{1+\lambda}X\right)$, so $(Z^{(\lambda)})^{1-\lambda^2} = \exp(\lambda(1-\lambda)X)$. We note that $(Z_T^{(\lambda)})^{(1+\lambda)/2\lambda} \le 1 + \exp(X_T/2)$, so $\sup_{T\in\mathcal{T}_b} E[(Z_T^{(\lambda)})^{(1+\lambda)/2\lambda}] < \infty$. As $(1+\lambda)/2\lambda > 1$, we see that $\{Z_T^{(\lambda)}\}_{T\in\mathcal{T}_b}$ is uniformly integrable. By Jensen's inequality and convexity of $x \mapsto \exp(\frac{\lambda}{1+\lambda}x)$, we see that $Z^{(\lambda)}$ is a submartingale. Therefore, $Z^{(\lambda)}$ is a uniformly integrable submartingale, and $Z_\infty^{(\lambda)}$ is well defined as an a.s. or L^1 limit (by Theorem 5.2.1 applied to $-Z^{(\lambda)}$). Rearrangement yields the limiting random variable $X_\infty = \frac{1+\lambda}{\lambda}\log(Z_\infty)$ and, by Theorem 11.4.7, as X is a local martingale, this implies that $\langle X\rangle_\infty < \infty$, and so $\mathcal{E}(X)_\infty > 0$.

From Fatou's lemma, we see that

$$E[\exp(X_\infty/2)] = E[\liminf_t \exp(X_t/2)] \le \liminf_t E[\exp(X_t/2)] < \infty.$$

Applying Hölder's inequality, for any set $A \in \mathcal{F}$ and any bounded stopping time T, as we know $E[\mathcal{E}(X)_T] \le 1$, we have

$$E[I_A\mathcal{E}(\lambda X)_T] \le E[\mathcal{E}(X)_T]^{\lambda^2} E[I_A Z_T^{(\lambda)}]^{1-\lambda^2} \le E[I_A Z_T^{(\lambda)}]^{1-\lambda^2}.$$

As $\{Z_T^{(\lambda)}\}_{T\in\mathcal{T}_b}$ is uniformly integrable, it follows that $\{\mathcal{E}(\lambda X)_T\}_{T\in\mathcal{T}_b}$ is uniformly integrable. Therefore, $\mathcal{E}(\lambda X)$ is of class (D) and hence a uniformly integrable martingale.

Using Hölder's inequality again,

$$\begin{aligned} 1 = E[\mathcal{E}(\lambda X)_\infty] &\le E[\mathcal{E}(X)_\infty]^{\lambda^2} E[Z_\infty^{(\lambda)}]^{1-\lambda^2} \\ &= E[\mathcal{E}(X)_\infty]^{\lambda^2} E\Big[\exp\Big(\frac{\lambda}{1+\lambda}X_\infty\Big)\Big]^{1-\lambda^2} \\ &\le E[\mathcal{E}(X)_\infty]^{\lambda^2} E[\exp(X_\infty/2)]^{2\lambda(1-\lambda)}. \end{aligned}$$

Taking $\lambda \to 1$, we see that $E[\mathcal{E}(X)_\infty] \ge 1$. As $\mathcal{E}(X)$ is a supermartingale we know that $E[\mathcal{E}(X)_\infty] \le 1$, so we conclude that $E[\mathcal{E}(X)_\infty] = 1$ and $\mathcal{E}(X)$ is a uniformly integrable martingale. □

Theorem 15.4.2 (Novikov's Criterion). *Let X be a continuous local martingale. Suppose that*

$$E[\exp(\langle X\rangle_\infty/2)] < \infty.$$

Then $\mathcal{E}(X)$ is a uniformly integrable martingale and $\mathcal{E}(X)_\infty > 0$.

Proof. For T any bounded stopping time,

$$\big(\mathcal{E}(X)_T\big)^{1/2} = \big(\exp(X_T - \langle X\rangle_T/2)\big)^{1/2} = \exp(X_T/2)\big(\exp(-\langle X\rangle_T/2)\big)^{1/2}$$

and, therefore,

$$\exp(X_T/2) = \big(\mathcal{E}(X)_t\big)^{1/2}\big(\exp(\langle X\rangle_T/2)\big)^{1/2}.$$

Applying the Cauchy–Schwarz inequality and recalling that $E[\mathcal{E}(X)_T] \leq 1$,

$$E[\exp(X_T/2)] \leq E[\mathcal{E}(X)_T]^{1/2} E[\exp(\langle X \rangle_T/2)]^{1/2} \leq E[\exp(\langle X \rangle_\infty/2)]^{1/2},$$

so we see that Kazamaki's criterion holds, and the result follows from Theorem 15.4.1. □

In the presence of jumps, these conditions become more involved, as one can consider providing conditions based either on the predictable compensator of the jumps, or on the jumps themselves. The following general results, due to Lépingle and Mémin [125], are proven in Appendix A.7.

Theorem 15.4.3. *Suppose that X is a local martingale with $\Delta X \geq -1$, and let $T = \inf\{t : \Delta X_t = -1\} = \inf\{t : \mathcal{E}(X)_t = 0\}$. If the increasing process*

$$\frac{1}{2}\langle X^c \rangle_{t\wedge T} + \sum_{s \leq t\wedge T} \big((1+\Delta X_s)\log(1+\Delta X_s) - \Delta X_s\big)$$

has a predictable compensator B and $E[\exp(B_\infty)] < \infty$, then $\mathcal{E}(X)$ is a uniformly integrable martingale and $\{\mathcal{E}(X)_\infty > 0\} = \{T = \infty\}$ almost surely.

The following simple case is often easier to verify.

Corollary 15.4.4. *For the above conditions to hold, it is sufficient that X be a locally square integrable local martingale, $\Delta X \geq -1$ and that*

$$E\Big[\exp\Big(\frac{1}{2}\langle X^c \rangle_T + \langle X^d \rangle_T\Big)\Big] < \infty.$$

Proof. Write $B = B^c + B^d$, where $B^c = \frac{1}{2}\langle X^c \rangle_{t\wedge T}$ and B^d is the compensator of

$$\sum_{s \leq t\wedge T} \big((1+\Delta X_s)\log(1+\Delta X_s) - \Delta X_s\big) \leq \sum_{s \leq t\wedge T} (\Delta X_s)^2.$$

We then see that $B^d_t \leq \langle X^d \rangle_{t\wedge T}$, and the result follows by monotonicity. □

One can also give a condition involving the jumps directly.

Theorem 15.4.5. *Suppose that X is a local martingale with $\Delta X > -1$ and*

$$E\Big[\exp\Big(\frac{1}{2}\langle X^c \rangle_\infty\Big) \prod_{t<\infty} (1+\Delta X_t)\exp\Big(-\frac{\Delta X_t}{1+\Delta X_t}\Big)\Big] < \infty.$$

Then $\mathcal{E}(X)$ is a uniformly integrable martingale and $\mathcal{E}(X)_\infty > 0$ a.s.

Given the stronger assumption in the following theorem, Lépingle and Mémin [124] also show that $\mathcal{E}(X) \in \mathcal{H}^p$ for some $p > 1$. Note that, in case (i) of this theorem, we do not need $\Delta X > -1$ (and so $\mathcal{E}(X)$ may be negative).

Theorem 15.4.6. *(i) Suppose that M is any local martingale and that $E[\exp(k\langle M\rangle_\infty)] < \infty$ for some $k > 1$. Then*

$$\|\mathcal{E}(M)\|_{\mathcal{H}^p} \le \frac{p}{p-1} E[\exp(k\langle M\rangle_\infty)]^{1/p-1/2},$$

where $p = 2k/(1+k) > 1$.

(ii) Suppose that X is a continuous local martingale and that $E[\exp(\frac{k}{2}\langle X\rangle_\infty)] < \infty$ for some $k > 1$. Then

$$\|\mathcal{E}(X)\|_{\mathcal{H}^p} \le \frac{p}{p-1} E\Big[\exp\Big(\frac{k}{2}\langle X\rangle_\infty\Big)\Big]^{(\sqrt{k}-1)/k},$$

where $p = k/(2\sqrt{k}-1) > 1$.

Proof. (i) Expanding the stochastic exponential, we obtain

$$\mathcal{E}(M)^2 = \mathcal{E}(2M + [M]) = \mathcal{E}(N + \langle M\rangle) = \mathcal{E}(\tilde{N})\mathcal{E}(\langle M\rangle)$$

where $N = 2M + [M] - \langle M\rangle$ and $\tilde{N}_t = \int_{[0,t]}(1 + \Delta\langle M\rangle_s)^{-1}dN_s$. It is easy to check that $\Delta\tilde{N} \ge -1$, so the local martingale $\mathcal{E}(\tilde{N})$ is nonnegative. Therefore, as $\mathcal{E}(\langle M\rangle)$ is increasing and is bounded by $\exp(\langle M\rangle)$, for any $p \ge 0$ we have the bound

$$|\mathcal{E}(M)|^p \le \mathcal{E}(\tilde{N})^{p/2}\exp(p\langle M\rangle/2).$$

For $p < 2$, applying Hölder's inequality with exponent $2/p$, we obtain, for any $t \le \infty$,

$$E\big[|\mathcal{E}(M)_t|^p\big] \le E\big[\mathcal{E}(\tilde{N})_t\big]^{p/2} E\Big[\exp\Big(\frac{p}{2-p}\langle M\rangle_t\Big)\Big]^{1-p/2}.$$

As $\mathcal{E}(\tilde{N})$ is a nonnegative local martingale it is a supermartingale, so $E[\mathcal{E}(\tilde{N})] \le 1$ and, taking $k = p/(2-p)$,

$$E\big[|\mathcal{E}(M)_t|^p\big] \le E\big[\exp(k\langle M\rangle_\infty)\big]^{1-p/2}.$$

By Doob's L^p inequality, we conclude

$$\begin{aligned}\|\mathcal{E}(M)\|_{\mathcal{H}^p} &= \|\sup_t(\mathcal{E}(M)_t)\|_{L^p} \le \frac{p}{p-1}\sup_t\|\mathcal{E}(M)_t\|_{L^p}\\ &\le \frac{p}{p-1}E\big[\exp(k\langle M\rangle_\infty)\big]^{1/p-1/2}.\end{aligned}$$

(ii) For any $\lambda \ge 0$, expanding the stochastic exponential we have

$$\mathcal{E}(X)^p = \mathcal{E}(p\lambda X)^{1/\lambda}\exp\Big(-\frac{1}{2}p(p\lambda - 1)\langle X\rangle\Big).$$

Applying Hölder's inequality, along with the fact $E[\mathcal{E}(p\lambda X)] \le 1$, we see

$$E\big[\mathcal{E}(X)^p\big] \le \exp\Big(-\frac{1}{2}p(p\lambda-1)\frac{\lambda}{\lambda-1}\langle X\rangle\Big)^{1-1/\lambda}.$$

Taking $\lambda = 2 - k^{-1/2}$ and $p = k/(2\sqrt{k}-1)$, we have

$$E[\mathcal{E}(X)^p]^{1/p} \le E\Big[\exp\Big(\frac{k}{2}\langle X\rangle_\infty\Big)\Big]^{(\sqrt{k}-1)/k},$$

and the result follows from Doob's L^p inequality as before. □

Remark 15.4.7. In fact, Lépingle and Mémin [124] show that (ii) holds for X discontinuous, provided $\Delta X \ge 0$, while Yan shows that (ii) holds for X discontinuous with $\Delta X \ge 0$, with the predictable variation $\langle X\rangle$ replaced by the optional variation $[X]$. See Yan [186] and references therein for further extensions and variations on these conditions.

One reason why this result is of interest is that it implies the following 'reverse Hölder inequality'.

Corollary 15.4.8 (Reverse Hölder Inequality). *Under the conditions of Theorem 15.4.6, let $Z = \mathcal{E}(X)$. Then there exists $c > 0$ such that, for every stopping time T,*

$$E[Z_\infty^p|\mathcal{F}_T] \le cZ_T^p.$$

Proof. On the set $\{Z_T = 0\}$, the result is trivial. We apply Theorem 15.4.6 to $\frac{Z_{T+t}}{Z_T} = \mathcal{E}(I_{]T,\infty]}\bullet X)$, in the filtration $\mathcal{G}_t = \mathcal{F}_{T+t}$, with the probability measure $P(\cdot|A)$, for any $A \in \mathcal{F}_T$ such that $Z_T \neq 0$ on A. The result follows. □

15.5 Extensions of Novikov's and Kazamaki's Criteria

We shall now explore how these conditions can be extended and applied in various contexts. We do this through a series of examples.

Example 15.5.1. Let H be a bounded predictable process and B a one-dimensional Brownian motion. Then $E[\mathcal{E}(H\bullet B)_t] = 1$ for all deterministic t, as

$$E[\exp(\langle H\bullet B\rangle_t/2)] = E\Big[\exp\Big(\frac{1}{2}\int_{[0,t]} H_s^2 ds\Big)\Big] \le \exp\Big(\frac{tk^2}{2}\Big) < \infty,$$

for k a bound on $|H|$. Furthermore, the stopped process satisfies $\mathcal{E}(H\bullet B)^t \in \mathcal{H}^p$ for all $p < \infty$.

Example 15.5.2. Let H be a predictable process and B a Brownian motion. Define

$$T_n = \inf\Big\{t : \int_{[0,t]} H_s^2 ds = n\Big\}.$$

Then

$$E[\exp(\langle H \bullet B\rangle_{T_n}/2)] = E\Big[\exp\Big(\frac{1}{2}\int_{[0,T_n]} H_s^2 ds\Big)\Big] \leq e^{n/2} < \infty,$$

so $E[\mathcal{E}(H \bullet B)_{T_n}] = 1$ for all n and $\mathcal{E}(H \bullet B)^{T_n} \in \mathcal{H}^p$ for every $p < \infty$.

Example 15.5.3. Let H be a predictable process, B a one-dimensional Brownian motion and $T > 0$. Suppose

$$\sup_{s \leq T} E[\exp(\delta H_s^2)] < \infty$$

for some $\delta > 0$. Then, by Jensen's inequality,

$$\exp\Big(\frac{1}{2}\int_{[0,T]} H_s^2 ds\Big) = \exp\Big(\frac{1}{T}\int_{[0,T]} \frac{TH_s^2}{2}\, ds\Big) \leq \frac{1}{T}\int_{[0,T]} \exp\Big(\frac{TH_s^2}{2}\Big) ds.$$

Therefore, if $T \leq 2\delta$, we have

$$E\Big[\exp\Big(\frac{1}{2}\int_{[0,T]} H_s^2 ds\Big)\Big] \leq \sup_{0 \leq s \leq T} E[\exp(\delta H_s^2)] < \infty,$$

and hence $E[\mathcal{E}(H \bullet B)_T] = 1$.

Now suppose $T > 2\delta$. In this case, write

$$Z_{a,b} = \exp\Big(\int_{[a,b]} H_s dB_s - \frac{1}{2}\int_{[a,b]} H_s^2 ds\Big) = \mathcal{E}((I_{[a,b]}H) \bullet B)_b,$$

so that

$$\mathcal{E}(H \bullet B)_T = Z_{0,t_1} Z_{t_1,t_2} \dots Z_{t_{n-1},t_n},$$

where $0 < t_1 < t_2 < \cdots < t_n = T$ and $\max_i(t_{i+1} - t_i) \leq 2\delta$. Then

$$E[Z_{t_i,t_{i+1}} | \mathcal{F}_{t_i}] = 1 \text{ a.s.}$$

and so,

$$\begin{aligned} E[\mathcal{E}(H \bullet B)_T] &= E\big[Z_{0,t_1} Z_{t_1,t_2} \dots Z_{t_{n-2},t_{n-1}} E[Z_{t_{n-1},t_n} | \mathcal{F}_{t_{n-1}}]\big] \\ &= E[Z_{0,t_1} Z_{t_1,t_2} \dots Z_{t_{n-2},t_{n-1}}] = \cdots = E[Z_{0,t_1}] = 1. \end{aligned}$$

It follows that $\mathcal{E}(H \bullet B)$ is a true martingale.

Example 15.5.4. For a local martingale X with $\Delta X > -1$, we know $\mathcal{E}(X)$ is a true martingale whenever there exists a sequence of stopping times $0 = \tau_0 \leq \tau_1 \leq \cdots \leq \tau_m = T$ such that

$$E[\exp(\langle X\rangle_{\tau_n} - \langle X\rangle_{\tau_{n-1}})] < \infty$$

for all $n \leq m$. (Simply take $Z_n = \mathcal{E}(I_{]\!]\tau_{n-1},\tau_n]\!]} \bullet X)$ and argue as in the previous example).

Example 15.5.5. Let X be a square integrable Lévy martingale (that is, a Lévy process which is a square integrable martingale) with $\Delta X > -1$. Then $\mathcal{E}(X)$ is a martingale, by applying the previous example (as the quadratic variation of X is of the form $d\langle X\rangle = kdt$ for some k).

The following example is a variant of one due to Beneš [11] (with a different proof).

Example 15.5.6. Suppose X is an m-dimensional martingale (or nonnegative submartingale) such that, for every t, $E[e^{a\|X_t\|^2}] < \infty$ for some $a > 0$. Suppose H is a predictable process such that

$$\|H_t\| \leq K(1 + X_t^*),$$

where $X_t^* = \sup_{s\leq t} \|X_s\|$. Examples of H satisfying this constraint include processes of the form $H_t = f(t, X_t, X_t^*)$ where f is a Borel measurable function of linear growth. Then, for some constants K_1, K_2,

$$\exp(\delta H_s^2) \leq K_1 \exp\left(\delta K_2 (X_s^*)^2\right).$$

By Jensen's inequality, $e^{a\|X\|^2}$ is a nonnegative submartingale, so Doob's maximal inequality implies $E[e^{a(X_t^*)^2}] < \infty$. Therefore,

$$E[\exp(\delta K_2 (X_T^*)^2)] < \infty$$

for small enough $\delta > 0$. Therefore, $\sup_{s\leq T} E[\exp(\delta H_s^2)] < \infty$ for this $\delta > 0$. By Example 15.5.3, the stochastic exponential

$$\mathcal{E}(H^\top \bullet X)$$

is then a martingale, where $H^\top \bullet X$ is the vector stochastic integral.

To apply the above example, the following lemma is sometimes useful.

Lemma 15.5.7. *Let $X = \sigma \bullet W$, for W an N-dimensional Brownian motion and σ a bounded predictable $\mathbb{R}^{m\times N}$-valued process (in particular $\int_{[0,t]} \sigma_s \sigma_s^\top ds \leq k(t)$ for some deterministic function k). Then for each t, we know $E[e^{a\|X_t\|^2}] < \infty$ for some $a > 0$.*

Proof. We begin by assuming that X is scalar and $\sigma = 1$, that is, X is a Brownian motion. Then, we know $X_t \sim N(0,t)$. Therefore, for any $a < (2t)^{-1}$

$$E[e^{aX_t^2}] = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi t}} e^{ax^2 - \frac{x^2}{2t}} dx$$
$$= \sqrt{\frac{\rho}{t}} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\rho}} e^{-\frac{x^2}{2\rho}} dx = \sqrt{\frac{\rho}{t}} = \frac{1}{\sqrt{1-2at}} < \infty.$$

where $1/\rho = 1/t - 2a > 0$.

Now assume only that X is scalar. Then we can define the time change $C(t) = \inf\{s : \int_{[0,s]} \sigma_u^2 du \geq t\}$, so that $B_t := X_{C(t)}$ is a Brownian motion (in the time changed filtration). As X does not vary on any interval where C is constant (Exercise 11.7.11), we define $C^{(-1)}(t) = \inf\{s : C(t) \geq s\}$ and observe $B_\tau = X_t$, where $\tau := C^{(-1)}(t) \leq k(t)$. By Jensen's inequality, as $E[e^{aB_t^2}] < \infty$ we know e^{aB^2} is a submartingale, so by the optional stopping theorem and our previous results,

$$E[e^{aX_t^2}] = E[e^{aB_\tau^2}] \leq E[e^{aB_{k(t)}^2}] = (1 - 2ak(t))^{-1/2} < \infty.$$

Finally, for X a vector valued process with components X^i, we know that $\|X\|^2 = \sum_{i=1}^m (X^i)^2$, and we have the general inequality $\prod_{i=1}^m b_i \leq \frac{1}{m} \sum_{i=1}^m b_i^m$ for $b_i \geq 0$ (a variation of the arithmetic-geometric mean inequality, which follows from Jensen's inequality). Therefore,

$$e^{a\|X_t\|^2} \leq e^{a\sum_{i=1}^m (X_t^i)^2} \leq \frac{1}{m} \sum_{i=1}^m e^{am(X_t^i)^2}.$$

The result then follows from the scalar case, provided $a < (2mk(t))^{-1}$. □

Example 15.5.6 can also be obtained using the technique of the following example, which is a variant of that presented in Protter and Shimbo [153]. Note that the following example places a stronger growth bound on H, but weaker integrability conditions on X (which are often necessary in the presence of jumps, see [153]).

Example 15.5.8. Let X be an m-dimensional martingale (or nonnegative submartingale) with $|\langle X\rangle_t - \langle X\rangle_s| \leq k(t-s)$ componentwise for all $t \geq s$, and H an X-integrable process with $H\Delta X > -1$. Suppose there exists $a > 0$ such that $E[e^{a\|X_t\|}] < \infty$ for all t, and a constant K such that

$$|H_t| \leq K(1 + (X_t^*)^{1/2}),$$

where $X_t^* = \sup_{s\leq t} \|X_s\|$. Then, for some fixed $k > 0$, for any $t > s$,

$$\langle H \bullet X\rangle_t - \langle H \bullet X\rangle_s = \int_{]s,t]} H_u^2 d\langle X\rangle_u \leq k \int_{]s,t]} H_u^2 du$$
$$\leq kK^2(t-s)(1 + (X_T^*)^{1/2})^2 \leq 2kK^2(t-s)(1 + X_T^*).$$

Therefore, writing $c = 2kK^2(t-s)$,

$$E[\exp(\langle H \bullet X\rangle_t - \langle H \bullet X\rangle_s)] \leq kK^2 E[\exp(cX_T^*)].$$

As we know $E[e^{a\|X_t\|}] < \infty$ for some $a > 0$, from Jensen's inequality we see that $e^{a\|X\|}$ is a submartingale. By Doob's maximal inequality, provided $c \leq a$, we have $E[\exp(cX_T^*)] \leq 4E[\exp(c\|X_T\|)] < \infty$. Therefore, by Example 15.5.4, taking $t - s$ sufficiently small, we see that $\mathcal{E}(H \bullet X)$ is a martingale.

The next example is closely related to the previous one, but has a more abstract restriction, and is in terms of random measures.

Example 15.5.9. Let μ be a random measure on a Blackwell space $\mathcal{Z}$ with compensator $\mu_p(d\zeta, dt) = \nu(d\zeta)dt$, and W an N-dimensional Brownian motion. Let α be a W-integrable process and β be a $\tilde{\mu}$-stochastically integrable map with $\beta > -1$. Let X be a submartingale such that there exists $a > 0$ with $E[e^{a\|X_T\|}] < \infty$. If, up to indistinguishability,

$$\|\alpha_t\|^2 + \int_{\mathcal{Z}} \beta_t^2(\zeta)\nu(d\zeta) \leq K(1 + (X_t^*)^{1/2})$$

it follows that $\mathcal{E}(\alpha \bullet W + \beta * \tilde{\mu})$ is a true martingale. The proof of this is almost identical to the previous example, given the inequality

$$\langle \alpha \bullet W + \beta * \tilde{\mu}\rangle_t - \langle \alpha \bullet W + \beta * \tilde{\mu}\rangle_s \leq 2K^2(t-s)(1 + X_T^*).$$

Remark 15.5.10. Note that in the two previous examples, if $\mathcal{E}(Y)$ is the stochastic exponential of interest, then by taking smaller timesteps we can also show that, for any $k > 1$,

$$E[\exp(k(\langle Y\rangle_t - \langle Y\rangle_s))] \leq k' E[\exp(cX_T^*)]$$

for some constant $k' > 1$. Therefore, from Theorem 15.4.5 and simple calculations, the stopped processes $\mathcal{E}(Y^T) = \mathcal{E}(Y)^T$ are in $\mathcal{H}^p$ for any $p < \infty$, with an $\mathcal{H}^p$ norm which is bounded by a function of $E[e^{c\|X_T\|}]$. Hence the Reverse Hölder inequality still holds.

Remark 15.5.11. Under the conditions of Example 15.5.6, we know that $\|X\|^2$ is a nonnegative submartingale. Hence, replacing X with $\|X\|^2$, we immediately see that Example 15.5.9 implies Example 15.5.6.

In order to make use of the previous examples, the following lemma is sometimes useful.

Lemma 15.5.12. *Suppose $X = x + \sigma \bullet W + \int_{\mathbb{R}^m} \zeta\, \tilde{\mu}(d\zeta, dt)$, where $x \in \mathbb{R}^m$, W is an N-dimensional Brownian motion, $\tilde{\mu} = \mu - \mu_p$ for $\mu_p(d\zeta, dt) = \nu(d\zeta)dt$, σ is a* bounded *predictable process and ν is a* compactly supported *measure on $\mathbb{R}^m$. Then $E[e^{a\|X_t\|}] < \infty$ for every $a < \infty$ and $t < \infty$.*

Proof. Suppose first that X is a scalar process. As ν is compactly supported and defines the compensator of μ, the jumps of X are bounded, up to an evanescent set. Therefore, $e^{a\Delta X} - 1 - a\Delta X \leq ka^2\Delta X^2/2$ for some $k > 1$. Using Itô's rule, we write

$$\begin{aligned} e^{aX_t} &= 1 + \int_{]0,t]} ae^{aX_{s-}} dX_s + \int_{[0,t]} \frac{a^2}{2} e^{aX_{s-}} d\langle X^c\rangle_s \\ &\quad + \sum_{0<s\leq t} e^{aX_{s-}}(e^{a\Delta X_s} - 1 - a\Delta X_s) \\ &\leq 1 + \int_{]0,t]} ae^{aX_{s-}} dX_s + k\frac{a^2}{2}\int_{[0,t]} e^{aX_{s-}} d[X]_s. \end{aligned}$$

Let $\{\tau_n\}_{n\in\mathbb{N}}$ be a localizing sequence for the local martingales $\{\int_{]0,t]} e^{aX_{s-}} dX_s\}_{t\geq 0}$ and $\{\int_{[0,t]} e^{aX_{s-}} d([X]_s - \langle X\rangle_s)\}_{t\geq 0}$. By assumption, we know that there exists $K > 0$ such that

$$\frac{d\langle X\rangle}{dt} = \sigma^2 + \int_{\mathcal{Z}} g(\zeta)^2 \nu(d\zeta) \leq K.$$

Therefore,

$$\begin{aligned} E[I_{\{t\leq\tau_n\}} e^{aX_t}] \leq E[e^{aX_{t\wedge\tau_n}}] &\leq 1 + \frac{a^2}{2} E\Big[\int_{]0,t\wedge\tau_n]} e^{aX_{s-}} d\langle X\rangle_s\Big] \\ &\leq 1 + \frac{a^2K}{2}\int_{]0,t]} E[I_{\{t\leq\tau_n\}} e^{aX_{s-}}]dt. \end{aligned}$$

From Grönwall's inequality, it follows that $E[I_{\{t\leq\tau_n\}}e^{aX_t}] \leq e^{a^2Kt/2}$, and by monotone convergence,

$$E[e^{aX_t}] = \lim_n E[I_{\{t\leq\tau_n\}} e^{aX_t}] \leq e^{a^2Kt/2} < \infty.$$

As a can be any real number, the inequality $e^{a|x|} \leq e^{ax} + e^{-ax}$ yields the result for X scalar.

The result for X vector-valued then follows using the arithmetic-geometric mean inequality, essentially in the same way as in Lemma 15.5.7. □

Finally we show that the constant $\frac{1}{2}$ which appears in Novikov's condition $E\big[\exp\big(\frac{1}{2}\langle M\rangle_T\big)\big] < \infty$ cannot be weakened, by giving an example where, for an arbitrary $\epsilon \in]0, 1/2[$,

$$E\Big[\exp\Big(\frac{1}{2} - \epsilon\Big)\int_{[0,\infty[} H_s^2 ds\Big] < \infty$$

yet $E[\mathcal{E}(H\bullet B)_\infty] \neq 1$.

Example 15.5.13. Let B be a one-dimensional Brownian motion, $0 < \epsilon < \frac{1}{2}$, and $a > 0$. Write

$$T_\epsilon = \inf \big\{t : B_t - (1-\epsilon)t = -a\big\},$$
$$T_\epsilon^n = \inf\{t : B_t \geq n\} \wedge T_\epsilon.$$

By the law of the iterated logarithm, in particular Remark 5.5.15, we see that T_ϵ and T_ϵ^n are a.s. finite. We first show that

$$E\Big[\exp\Big(\Big(\frac{1}{2} - \epsilon\Big)T_\epsilon^n\Big)\Big] = V_n(0), \tag{15.3}$$

where

$$V_n(x) = \frac{e^x\big(e^{-2\epsilon n} - e^{-(1-2\epsilon)a-n}\big)}{e^{-(a+2\epsilon n)} - e^{-(1-2\epsilon)a}} + \frac{e^{(1-2\epsilon)x}\big(e^{-(a+n)} - 1\big)}{e^{-(a+2\epsilon n)} - e^{-(1-2\epsilon)a}}$$

is the solution of the differential equation

$$V_n''(x) - 2(1-\epsilon)V_n'(x) + (1-2\epsilon)V_n(x) = 0, \tag{15.4}$$

satisfying $V_n(-a) = V_n(n) = 1$. To establish (15.3) consider the function $V_n(x)e^{(1/2-\epsilon)t}$. By Itô's rule, if $X_t = B_t - (1-\epsilon)t$, for any integer $N \geq 1$,

$$\begin{aligned} &V_n(X_{T_\epsilon^n \wedge N}) \exp\Big[\Big(\frac{1}{2} - \epsilon\Big)(T_\epsilon^n \wedge N)\Big] \\ &\quad = V_n(0) + \int_{[0, T_\epsilon^n \wedge N]} V_n'(X_s) \exp\Big[\Big(\frac{1}{2} - \epsilon\Big)s\Big] dB_s, \end{aligned}$$

using (15.4). Now $V_n'(x)$ is bounded if $-a \leq x \leq n$, so

$$E\Big[\int_{[0, T_\epsilon^n \wedge N]} V_n'(X_s) \exp\Big(\Big(\frac{1}{2} - \epsilon\Big)s\Big) dB_s\Big] = 0.$$

Therefore,

$$E\Big[V_n(X_{T_\epsilon^n \wedge N}) \exp\Big(\Big(\frac{1}{2} - \epsilon\Big)(T_\epsilon^n \wedge N)\Big)\Big] = V_n(0). \tag{15.5}$$

From the explicit form for $V_n(x)$ we see that

$$0 < \inf_{-a \leq x \leq n} V_n(x) < \sup_{-a \leq x \leq n} V_n(x) < \infty,$$

and so

$$E\Big[\exp\Big(\Big(\frac{1}{2} - \epsilon\Big)(T_\epsilon^n \wedge N)\Big)\Big] \leq \frac{V_n(0)}{\inf_{-a \leq x \leq n} V_n(x)} < \infty.$$

Consequently, letting $N \to \infty$, we have

$$E\Big[\exp\Big(\Big(\frac{1}{2}-\epsilon\Big)T^n_\epsilon\Big)\Big] \le \frac{V_n(0)}{\inf_{-a\le x\le n} V_n(x)} < \infty.$$

Furthermore, because

$$V_n(X_{T^n_\epsilon\wedge N})\exp\Big(\Big(\frac{1}{2}-\epsilon\Big)(T^n_\epsilon\wedge N)\Big) \le \exp\Big(\Big(\frac{1}{2}-\epsilon\Big)T^n_\epsilon\Big)\sup_{-a\le x\le n} V_n(x),$$

by dominated convergence we can let $N\to\infty$ in (15.5) to obtain

$$E\Big[V_n(X_{T^n_\epsilon})\exp\Big(\Big(\frac{1}{2}-\epsilon\Big)T^n_\epsilon\Big)\Big] = V_n(0).$$

However, $V_n(X_{T^n_\epsilon}) = 1$ a.s., so we have

$$E\Big[\exp\Big(\Big(\frac{1}{2}-\epsilon\Big)T^n_\epsilon\Big)\Big] = V_n(0).$$

Letting $n\to\infty$, we see that

$$E\Big[\exp\Big(\Big(\frac{1}{2}-\epsilon\Big)T_\epsilon\Big)\Big] = \exp\big((1-2\epsilon)a\big) < \infty.$$

However, if $H^\epsilon_t := I_{\{t\le T_\epsilon\}}$, then

$$\begin{aligned}\mathcal{E}(H^\epsilon\bullet B)_\infty &= \exp\Big(B_{T_\epsilon}-\frac{1}{2}T_\epsilon\Big) = \exp\big(B_{T_\epsilon}-(1-\epsilon)T_\epsilon\big)\exp\Big(\Big(\frac{1}{2}-\epsilon\Big)T_\epsilon\Big)\\ &= e^{-a}\exp\Big(\Big(\frac{1}{2}-\epsilon\Big)T_\epsilon\Big).\end{aligned}$$

Therefore

$$E\big[\mathcal{E}(H^\epsilon\bullet B)_{T_\varepsilon}\big] = e^{-2\varepsilon a} < 1.$$

Remark 15.5.14. Protter and Shimbo [153] also show that the constant 1 in front of $\langle X^d\rangle$ in the condition with jumps $E[\exp(\frac{1}{2}\langle X^c\rangle + \langle X^d\rangle)] \le \infty$ cannot be reduced, unless a stronger bound is placed on the jumps of X. If the jumps are known to be *nonnegative*, then a coefficient $1/2$ can be used, giving the condition $E[\exp(\frac{1}{2}(\langle X^c\rangle + \langle X^d\rangle))] \le \infty$. Sokol [169] gives optimal coefficients for the intermediate cases.

15.6 Exercises

Exercise 15.6.1. Let B be a Brownian motion. Show that $\mathcal{E}(\alpha B)$ is not uniformly integrable for any constant $\alpha \ne 0$, but that for any deterministic $T<\infty$ we know $\mathcal{E}(\alpha I_{[0,T]}\bullet B)$ is uniformly integrable.

Exercise 15.6.2. Let $X = \{X^1, X^2, \ldots, X^m\}$ be an m-dimensional continuous martingale, with $\langle X^i, X^j \rangle = A_{ij}t$, for A a symmetric strictly positive definite matrix. Give general sufficient conditions on a predictable process H in $\mathbb{R}^m$ such that there exists a measure Q with

$$\Big\{X_t - \int_{[0,t]} H_s ds\Big\}_{t\geq 0}$$

a continuous Q-martingale. Under what conditions is this process a Q-Brownian motion?

Exercise 15.6.3. Let

$$X_t = X_0 + \int_{]0,t]} H_s ds + \int_{]0,t]} \sigma_s dB_s + \int_{]0,t]} \eta_s dW_s$$

where B and W are Brownian motions, H, σ and η are bounded predictable processes, and σ, η are bounded above zero. Construct at least two distinct measures such that $\{X_t - rt\}_{t\geq 0}$ is a martingale under the new measures, for r a constant.

Exercise 15.6.4. Consider the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$, where $\{\mathcal{F}_t\}_{t\geq 0}$ is the (completed, right-continuous) filtration generated by a Brownian motion B. Let H be a predictable process such that

$$E\Big[\exp\Big(\frac{1}{2}\int_{[0,\infty]} H_s^2 ds\Big)\Big] < \infty \text{ for all } t.$$

Using the martingale representation theorem, show that the measure Q equivalent to P under which $\{B_t - \int_{[0,t]} H_s ds\}_{t\geq 0}$ is a martingale is unique.

Exercise 15.6.5. Let B be a Brownian motion, and for constants x, y, let X be the process defined for $t \leq 1$ by

$$X_t = yt + (1-t)\Big(x + \int_{[0,t]} (1-s)^{-1} dB_s\Big)$$

with $X_t = X_1$ for $t \geq 1$.

(i) Show that X is a well-defined semimartingale and determine its semimartingale decomposition and its behaviour as $t \to 1$.
(ii) Show that there is a sequence of stopping times $T_n \to 1$, and a family of measures $\{Q_n\}_{n\in\mathbb{N}}$ equivalent to P, such that X^{T_n} is a Q_n-Brownian motion for all n.
(iii) Show that there exists no measure Q equivalent to P such that X is a Q-martingale.

(Here, X is called the Brownian bridge between x and y.)

Exercise 15.6.6. Let B and W be independent Brownian motions, and let $X_t = \int_{[0,t]} f(W_t)dt + B_t$, for f a Borel measurable function of linear growth. Show that there exists a measure Q under which X and W are independent.

Exercise 15.6.7. Let N be a Poisson process and B be a Brownian motion. Assuming $\mathcal{E}(H \bullet B)$ and $\mathcal{E}(\eta \bullet \tilde{N})$ are nonnegative, uniformly integrable martingales, define the measure changes $dQ/dP = \mathcal{E}(H \bullet B)_\infty$ and $dQ'/dP = \mathcal{E}(\eta \bullet \tilde{N})_\infty$. Show that B is a Q'-Brownian motion and N is a Q-Poisson process.

16

Lipschitz Stochastic Differential Equations

As is now usual, all (in)equalities in this chapter should be read as 'up to an evanescent set', unless otherwise specified, martingales are càdlàg and we assume we have a filtered probability space satisfying the usual conditions. In this chapter, we consider stochastic differential equations (SDEs), that is, m-dimensional processes X satisfying an equation of the form

$$X_t = H_t + \int_{[0,t]} f(\omega, u, X) dY_u \tag{16.1}$$

where Y is a d-dimensional semimartingale and f a prescribed function taking values in $\mathbb{R}^{m\times d}$. The integral here is the vector stochastic integral considered in Section 12.5. An important special case of this is

$$X_t = X_0 + \int_{[0,t]} f(\omega, u, X_{u-}) du + \int_{]0,t]} \tilde{f}(\omega, u, X_{u-}) dW$$

where W is a Brownian motion.

Fundamentally, the term "stochastic *differential* equation" is a misnomer, as these equations are 'integral' rather than 'differential' equations. Nevertheless, given Remark 12.3.16 (see also Remark 14.2.7), we can express (16.1) in a 'differential' form,

$$dX_t = dH_t + f(\omega, u, X) dY_u,$$

and, together with the initial condition $X_{0-} = H_{0-}$, these formulations can be seen to be equivalent. The main result of this chapter is Theorem 16.3.11, which gives the existence and uniqueness of solutions to this equation whenever f satisfies an appropriate Lipschitz continuity condition.

S.N. Cohen, R.J. Elliott, *Stochastic Calculus and Applications*,
Probability and Its Applications, DOI 10.1007/978-1-4939-2867-5_16

Definition 16.0.1. *For X a measurable process with left limits and T a stopping time, we define the process*

$$X^{T-} = I_{[\![0,T[\![}X + I_{[\![T,\infty[\![}X_{T-} = \begin{cases} X_t & \text{for } t < T \\ X_{T-} & \text{for } t \geq T. \end{cases}$$

We say that a property holds prelocally *for X if there exists a sequence of stopping times $T_n \to \infty$ such that the property holds for each X^{T_n-}.*

The surprising power of this definition will become apparent in Lemma 16.2.7.

Remark 16.0.2. One can easily check that if X is progressive/optional/predictable/càdlàg/left continuous/a (special) semimartingale, so is X^{T-}; however X being a local martingale does not generally imply that X^{T-} is a local martingale. Conversely, if X is prelocally progressive/optional/predictable/càdlàg/left continuous/a semimartingale, the same must hold for X without prelocalization. On the other hand, we shall see that *any* semimartingale is prelocally a special semimartingale, so checking that X is prelocally a special semimartingale does not guarantee that X is a special semimartingale.

Definition 16.0.3. *Write $\mathcal{D}$ for the space of càdlàg $\mathbb{R}^m$-valued adapted processes. A map*

$$f : \Omega \times [0, \infty[\times \mathcal{D} \to \mathbb{R}^d$$

will be called the coefficient *of an SDE in Y if*

(i) for any $x \in \mathcal{D}$, the process $\{f(\omega, t, x)\}_{t\geq 0}$ is Y-integrable (and hence predictable),
(ii) f is 'non-anticipative', that is, for any stopping time T, if $X^{T-} = \tilde{X}^{T-}$, then $f(\omega, t, X) = f(\omega, t, \tilde{X})$ on the set $\{t \leq T\}$.

For notational simplicity, we write $f_t(X) = f(\omega, t, X)$, and $f(X)$ for the process $\{f_t(X)\}_{t\geq 0}$, whenever this does not lead to confusion.

Remark 16.0.4. Note that we allow f to depend on the whole path of X up (but not including) time t. Our continuity assumption below will make it clear that f depends only on the *realized* path $\{X_s(\omega)\}_{s<t}$, and not on the path as a random object.

Remark 16.0.5. Through the non-anticipative assumption, we formally insist that, at time t, f does *not* depend on X_t, but only on X^{t-} or, equivalently, on the set $\{X_s\}_{s<t}$. This is important when dealing with jump processes. However, when Y is continuous, as a càdlàg process has only countably many jumps, and the stochastic integral with respect to a continuous process will not charge any countable set, this condition can be relaxed without altering the equation considered.

Remark 16.0.6. Note that f being non-anticipative does not necessarily imply that for a stopping time T we should have $f_{T-}(X^{T-}) = f_T(X^{T-})$, but does imply that $f_{T-}(X^{T-}) = f_{T-}(X)$. On the other hand, the process $f(X)$ is required be predictable, as it is Y-integrable.

Definition 16.0.7. *Let Y be a semimartingale and H be a càdlàg $\mathbb{R}^m$-valued adapted process in the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,\infty[}, P)$ (that is, $H \in \mathcal{D}$). A process $X \in \mathcal{D}$ is called a* solution *to the SDE if*

$$X_t = H_t + \int_{[0,t]} f(\omega, u, X)dY_u \qquad \text{a.s.} \qquad \textit{for all } t \geq 0 \tag{16.2}$$

If H is assumed to be a semimartingale, then a solution X is clearly also a semimartingale.

Remark 16.0.8. Recalling Lemma 3.2.10, as our processes are càdlàg we could replace 'a.s. for all $t \geq 0$' by 'up to indistinguishability' or '$dP \times dt$ almost everywhere' in the previous definition without any change.

We now give a few examples of SDEs for which we can easily demonstrate the existence of solutions.

Example 16.0.9. If $g : \mathbb{R} \to \mathbb{R}$ is bounded and measurable and we take $X_t = x_0$ for $t \leq 0$, then, for any $\epsilon > 0$, the 'delayed' SDE

$$X_t = x_0 + \int_{[0,t]} g(X_{t-\epsilon})dY_t$$

has a unique solution, which can be obtained by integrating over the intervals $[0, \epsilon]$, $]\epsilon, 2\epsilon]$, ..., etc.

Example 16.0.10. We have already seen the equation $X_t = 1 + \int_{[0,t]} X_{s-}dY_s$, for Y a semimartingale, has unique solution $X = \mathcal{E}(Y)$.

Example 16.0.11. Let W be a Brownian motion and consider the equation

$$X_t = \int_{[0,t]} 3X_s^{1/3}ds + \int_{[0,t]} 3X_s^{2/3}dW_s.$$

Let $S \leq T$ be any stopping times such that $W_S = W_T = 0$ on $\{T < \infty\}$, then the process

$$X_t = I_{t\in[S,T]}W_t^3$$

is a solution. (Note that the solution is not unique.)

Example 16.0.12. Let W be a Brownian motion and consider the equation

$$X_t = 1 - \int_{[0,t]} \Big(\int_{[0,s]} \frac{1}{(s-u+1)^2}dW_u\Big)X_s ds + \int_{[0,t]} X_s dW_s.$$

At time t, this equation depends nontrivially on the entire path $\{W_s\}_{s<t}$ and has solution

$$X_t = \exp\Big(\int_{[0,t]} \frac{1}{t-s+1} dW_s\Big).$$

Example 16.0.13. Let N be a jump process with jumps of size 1 (for example, a Poisson process). Then for some predictable nondecreasing process λ_t with $\lambda_{0-} = 0$, consider the equation

$$X_t = \int_{[0,t]} ((1 + N_{s-} - \lambda_s)^+ - X^*_{s-})^+ dN_s,$$

which has solution

$$X_t = ((N_t - \lambda_t)^+)^*.$$

16.1 A Simple Case

We now consider a simple continuous and scalar case, where our SDE has the particular form

$$X_t = X_0 + \int_{[0,t]} \mu(\omega, s, X_{s-}) ds + \int_{[0,t]} \sigma(\omega, s, X_{s-}) dW_s, \tag{16.3}$$

for W a Brownian motion. When μ and σ are deterministic functions, such a process is commonly called an *Itô process*. This case arises in many applications and, in this setting, one can avoid many technicalities which arise for general semimartingales. For notational simplicity, we write $\mu_s(x)$ for $\mu(\omega, s, x)$, and similarly $\sigma_s(x)$.

Definition 16.1.1. *We say that a function $\mu : \Omega \times [0,\infty] \times \mathbb{R} \to \mathbb{R}$ is $dt \times dP$ uniformly Lipschitz if there exists a constant $K \in [0,\infty[$ such that*

$$|\mu_t(x) - \mu_t(x')| \leq K|x - x'| \qquad dt \times dP - a.s.$$

The following result is not the most general we shall prove; however, it is a useful simple case.

Theorem 16.1.2. *Let μ and σ be uniformly Lipschitz functions with*

$$\int_{[0,T]} E\big[|\mu_s(0)|^p + |\sigma_s(0)|^p\big] ds < \infty$$

for some $p \geq 2$. Then (16.3) has a unique[1] solution X.

[1] Here and elsewhere, when stating that an equation has a unique solution, we mean both that a solution exists and that the solution is unique.

Our method of proof depends on establishing a useful stability result for this equation, under some additional assumptions.

Remark 16.1.3. The following basic inequalities will be useful in many of our calculations: for any $a, b \in \mathbb{R}$, we know that $2ab \leq a^2 + b^2$ and $(a+b)^2 \leq 2(a^2+b^2)$. More generally, for any $a, b, c \in \mathbb{R}$ and any $p \geq 1$ we know that $(a+b)^p \leq 2^{p-1}(|a|^p + |b|^p)$ and $(a+b+c)^p \leq 3^{p-1}(|a|^p + |b|^p + |c|^p)$.

Lemma 16.1.4. *Let X be a solution of (16.3) with μ, σ functions satisfying the linear growth condition*

$$|\mu_s(x)| \leq \tilde{\mu}_s + K|x|, \qquad |\sigma_s(x)| \leq \tilde{\sigma}_s + K|x|,$$

for some constant K and some processes $\tilde{\mu}$ and $\tilde{\sigma}$. (Note that if μ and σ are uniformly Lipschitz, then $\tilde{\mu} = |\mu(0)|$ and $\tilde{\sigma} = |\sigma(0)|$ satisfy these requirements, with K the Lipschitz constant of the functions.)

Then X is continuous and, for any deterministic time T and any $p \geq 2$, there exists a real constant C depending on T, K and p such that

$$E[(X_T^*)^p] \leq C\Big(E\big[|X_0|^p\big] + \int_{[0,T]} E\big[|\tilde{\mu}_s|^p + |\tilde{\sigma}_s|^p\big]ds\Big).$$

Proof. Continuity of X follows immediately from the continuity of the integrals in (16.3). If $E\big[|X_0|^p\big] + \int_{[0,T]} E\big[|\tilde{\mu}_s|^p + |\tilde{\sigma}|^p\big]ds = \infty$, then the result is trivial, so we can assume this quantity is finite. In the following, C denotes a constant which can depend on T, K and p, and may vary from line to line. We observe, for $t \leq T$,

$$\begin{aligned}
&E[(X_t^*)^p] \\
&= E\Big[\sup_{r\leq t}\Big|X_0 + \int_{[0,r]} \mu_s(X_s)ds + \int_{[0,r]} \sigma_s(X_s)dW_s\Big|^p\Big] \\
&\leq CE\big[|X_0|^p\big] + C\int_{[0,t]} E\big[|\mu_s(X_s)|^p\big]ds + CE\Big[\Big(\Big(\int_{[0,\cdot]} \sigma_s(X_s)dW_s\Big)_t^*\Big)^p\Big] \\
&\leq CE\big[|X_0|^p\big] + C\int_{[0,t]} E\big[|\mu_s(X_s)|^p\big]ds + CE\Big[\Big(\int_{[0,t]} (\sigma_s(X_s))^2 ds\Big)^{p/2}\Big] \\
&\leq CE\big[|X_0|^p\big] + C\int_{[0,t]} E\big[|\tilde{\mu}_s|^p + K^p|X_s|^p\big]ds + C\int_{[0,t]} E\big[|\tilde{\sigma}|^p + K^p|X_s|^p\big]ds \\
&\leq C\Big(E\big[|X_0|^p\big] + \int_{[0,t]} E\big[|\tilde{\mu}_s|^p + |\tilde{\sigma}|^p\big]ds\Big) + C\int_{[0,t]} E[(X_s^*)^p]ds
\end{aligned}$$

where on the third and fifth lines we have used Jensen's inequality, and on the fourth we have used the BDG inequality. By Grönwall's inequality (Corollary 15.1.7), this implies that

$$E[(X_T^*)^p] \leq C\Big(E\big[|X_0|^p\big] + \int_{[0,T]} E\big[|\tilde{\mu}_s|^p + |\tilde{\sigma}|^p\big]ds\Big)e^{CT} < \infty.$$

Writing C for Ce^{CT} gives the result. □

Remark 16.1.5. One can easily check that this argument works equally well when μ and σ are permitted to depend on the entire path of X, with the growth condition $|\mu_t(X)| \leq \mu_s + KX_t^*$ for some K, and similarly for σ. One can also check that a multivariate version of this result holds (in finite dimensions), with $E[(X_t^*)^p]$ replaced by $E[(\|X\|_t^*)^p]$, for $\|\cdot\|$ the Euclidean norm or max norm ($\|x\| = \max_i |x^i|$).

One approach to solving SDEs is to apply the above argument to the difference of two SDEs, and then use the resulting estimate to solve the SDE over a short time interval. The existence of a solution for all time follows by pasting. We shall use this approach in a general setting (Section 16.3). Here, we can instead give a more elegant approach using the following, more careful, estimate.

Lemma 16.1.6. *Let $\mu, \tilde{\mu}, \sigma, \tilde{\sigma}$ be uniformly Lipschitz functions satisfying the conditions of Theorem 16.1.2. Let X and $\tilde{X}$ be solutions of (16.3) with coefficients (μ, σ) and $(\tilde{\mu}, \tilde{\sigma})$ respectively. For any $\beta > 0$,*

$$\begin{aligned}
&E[e^{-\beta t}(X_t - \tilde{X}_t)^2] \\
&\leq e^{-(\beta-1-4K^2)t}\Big(E[(X_0 - \tilde{X}_0)^2] + \int_{[0,t]} 2e^{-\beta s}E[(\mu_s(X_s) - \tilde{\mu}_s(X_s))^2] \\
&\qquad\qquad + 2e^{-2\beta s}E[(\sigma_s(X_s) - \tilde{\sigma}_s(X_s))^2]ds\Big).
\end{aligned}$$

Proof. Write $Y_t = e^{-\beta t}(X_t - \tilde{X}_t)^2$. As our processes are continuous, using the product rule we see

$$\begin{aligned}
Y_t = (X_0 - \tilde{X}_0)^2 &- \beta \int_{[0,t]} e^{-\beta s}(X_s - \tilde{X}_s)^2 ds \\
&+ 2\int_{[0,t]} e^{-\beta s}(X_s - \tilde{X}_s)(\mu_s(X_s) - \tilde{\mu}_s(\tilde{X}_s))ds \\
&+ 2\int_{[0,t]} e^{-\beta s}(X_s - \tilde{X}_s)(\sigma_s(X_s) - \tilde{\sigma}_s(\tilde{X}_s))dW_s \\
&+ \int_{[0,t]} e^{-2\beta s}(\sigma_s(X_s) - \tilde{\sigma}_s(\tilde{X}_s))^2 ds.
\end{aligned} \tag{16.4}$$

Calculating the quadratic variation of Y, we have

$$\langle Y \rangle_t = 4\int_{[0,t]} e^{-2\beta s}(X_s - \tilde{X}_s)^2(\sigma(\omega, s, X_s) - \tilde{\sigma}(\omega, s, \tilde{X}_s))^2 ds.$$

From Lemma 16.1.4, we see that $E[((X - \tilde{X})_t^*)^2] < \infty$, so $E[\int_{[0,t]}(X_s - \tilde{X}_s)^2 ds] < \infty$ and

$$\begin{aligned}
E[\langle Y \rangle_t^{1/2}] &\leq 4E\Big[(X - \tilde{X})_t^*\Big(\int_{[0,t]}\big((\sigma_s(0))^2 + K^2(X_s - \tilde{X}_s)^2\big)ds\Big)^{1/2}\Big] \\
&\leq 2E\Big[\big((X - \tilde{X})_t^*\big)^2 + \int_{[0,t]}\big((\sigma_s(0))^2 + K^2(X_s - \tilde{X}_s)^2\big)ds\Big] \\
&< \infty.
\end{aligned}$$

By the BDG inequality, we see that the 'dW' term in (16.4) is a true martingale.

Write $\delta\mu_s = \mu_s(X_s) - \tilde{\mu}_s(X_s)$ and $\delta\sigma_s = \sigma_s(X_s) - \tilde{\sigma}_s(X_s)$. Taking an expectation and applying the Cauchy–Schwarz inequality and those of Remark 16.1.3 to (16.4), we know that

$$\begin{aligned}
E[Y_t] \leq E[Y_0] &- \beta \int_{[0,t]} E[Y_s]ds + \int_{[0,t]} E[Y_s]ds \\
&+ \int_{[0,t]} e^{-\beta s} E[(\mu_s(X_s) - \tilde{\mu}_s(\tilde{X}_s))^2]ds \\
&+ \int_{[0,t]} e^{-2\beta s} E[(\sigma_s(X_s) - \tilde{\sigma}_s(\tilde{X}_s)^2]ds \\
\leq{}& E[Y_0] - (\beta - 1) \int_{[0,t]} E[Y_s]ds + \int_{[0,t]} \big(2E[(\delta\mu_s)^2] + 2K^2 E[Y_s]\big)ds \\
&+ \int_{[0,t]} \big(e^{-\beta s} 2E[(\delta\sigma)^2] + 2K^2 E[Y_s]\big)ds \\
\leq{}& E[Y_0] - (\beta - 1 - 4K^2) \int_{[0,t]} E[Y_s]ds \\
&+ \int_{[0,t]} \big(2e^{-\beta s} E[(\delta\mu_s)^2] + e^{-2\beta s} 2E[(\delta\sigma_s)^2]\big)ds.
\end{aligned}$$

Applying Grönwall's inequality (Corollary 15.1.7), we conclude

$$E[Y_t] \leq \Big(E[Y_0] + \int_{[0,t]} 2e^{-\beta s} E[(\delta\mu_s)^2] + e^{-2\beta s} 2E[(\delta\sigma_s)^2]ds\Big) e^{-(\beta-1-4K^2)t}.$$

□

Using this estimate, we now prove existence and uniqueness of the solution.

Proof of Theorem 16.1.2. Fix the initial condition $X_0 = x_0$. For a given predictable process X, consider the map F defined by

$$F(X)_t = x_0 + \int_{[0,t]} \mu_s(X_s)ds + \int_{[0,t]} \sigma_s(X_s)dW_s.$$

The process $F(X)$ then satisfies an SDE of the form (16.3), with μ and σ which do not depend on $F(X)$. From Lemma 16.1.6, we can see that, for any $X, \tilde{X}$ and any $\beta > 0$,

$$\begin{aligned}
&E[e^{-\beta t}(F(X)_t - F(\tilde{X})_t)^2] \\
&\leq 2e^{-(\beta-1)t} \int_{[0,t]} \Big(e^{-\beta s} E[(\mu_s(X_s) - \mu_s(\tilde{X}_s))^2] \\
&\qquad\qquad + e^{-2\beta s} E[(\sigma_s(X_s) - \sigma_s(\tilde{X}_s))^2]\Big)ds \\
&\leq 4e^{-(\beta-1)t} \int_{[0,t]} K^2 E[e^{-\beta s}(X_s - \tilde{X}_s)^2]ds.
\end{aligned}$$

Hence, by Fubini's theorem,

$$\begin{aligned}
&\int_{[0,\infty[} E[e^{-\beta t}(F(X)_t - F(\tilde{X})_t)^2]dt \\
&\leq \int_{[0,\infty[} 4e^{-(\beta-1)t} \int_{[0,t]} K^2 E[e^{-\beta s}(X_s - \tilde{X}_s)^2]ds\, dt \\
&\leq \int_{[0,\infty[} \frac{4K^2}{\beta-1} E[e^{-\beta s}(X_s - \tilde{X}_s)^2]ds.
\end{aligned}$$

Therefore, for $\beta > 4K^2 + 1$, F is a contraction on the space of predictable processes $X : [0,\infty[\to \mathbb{R}$, under the norm

$$\|X\|_\beta = \int_{[0,\infty[} E[e^{-\beta t} X_t^2]dt.$$

As this is simply a (weighted) L^2 norm, the space is complete. By the fixed point theorem for contractions (Lemma 1.5.18), we know that there is a unique process which satisfies (16.3), up to equality in this norm, that is, $X = F(X)$ $dt \times dP$-a.e. By continuity of the integrals, $F(X)$ is continuous, which implies the solution satisfies (16.3) and is unique, up to indistinguishability. □

Corollary 16.1.7. *Let F be as in the proof of Theorem 16.1.2, and let $\{X^{(n)}\}_{n\in\mathbb{N}}$ be the sequence of Picard iterations approximating this SDE, as defined by $X^{(n)} = F(X^{(n-1)})$ for some initial approximation $X^{(0)}$. Then the sequence $\{X^{(n)}\}_{n\in\mathbb{N}}$ converges ucp to the solution of the SDE.*

Proof. We know that $\int_{[0,\infty[} E[e^{-\beta t}(X_t^{(n)} - X_t^{(n-1)})^2]dt \to 0$, by construction. From an argument essentially the same as the proof of Lemma 16.1.4, we also see that we can bound $E[((X^{(n)} - X^{(n-1)})_T^*)^2]$ in terms of $\int_{[0,T]} E[(X_t^{(n)} - X_t^{(n-1)})^2]dt$. Therefore, we deduce that $e^{-\beta t}E[((X^{(n)} - X^{(n-1)})_t^*)^2] \to 0$ uniformly in t, which implies convergence ucp. □

One can also show that the convergence is in the semimartingale topology, but we leave this to the general setting (Theorem 16.5.1).

16.2 The S^p and $\mathcal{H}^p_{\mathcal{S}}$ Norms

In order to consider SDEs where the integral is with respect to a general semimartingale, and with weaker assumptions on the coefficients, we require a more careful choice of norm under which we prove convergence. The norms needed were developed by Protter [151], Meyer [135] and Émery [73, 74, 75], and we broadly follow their presentation. Working with these norms is not necessarily the most direct method to construct the solutions to SDEs, but they naturally arise when considering stability of solutions.

Recall that $\mathcal{D}$ denotes the space of (scalar) càdlàg adapted processes and $\mathcal{S}$ the space of (scalar) semimartingales. These spaces can be respectively given the topology of ucp convergence and Émery's semimartingale topology (Section 12.4). In order to prove convergence results, it is convenient to work with a somewhat stronger topology than ucp, which we now define. We shall focus on scalar semimartingales; however all our results can be adjusted to the vector case, see Remark 16.2.10.

Definition 16.2.1. *For $p \in [1, \infty]$, we denote by S^p the class of càdlàg processes X such that*

$$\|X\|_{S^p} := \|X^*_\infty\|_{L^p} < \infty$$

where $\|\cdot\|_{L^p}$ is the usual L^p norm on random variables (so $\|X\|_{S^p} = E[(X^*_\infty)^p]^{1/p}$ for $p < \infty$).

Note that $S^p \not\subset \mathcal{S}$ in general, that is, processes in S^p do not have to be semimartingales. (The 'S' in S^p refers to 'sup', rather than 'semimartingale', which is why we do not use the script $\mathcal{S}$.)

Remark 16.2.2. We need to work with optional or predictable processes (such as integrands in stochastic integrals). However, in general, these are not càdlàg. Nevertheless, recall that if X is an optional process, we defined $X^*_t := \lim_{s\downarrow t}\{\sup_{u<s}|X_u|\}$. This ensures that X^* is càdlàg, and so $\|X^*\|_{S^p}$ is well defined. For simplicity, we shall write $\|X\|_{S^p}$ in these cases also.

We defined, for $p \in [1, \infty[$, the space $\mathcal{H}^p$ of martingales M such that $E[(M^*_\infty)^{1/p}] < \infty$, with the corresponding convergence $M^{(n)} \to M$ in $\mathcal{H}^p$ if $E[((M^{(n)} - M)^*_\infty)^{1/p}] \to 0$. That is, $\mathcal{H}^p = S^p \cap \mathcal{M}$. As stated in Remark 11.5.9, the BDG inequality shows that this is equivalent to those martingales such that $E\big[[M]^{p/2}_\infty\big] < \infty$, which also leads to the definition of $\mathcal{H}^\infty$.

In order to give estimates on the behaviour of stochastic integrals, we define another norm on semimartingales, related to the $\mathcal{H}^p$ norm for local martingales in terms of the quadratic variation. We shall see (Theorem 16.2.6) that this defines the natural analogue of the semimartingale topology, when we replace ucp convergence with convergence in S^p.

Definition 16.2.3. *For $p \in [1, \infty]$, we denote by $\mathcal{H}^p_{\mathcal{S}}$ the space of semimartingales X such that*

$$\|X\|_{\mathcal{H}^p_{\mathcal{S}}} := \inf\left\{\left\|[M]^{1/2}_\infty + \int_{[0,\infty]} |dA|_s\right\|_{L^p}\right\} < \infty$$

where the infimum is taken over all decompositions of X into a local martingale $M \in \mathcal{M}_{0,\mathrm{loc}}$ and a finite variation process A with $A_0 = X_0 = \Delta A_0$.

Remark 16.2.4. If X is a local martingale, then, for any decomposition $X = M + A$, we have

$$[X]_\infty^{1/2} \le [M]_\infty^{1/2} + [A]_\infty^{1/2} \le [M]_\infty^{1/2} + \int_{[0,\infty]} |dA|_s.$$

Therefore, as $X = X + 0$ is a decomposition of X, we have $\|X\|_{\mathcal{H}^p_\mathcal{S}} = \|[X]_\infty^{1/2}\|_{L^p}$, which is (by the BDG inequality) equivalent to our earlier $\mathcal{H}^p$-norm for local martingales, for all $p \in [1, \infty[$. Therefore, the $\mathcal{H}^p_\mathcal{S}$ and S^p norms are equivalent for the local martingales, but this is not the case for semimartingales generally.

Lemma 16.2.5. *Any semimartingale in $\mathcal{H}^p_\mathcal{S}$ is a special semimartingale, and for $p < \infty$, the $\mathcal{H}^p_\mathcal{S}$ norm is equivalent to that given by*

$$X \mapsto \|[\tilde{M}]^{1/2}\|_{L^p} + \Big\| \int_{[0,\infty]} |d\tilde{A}_s| \Big\|_{L^p} = \|\tilde{M}\|_{\mathcal{H}^p_\mathcal{S}} + \|\tilde{A}\|_{\mathcal{H}^p_\mathcal{S}},$$

where $X = \tilde{M} + \tilde{A}$ is the canonical decomposition of X. Therefore $\mathcal{H}^p_\mathcal{S} = \mathcal{H}^p_0 \oplus \mathcal{A}^p_{\text{pred}}$, where $\mathcal{A}^p_{\text{pred}}$ denotes the predictable processes in $\mathcal{A}^p$, that is, with $E\big[\big(\int_{[0,\infty]} |dA|\big)^p\big] < \infty$.

Proof. Recall, from Definition 11.6.9, that a semimartingale is special if it has a decomposition $X = M + A$, where M is a local martingale, $M_0 = 0$ and A is locally of integrable variation. For a process $X \in \mathcal{H}^p_\mathcal{S}$, we see that there is a decomposition where A has total variation in L^p, so clearly X is special.

Considering the canonical decomposition $X = \tilde{M} + \tilde{A}$, we have

$$\|X\|_{\mathcal{H}^p_\mathcal{S}} \le \Big\| [\tilde{M}]_\infty^{1/2} + \int_{[0,\infty]} |d\tilde{A}|_s \Big\|_{L^p} \le \|[\tilde{M}]^{1/2}\|_{L^p} + \Big\| \int_{[0,\infty]} |d\tilde{A}_s| \Big\|_{L^p}.$$

For $X = M + A$ an arbitrary semimartingale decomposition with $A_0 = 0$, we know that $\tilde{A} = \Pi_p^* A$ and $\tilde{M} = M + A - \tilde{A}$. By Theorem 8.2.19, there exists a constant C_p independent of A such that

$$E\Big[\Big(\int_{[0,\infty]} |d\tilde{A}|_s\Big)^p\Big] \le C_p E\Big[\Big(\int_{[0,\infty]} |dA|_s\Big)^p\Big]$$

and, as in Remark 16.2.4,

$$[\tilde{M}]_\infty^{1/2} \le [M]_\infty^{1/2} + \int_{[0,\infty]} |d(A - \tilde{A})|_s \le [M]_\infty^{1/2} + \int_{[0,\infty]} |dA|_s + \int_{[0,\infty]} |d\tilde{A}|_s.$$

Therefore, taking the L^p norm and the infimum over all decompositions, there exists a $k > 0$ such that

$$\|[\tilde{M}]^{1/2}\|_{L^p} + \Big\| \int_{[0,\infty]} |d\tilde{A}_s| \Big\|_{L^p} \le k \|X\|^p_{\mathcal{H}^p_\mathcal{S}}$$

and we see the norms are equivalent. □

Using Lemma 16.2.5 and the BDG inequality, it is relatively straightforward to show that S^p and $\mathcal{H}^p_\mathcal{S}$ are Banach spaces under these norms, and we omit the proof.

We now see that $\mathcal{H}^p_\mathcal{S}$ and S^p are connected in the same way as the semimartingale and ucp 'norms' (see Definition 12.4.3).

Theorem 16.2.6. *For $p < \infty$, the $\mathcal{H}^p_\mathcal{S}$ norm is equivalent to the operator norm*

$$\sup\big\{\|H\bullet X\|_{S^p} : |H| \leq 1, H \text{ predictable}\big\}.$$

As a consequence, the $\mathcal{H}^p_\mathcal{S}$ norm is stronger than the S^p norm on the semimartingales, and there exists a constant c_p such that $\|X\|_{S^p} \leq c_p\|X\|_{\mathcal{H}^p_\mathcal{S}}$.

Proof. By Lemma 16.2.5, $\mathcal{H}^p_\mathcal{S} = \mathcal{H}^p_0 \oplus \mathcal{A}^p_{\text{pred}}$, so it is enough to prove the statement under the assumption that X is either a $\mathcal{H}^p_0$-martingale or a predictable process of finite variation (the equivalence of the norms then follows from Lemma 1.5.9).

First suppose X is a martingale. From the BDG inequality, for any $|H| \leq 1$ we have

$$\|H\bullet X\|_{S^p} = E\big[((H\bullet X)^*_\infty)^p\big]^{1/p} \leq C_p E\big[[X]^{p/2}_\infty\big]^{1/p} = C_p\|X\|_{\mathcal{H}^p_\mathcal{S}}.$$

Similarly, using the other side of the BDG inequality, for any $|H| = 1$,

$$\|X\|_{\mathcal{H}^p_\mathcal{S}} \leq c_p^{-1}\|H\bullet X\|_{S^p}.$$

Taking suprema over H yields the equivalence of the norms.

Now suppose X is a predictable process of finite variation. We know that

$$\sup_{|H|\leq 1}\{(H\bullet X)^*_\infty\} = \int_{[0,\infty]} \operatorname{sign}(dX)dX = \int_{[0,\infty]} |dX|$$

and the equivalence is immediate.

Therefore, by Lemmata 16.2.5 and 1.5.9, the $\mathcal{H}^p_\mathcal{S}$ and operator norms are equivalent. That the $\mathcal{H}^p_\mathcal{S}$ norm is stronger than the S^p norm, and the desired inequality, follows by taking $H = 1$. □

Extending Definition 16.0.1, we say that a process X is prelocally in $\mathcal{H}^p_\mathcal{S}$ if there exists a sequence of stopping times $T_n \to \infty$ such that $X^{T_n-} \in \mathcal{H}^p_\mathcal{S}$. We will say a sequence of processes $\{X^{(m)}\}_{m\in\mathbb{N}}$ converges prelocally to X in $\mathcal{H}^p_\mathcal{S}$ if there is a sequence of stopping times $T_n \to \infty$ such that $(X^{(m)})^{T_n-} \to X^{T_n-}$ in $\mathcal{H}^p_\mathcal{S}$ for all $n \in \mathbb{N}$. The following lemma gives a key connection between the various topologies.

Lemma 16.2.7. *(i) For all $p \in [1,\infty]$, any semimartingale is prelocally in $\mathcal{H}^p_{\mathcal{S}}$ and any process in $\mathcal{D}$ is prelocally in S^p*

(ii) If a sequence $X^{(m)}$ in $\mathcal{D}$ converges ucp, then for any $p \in [1,\infty[$ there is a subsequence which converges prelocally in S^p.

(iii) For any $p \in [1,\infty[$, if a sequence $X^{(m)}$ converges prelocally in S^p, then it converges in ucp.

(iv) The above also hold if $X^{(m)}$ is a sequence of semimartingales, and we replace 'convergence in S^p' with 'convergence in $\mathcal{H}^p_{\mathcal{S}}$' and 'convergence ucp' with 'convergence in the semimartingale topology'.

Proof. (i) For a process $X \in \mathcal{D}$, let $T_n = \inf\{t : |X_t| > n\}$. Then $\|X^{T_n-}\|_{S^p} \leq n$ and $T_n \to \infty$ as $n \to \infty$. Therefore X is prelocally in S^p.

For a process $X \in \mathcal{S}$, recall that $[M]_t^{1/2} + \int_{[0,t]} |dA|_s$ is a.s. finite, for any decomposition $X = M + A$ and any time $t < \infty$, and is a càdlàg process in t. Therefore, if we define

$$T_n := \inf\Big\{t : [M]_t^{1/2} + \int_{[0,t]} |dA|_s > n\Big\}$$

then $T_n \to \infty$, and we see that $\|X^{T_n-}\|_{\mathcal{H}^p_{\mathcal{S}}} \leq n$.

(ii) Suppose $\{X^{(m)}\}_{m\in\mathbb{N}}$ converges ucp to X. By Lemma 1.3.38, for each k, there exists a subsequence $\{X^{(m,k)}\}_{m\in\mathbb{N}}$ such that

$$P\big(\lim_{m\to\infty} (X^{(m,k)} - X)^*_k = 0\big) = 1.$$

By selecting these subsequences iteratively, we can assume $\{X^{(m,k+1)}\}_{m\in\mathbb{N}} \subseteq \{X^{(m,k)}\}_{m\in\mathbb{N}}$ for all k. Taking $\hat{X}^{(m)} := X^{(m,m)}$, we have a subsequence $\{\hat{X}^{(m)}\}_{m\in\mathbb{N}} \subseteq \{X^{(m)}\}_{m\in\mathbb{N}}$ such that

$$P\big(\lim_{m\to\infty} (\hat{X}^{(m)} - X)^*_t = 0\big) = 1$$

for all $t < \infty$. Hence, if $T_n = \inf\{t : \sup_{m\geq n} |\hat{X}^{(m)}_t - X_t| > n\} \wedge n$, we know that $P(T_n < t) \to 0$ for all $t > 0$, and as T_n is increasing, $T_n \to \infty$ a.s. We can also calculate, for any $\epsilon > 0$,

$$E\big[(\hat{X}^{(m)} - X)^*_{T_n-}\big] \leq nP\big((\hat{X}^{(m)} - X)^*_n > \epsilon\big) + \epsilon \to \epsilon$$

as $m \to \infty$. Therefore, $\hat{X}^{(m)} \to X$ prelocally in S^p.

(iii) Suppose $X^{(m)}$ converges prelocally in S^p to X. Then there is a sequence of stopping times $\{T_n\}_{n\in\mathbb{N}}$ such that $\lim_{n\to\infty} E[((X^{(m)} - X)^*_{T_n-})^p] = 0$. By Exercise 3.4.16, we know that for any $\epsilon > 0$ we can find n such that $P(T_n > t) > 1 - \epsilon/2$. Therefore, for any $t \geq 0$, we can find m sufficiently large that

$$\begin{aligned} P\big((X^{(m)} - X)^*_t > \epsilon\big) &\leq P\big((X^{(m)} - X)^*_{T_n-} > \epsilon\big) + P(T_n \leq t) \\ &\leq \frac{E\big[\big((X^{(m)} - X)^*_{T_n-}\big)^p\big]}{\epsilon^p} + \epsilon/2 \end{aligned}$$

and so, taking $m \to \infty$, we see that $X^{(m)} \to X$ in ucp.

(iv) Using Lemma 12.4.6 and a similar argument for $\mathcal{H}^p_{\mathcal{S}}$ (with Theorem 16.2.6), we can extend our results to $\mathcal{H}^p_{\mathcal{S}}$ and the semimartingale topology by applying (ii) and (iii) to $H^m \bullet X^{(m)}$ for an arbitrary sequence of predictable processes with $|H^m| \leq 1$. □

Remark 16.2.8. Mémin [132] gives an alternative approach to working with the semimartingale topology, based on measure changes rather than prelocalization. In this approach, one first shows that the semimartingale topology does not change when considering a probability measure equivalent to P. Furthermore, for any Cauchy sequence $\{X^n\}_{n\in\mathbb{N}}$ in $\mathcal{S}$, there exists a probability Q equivalent to P, with dQ/dP bounded, and a subsequence $\{Y^n\}_{n\in\mathbb{N}}$ which is locally a Cauchy sequence in $\mathcal{H}^1_{\mathcal{S}}(Q)$ (or more specifically, in $\mathcal{H}^2(Q) \oplus \mathcal{A}(Q)$, where $\mathcal{H}^1_{\mathcal{S}}(Q)$ denotes the $\mathcal{H}^1_{\mathcal{S}}$ topology under measure Q). Using this fact (which has a nontrivial proof), one can usually avoid prelocalization arguments.

We now obtain a simple, but surprisingly powerful, inequality for the norms of stochastic integrals.

Lemma 16.2.9 (Émery's Inequality). *For $X \in \mathcal{S}$, and H an X-integrable process, for any $p, q, r \in [1, \infty]$ with $p^{-1} + q^{-1} = r^{-1}$, we have*

$$\|H \bullet X\|_{\mathcal{H}^r_{\mathcal{S}}} \leq \|H\|_{S^p} \|X\|_{\mathcal{H}^q_{\mathcal{S}}}.$$

Proof. For any semimartingale decomposition $X = M + A$,

$$\begin{aligned}\Big\|(H^2 \bullet [M])^{1/2}_\infty + \int_{[0,\infty]} |H_s||dA|_s\Big\|_{L^r} &\leq \Big\|H^*_\infty\Big([M]^{1/2}_\infty + \int_{[0,\infty]} |dA|_s\Big)\Big\|_{L^r} \\ &\leq \|H^*_\infty\|_{L^p}\Big\|[M]^{1/2}_\infty + \int_{[0,\infty]} |dA|_s\Big\|_{L^q}\end{aligned}$$

where the last line is an application of Hölder's inequality. Therefore, whenever the right-hand side is finite, $H \bullet M$ and $H \bullet A$ are well defined, and $H \bullet X$ has a semimartingale decomposition $H \bullet X = H \bullet M + H \bullet A$. Consequently,

$$\begin{aligned}\|H \bullet X\|_{\mathcal{H}^r_{\mathcal{S}}} &\leq \Big\|[H \bullet M]^{1/2}_\infty + \int_{[0,\infty]} |d(H \bullet A)|_s\Big\|_{L^r} \\ &\leq \|H^*_\infty\|_{L^p}\Big\|[M]^{1/2}_\infty + \int_{[0,\infty]} |dA|_s\Big\|_{L^q},\end{aligned}$$

and taking an infimum over decompositions yields the result. □

Remark 16.2.10. The above presentation has been for scalar processes, but the same results hold for vector processes, with only minor adjustments. For the sake of concreteness, we shall say that if $X = (X^1, X^2, \dots, X^d)^\top$ is a vector process, then we write

$$\|X\|_{S^p} = \max_i \|X^i\|_{S^p} \qquad \text{and} \qquad \|X\|_{\mathcal{H}^p_{\mathcal{S}}} = \max_i \|X^i\|_{\mathcal{H}^p_{\mathcal{S}}},$$

whenever these are defined.

If X is a $d \times m$-matrix-valued process with components X^{ij}, then it is convenient to write

$$\|X\|_{S^p} = m \max_{i,j} \|X^{ij}\|_{S^p},$$

which ensures that

$$\|H \bullet X\|_{\mathcal{H}^r_{\mathcal{S}}} = \max_i \Big\| \sum_{j=1}^m (H^{ij} \bullet X^j) \Big\|_{\mathcal{H}^r_{\mathcal{S}}} \leq \max_{i,j} m \|H^{ij} \bullet X^j\|_{\mathcal{H}^r_{\mathcal{S}}} \leq \|H\|_{S^p} \|X\|_{\mathcal{H}^q_{\mathcal{S}}},$$

and so Émery's inequality can be used without variation. The inequality in Theorem 16.2.6 still holds, but the constant c_p will depend on m.

16.3 Existence and Uniqueness

We now return to considering solutions to SDEs. The key assumption under which we will obtain our results is the following version of Lipschitz continuity.

Recall that we are considering the SDE (16.2), which we repeat here for clarity:

$$X_t = H_t + \int_{[0,t]} f(\omega, u, X) dY_u \qquad \text{a.s. for all } t \geq 0$$

or equivalently, we can simply write $X = H + f(X) \bullet Y$. Here H is a prescribed càdlàg adapted process taking values in $\mathbb{R}^d$, Y is a semimartingale taking values in $\mathbb{R}^m$, and $f : \Omega \times [0, \infty[\times \mathcal{D} \to \mathbb{R}^{d \times m}$. Leaving the dimension implicit throughout this section, we write $H \in \mathcal{D}$ and $Y \in \mathcal{S}$.

Remark 16.3.1. For notational simplicity, when X is a vector we write X^* for the scalar process $\max_i\{(X^i)^*\}$, and similarly in the matrix case.

Definition 16.3.2. *We say that a coefficient f is* stochastically Lipschitz *continuous if there exists a sequence of stopping times $\{T_n\}_{n \in \mathbb{N}}$ with $T_n \to \infty$, and an increasing measurable process K such that, for all n, for any $X, \tilde{X} \in \mathcal{D}$,*

$$(f(X) - f(\tilde{X}))^*_{t \wedge T_n -} \leq K_t(\omega)(X - \tilde{X})^*_{t \wedge T_n -}$$

Here all inequalities are to be read as 'up to an evanescent set'.

We say that a coefficient f is uniformly Lipschitz *when there exists a constant $K > 0$ such that, for any $X, \tilde{X} \in \mathcal{D}$,*

$$(f(X) - f(\tilde{X}))^*_t \leq K(X - \tilde{X})^*_{t-}.$$

We write $f \in \mathrm{Lip}(K)$ in this case.

Remark 16.3.3. As alluded to previously, if f is stochastically Lipschitz, then it is clear that f must only depend on the realized path $\{X_t(\omega)\}_{t \geq 0}$, and not on 'statistical' quantities of the random process X, such as $E[X_{t-}]$.

Remark 16.3.4. In a vector setting, for any $f \in \mathrm{Lip}(K/m)$, where m is the dimension of Y, it is easy to see that

$$\|f(X) - f(X')\|_{S^p} = m\|(f(X) - f(X'))^*\|_{S^p} \leq K\|X - X'\|_{S^p}$$

for any $p \leq \infty$, which is the key property which we shall use in the following.

Remark 16.3.5. As all norms on finite dimensional vector spaces are equivalent (as follows from repeated application of Lemma 1.5.9), we could equivalently define Lipschitz functions using the Euclidean norm on $\mathbb{R}^d$ (or any other norm), instead of maximizing over the components. In this case, we would replace X^* with $(\sum_i((X^i)^*)^2)^{1/2}$ in the definitions, and similarly for the matrix terms. The choice above is convenient for our purposes, as it agrees with our earlier extension of the S^p and $\mathcal{H}^p_{\mathcal{S}}$ norms (Remark 16.2.10).

We now see that, if we establish an existence and uniqueness result for uniformly Lipschitz coefficients, then the same holds whenever f is stochastically Lipschitz. Recall that we write $\mathcal{S}$ for the space of semimartingales.

Lemma 16.3.6. *Suppose that we can show that (16.2) admits a unique solution in $\mathcal{D}$ (resp. $\mathcal{S}$) whenever f is uniformly Lipschitz. Then (16.2) admits a unique solution in $\mathcal{D}$ (resp. $\mathcal{S}$) under the assumption f is stochastically Lipschitz.*

Proof. Suppose f is stochastically Lipschitz. Define

$$R_n = T_n \wedge \inf\left\{t : \operatorname{ess\,sup}_{Y,\tilde{Y}\in\mathcal{D}}\left\{\Big(\frac{(f(Y) - f(\tilde{Y}))^*_t}{(Y - \tilde{Y})^*_{t-}}\Big)^*\right\} > n\right\}.$$

As f is the coefficient of an SDE, it is easy to see that the process

$$\left\{\operatorname{ess\,sup}_{Y,\tilde{Y}\in\mathcal{D}}\left\{\Big(\frac{(f(Y) - f(\tilde{Y}))^*_t}{(Y - \tilde{Y})^*_{t-}}\Big)^*\right\}\right\}_{t\geq 0}$$

is adapted, càdlàg and nondecreasing (and hence optional), and so R_n is a stopping time (but may be zero with positive probability). Taking $n \to \infty$, as $K_t(\omega)$ is assumed to be a.s. finite in the definition of a stochastically Lipschitz coefficient, we see that $R_n \to \infty$ a.s.

Now note that $I_{[\![0,R_n[\![}f \in \mathrm{Lip}(n)$, so by assumption we have a unique solution $X^{(n)}$ in $\mathcal{D}$ (resp. $\mathcal{S}$) to the SDE with coefficient $I_{[\![0,R_n[\![}f$. By uniqueness, for $n' > n$ the solutions $X^{(n)}$ and $X^{(n')}$ must agree on the interval $[\![0, R_n[\![$, and so pasting yields a single process X such that $X = X^{(n)}$ on $[\![0, R_n[\![$ for all n. Therefore, a solution to the SDE with coefficient f exists.

Given any other solution $\tilde{X}$, uniqueness implies that $\tilde{X} = X = X^{(n)}$ on $[\![0, R_n[\![$ for all n. Therefore, $\tilde{X} = X$ a.s. for all t, and the solution is unique. □

The next lemma provides the fundamental existence result in this setting, assuming Y is 'small' in the $\mathcal{H}^\infty_\mathcal{S}$ norm. We use this to build solutions to SDEs.

Lemma 16.3.7. *Suppose $p < \infty$, Y is an m-dimensional semimartingale with $\|Y\|_{\mathcal{H}^\infty_\mathcal{S}} \leq \frac{1}{2c_pK}$, $H \in S^p$, $f \in \mathrm{Lip}(K/m)$ and $f(0) \in S^p$, where c_p is the constant in Theorem 16.2.6. Then the SDE $X = H + f(X)\bullet Y$ admits a unique solution in S^p, and this solution satisfies*

$$\|X\|_{S^p} \leq 2\|H + f(0)\bullet Y\|_{S^p} \leq 2\|H\|_{S^p} + K^{-1}\|f(0)\|_{S^p},$$

where 0 denotes the zero process.

Proof. We show that the map $X \mapsto F(X) := H + f(X)\bullet Y$ is a contraction in S^p. We have

$$F(X) - F(X') = (f(X) - f(X'))\bullet Y$$

so by Émery's inequality with $r = p$ and $q = \infty$,

$$\begin{aligned}\|F(X) - F(X')\|_{S^p} \leq c_p\|F(X) - F(X')\|_{\mathcal{H}^p_\infty} &\leq \|f(X) - f(X')\|_{S^p}\|Y\|_{\mathcal{H}^\infty_\mathcal{S}} \\ &\leq Kc_p\|X - X'\|_{S^p}\|Y\|_{\mathcal{H}^\infty_\mathcal{S}} \leq \frac{1}{2}\|X - X'\|_{S^p}.\end{aligned}$$

Therefore, we know that F is a contraction in S^p, so we have the existence of a unique solution in S^p (Lemma 1.5.18). The norm $\|X\|_{S^p} \leq 2\|F(0)\|_{S^p}$ follows, and applying Émery's inequality again gives the final expression. □

In the simple setting of Section 16.1, we were able to create a family of norms where we exponentially shrank the norm at time t, at a rate β. We then showed that we had a contraction whenever β was sufficiently large. In the general setting, we instead partition time into small pieces, and then use Lemma 16.3.7 to prove that our SDE has a solution on each piece sequentially.

Definition 16.3.8. *We say that a semimartingale $X \in \mathcal{H}^\infty_\mathcal{S}$ is α-sliceable, and write $X \in \mathrm{Sl}(\alpha)$, if there exists a finite sequence of stopping times $0 = T_1 \leq T_1 \leq \ldots \leq T_k \leq \infty$, such that $X = X^{T_k-}$ and, for all n,*

$$\|(X - X^{T_n})^{T_{n+1}-}\|_{\mathcal{H}^\infty_\mathcal{S}} \leq \alpha.$$

*The sequence $\{T_n\}_{n=1}^k$ is said to α-*slice *X.*

Note that the jumps of a process $X \in \mathrm{Sl}(\alpha)$ must be bounded (as $X \in \mathcal{H}^\infty_\mathcal{S}$) but they do not need to be small (as they may occur at one of the stopping times T_k).

Theorem 16.3.9. *Let X be any semimartingale.*

(i) If X is vector-valued with components $X^i \in \mathrm{Sl}(\alpha)$, then $X \in \mathrm{Sl}(\alpha)$.

(ii) If $X \in \mathrm{Sl}(\alpha)$, then for any stopping time T, $X^T \in \mathrm{Sl}(\alpha)$ and $X^{T-} \in \mathrm{Sl}(2\alpha)$.

(iii) For any $\alpha > 0$, there exists a sequence of stopping times $T_n \to \infty$ such that $X^{T_n -} \in \mathrm{Sl}(\alpha)$ for all n.

Proof. (i) For each component we have a sequence $\{T_n^i\}_{n=1}^{k_i}$. Reordering these (more particularly, applying the result of Exercise 6.5.10), we obtain a single sequence $\{T_n\}_{n=1}^{\sum_i k_i}$ which α-slices X^i for all i. From the definition of $\mathcal{H}_{\mathcal{S}}^\infty$ for a vector process (Remark 16.2.10), we see that $X \in \mathrm{Sl}(\alpha)$.

(ii) That $X^T \in \mathrm{Sl}(\alpha)$ follows easily from the inequality $\|X^T\|_{\mathcal{H}_{\mathcal{S}}^\infty} \leq \|X\|_{\mathcal{H}_{\mathcal{S}}^\infty}$. If we have a semimartingale decomposition $X = M + A$, then $X^{T-} = M^T + (A^{T-} - \Delta M_T I_{[\![T,\infty[\![})$ is a semimartingale decomposition of X^{T-}. Therefore, we can verify that $\|X^{T-}\|_{\mathcal{H}_{\mathcal{S}}^\infty} \leq 2\|X\|_{\mathcal{H}_{\mathcal{S}}^\infty}$. As T was arbitrary, this implies $X^{T-} \in \mathrm{Sl}(2\alpha)$.

(iii) By (i), it is sufficient to consider the scalar case. Fix $\beta > 0$. Using Theorem 10.3.3, we decompose X into a local martingale M with jumps bounded in absolute value by β and a finite variation process A.

First consider A. We define the sequence $\{S_n\}_{n\in\mathbb{N}}$ by $S_1 = 0$ and

$$S_{n+1} = \inf\Big\{t \geq S_n : \int_{]S_n,t]} |dA|_s \geq \beta\Big\}.$$

Then, for any n, $A^{S_n -}$ is in $\mathrm{Sl}(\beta)$.

Next, consider M. We define $R_1 = 0$, and

$$R_{n+1} = \inf\Big\{t \geq R_n : [M]_t - [M]_{R_n} \geq \beta^2\Big\}.$$

Writing

$$(M - M^{R_n})^{R_{n+1}-} = M^{R_{n+1}} - M^{R_n} - \Delta M_{R_{n+1}} I_{\{R_{n+1} > R_n\}} I_{[\![R_{n+1},\infty[\![},$$

as the jumps of M are bounded in absolute value by β, we have

$$\begin{aligned}\|(M - M^{R_n})^{R_{n+1}-}\|_{\mathcal{H}_{\mathcal{S}}^\infty} &\leq \big\|([M]_{R_{n+1}} - [M]_{R_n})^{1/2} + |\Delta M_{R_{n+1}}|\big\|_{L^\infty} \\ &= \big\|((\Delta M_{R_{n+1}})^2 + [M]_{R_{n+1}-} - [M]_{R_n})^{1/2} + |\Delta M_{R_{n+1}}|\big\|_{L^\infty} \\ &\leq (\beta^2 + \beta^2)^{1/2} + \beta,\end{aligned}$$

and it follows that $M^{R_n -} \in \mathrm{Sl}((\sqrt{2}+1)\beta)$ for all n.

Finally, we observe that if X and X' are two semimartingales in $\mathrm{Sl}(\beta)$, with corresponding β-slicing sequences $\{S_n\}_{n=1}^k$ and $\{R_n\}_{n=1}^{k'}$, then (ii) implies that $X + X'$ is in $\mathrm{Sl}(8\beta)$, with corresponding sequence defined by reordering $\{S_n\}_{n=1}^k$ and $\{R_n\}_{n=1}^{k'}$.

Combining these results, we have

$$X^{T_n -} = M^{T_n -} + A^{T_n -} \in \mathrm{Sl}(8(\sqrt{2}+1)\beta).$$

As β was arbitrary, we conclude that there exists a sequence $\{T_n\}_{n\in\mathbb{N}}$ such that $X^{T_n -} \in \mathrm{Sl}(\alpha)$ for all α. □

By slicing, we can extend Lemma 16.3.7 to a wider setting.

Lemma 16.3.10. *Let $H \in S^p$, $Y \in \mathrm{Sl}((2c_pK)^{-1})$ be of dimension m and $f \in \mathrm{Lip}(K/m)$. Then the SDE $X = H + f(X)\bullet Y$ admits a unique solution in S^p, and this solution satisfies*

$$\|X\|_{S^p} \le \gamma(\|H\|_{S^p} + \|f(0)\|_{S^p}),$$

where 0 denotes the zero process, and γ is a constant depending only on K and Y.

Proof. We write $y = \|Y\|_{\mathcal{H}^\infty_S}$, $h = 2\|H\|_{S^p} + (y + K^{-1})\|f(0)\|_{S^p}$ and $\alpha = (2c_pK)^{-1}$. Let $\{T_n\}_{n=1}^k$ be an α-slicing sequence for Y. We consider solving our SDE on the sequence of intervals $[\![0, T_i[\![$, $[\![0, T_i]\!]$, $[\![0, T_{i+1}[\![$, ..., that is, we consider the equations

$$X^{T_i-} = H^{T_i-} + f(X)\bullet Y^{T_i-} \tag{16.5}$$
$$X^{T_i} = H^{T_i} + f(X)\bullet Y^{T_i} \tag{16.6}$$

for increasing values of i. To begin, with $i = 1$, as $T_i = 0$ we have a unique solution to (16.5), namely $X = 0$, using our convention that $X_{0-} = H_{0-} = 0$.

Step 1: Suppose we have a unique solution $X^{(i-)}$ to (16.5) for a given value of i. This must agree with any solution to (16.6) on the interval $[\![0, T_i[\![$. As f is non-anticipative, we see that the unique solution $X^{(i-)}$ to (16.6) is then

$$X^{(i)} := I_{[\![0,T_i[\![}X^{(i-)} + I_{[\![T_i]\!]}\big(X^{(i-)}_{T_i-} + H_{T_i} - H_{T_i-} + f_{T_i}(X^{(i-)})\Delta Y_{T_i}\big).$$

The S^p norm of $X^{(i)}$ is bounded by

$$\begin{aligned}\|X^{(i)}\|_{S^p} &\le \|X^{(i-)}\|_{S^p} + \|H_{T_i} - H_{T_i-}\|_{L^p} + \|f_{T_i}(X^{(i-)})\Delta Y_{T_i}\|_{L^p} \\ &\le \|X^{(i-)}\|_{S^p} + 2\|H\|_{S^p} + Ky\|X^{(i-)}\|_{S^p} + y\|f(0)\|_{S^p} \\ &\le (1 + Ky)\|X^{(i-)}\|_{S^p} + h.\end{aligned}$$

Step 2: Suppose we have a unique solution $X^{(i-1)}$ to (16.6) with i replaced by $i - 1$. Then any solution to (16.5) must agree with $X^{(i-1)}$ on the interval $[\![0, T_{i-1}]\!]$. Rewriting (16.5), we have

$$X = \underbrace{X^{(i-1)} + H^{T_i-} - H^{T_{i-1}}}_{\in S^p} + f(X)\bullet \underbrace{(Y - Y^{T_{i-1}})^{T_i-}}_{\in \mathcal{H}^\infty_S}. \tag{16.7}$$

As $\{T_n\}_{n=1}^k$ α-slices Y, we know that $\|(Y - Y^{T_{i-1}})^{T_i-}\|_{\mathcal{H}^\infty_S} \le (2c_pK)^{-1}$. Therefore, applying Lemma 16.3.7, we know that (16.7) admits a unique solution $X^{(i-)}$, and that its norm in S^p is bounded by

$$\|X^{(i-)}\|_{S^p} \le 2\|X^{(i-1)} + H^{T_i-} - H^{T_{i-1}}\|_{S^p} + K^{-1}\|f(0)\|_{S^p} \le 2\|X^{(i-1)}\|_{S^p} + 2h.$$

Combining these two steps up to the final slicing time T_k, we obtain a solution $X^{(k-)}$ on the interval $[\![0, T_k[\![$ and, as $Y^{T_k-} = Y$, we have the global solution

$$X = X^{(k-)} + H - H^{T_k-}.$$

Combining our earlier estimates, the norm of $X^{(i-)}$ is bounded by the recursion

$$\|X^{(i-)}\|_{S^p} \leq 2(1+Ky)\|X^{(i-)}\|_{S^p} + 4h$$

and so, as $X^{(1-)} = 0$, we have the bound

$$\|X\|_{S^p} \leq \|X^{(k-)}\|_{S^p} + \|H - H_{T_k-}\|_{S^p} \leq 2h\frac{2^k(1+Ky)^k - 1}{1+Ky} + h.$$

The stated inequality then follows, with $\gamma = (2+y+K^{-1})(\frac{2^{k+1}(1+Ky)^k-2}{1+Ky}+1)$. $\square$

We can now state the overall existence result.

Theorem 16.3.11. *Let H be in $\mathcal{D}$, Y a semimartingale and $f \in \mathrm{Lip}(K)$ for some $K > 0$. Then the equation $X = H + f(X) \bullet Y$ admits a unique solution in $\mathcal{D}$.*

Proof. By Lemma 16.2.7(i) and Theorem 16.3.9(iii), there is a sequence of stopping times $\{T_n\}_{n\in\mathbb{N}}$ such that $T_n \to \infty$ a.s. and, for every n, H^{T_n-} and $(f(0))^{T_n-}$ are in S^p and $Y^{T_n-} \in \mathrm{Sl}\big((2c_pK)^{-1}\big)$. By Lemma 16.3.10, we have the existence of a unique solution in S^p, on the interval $[\![0, T_n[\![$, for each n. By pasting these together, we have the existence of a solution in $\mathcal{D}$.

To show uniqueness, suppose we have two solutions X and X' in $\mathcal{D}$. Then, by Lemma 16.2.7(i), we can find a sequence of stopping times $R_n \to \infty$ such that X^{R_n-} and $\tilde{X}^{R_n-}$ are bounded and, therefore, are in S^p. However, together with our earlier uniqueness result, this implies that $X^{T_n\wedge R_n-} = \tilde{X}^{T_n\wedge R_n-}$ for all n, so $X = \tilde{X}$. $\square$

Remark 16.3.12. We have given results for SDEs where the integral is with respect to a semimartingale. One can also consider SDEs where integrals are taken with respect to random measures, and with respect to both random measures and semimartingales. The proof of existence and uniqueness is similar to above; however, it is notationally much more cumbersome. A classical description of this theory can be found in Jacod [107, Chapter XIV], and a simple case which often appears in applications is considered in the coming chapter.

16.4 Stability of Solutions

In addition to knowing that solutions to SDEs exist and are unique, we also are usually interested in their stability, that is, knowing that a small perturbation in the parameters (in our setting, in H, f and Y) has only a small impact on the solution X. We shall do this by first showing that we have stability in S^p (or $\mathcal{H}^p_{\mathcal{S}}$), under appropriate conditions, and then using this to show that we generally have stability in ucp (or in the semimartingale topology).

Lemma 16.4.1. *Consider an SDE with $H \in S^p$ (resp. $\mathcal{H}^p_{\mathcal{S}}$), $f \in \mathrm{Lip}(K/m)$ and $Y \in \mathrm{Sl}\big((2Kc_p)^{-1}\big) \subset \mathcal{H}^\infty_{\mathcal{S}}$ an m-dimensional semimartingale. Let X be the unique solution to $X = H + f(X)\bullet Y$ in S^p.*

Take sequences $\{H^n\}_{n\in\mathbb{N}}$, $\{f^n\}_{n\in\mathbb{N}}$ and $\{Y^n\}_{n\in\mathbb{N}}$ such that:

- *$H^n \to H$ in S^p (resp. $\mathcal{H}^p_{\mathcal{S}}$),*
- – *$f^n \in \mathrm{Lip}(K/m)$,*
  – *there is $c > 0$ such that $(f^n(Z))^* \le c$ for all n and all $Z \in \mathcal{D}$,*
  – *$(f^n(X) - f(X))^* \to 0$ in S^p and*
- *$Y^n \to Y$ in $\mathcal{H}^p_{\mathcal{S}}$.*

Let X^n be the solution to the equation $X^n = H^n + f^n(X^n)\bullet Y^n$. Then $X^n \to X$ in S^p (resp. $\mathcal{H}^p_{\mathcal{S}}$). Furthermore, $X \in \mathcal{H}^p_{\mathcal{S}}$ whenever $H \in \mathcal{H}^p_{\mathcal{S}}$.

Proof. Define, for $n \in \mathbb{N}$,

$$\begin{aligned} J^n &= (f(X) - f^n(X))\bullet Y + f^n(X^n)\bullet(Y - Y^n), \\ g^n(\cdot) &= f^n\big(X\big) - f^n\big(X - (\cdot)\big). \end{aligned}$$

The semimartingales J^n are in $\mathcal{H}^p_{\mathcal{S}}$ and, from Émery's inequality and the boundedness of $f^n(X^n)$,

$$\|J^n\|_{\mathcal{H}^p_{\mathcal{S}}} \le \big\|\big(f(X) - f^n(X)\big)^*\big\|_{S^p}\|Y\|_{\mathcal{H}^\infty_{\mathcal{S}}} + \|f^n(X^n)\|_{S^\infty}\|Y - Y^n\|_{\mathcal{H}^p_{\mathcal{S}}} \to 0.$$

We also note that $g^n \in \mathrm{Lip}(K)$ and $g^n(0) = 0$ for all n.

We can write

$$\begin{aligned} X - X^n &= H - H^n + (f(X) - f^n(X))\bullet Y + f^n(X^n)\bullet(Y - Y^n) \\ &\qquad + (f^n(X) - f^n(X^n))\bullet Y \\ &= H - H^n + J^n + g^n(X - X^n)\bullet Y, \end{aligned}$$

and so, from Lemma 16.3.10, there is a constant γ independent of n such that

$$\|X - X^n\|_{S^p} \le \gamma\|H - H^n + J^n\|_{S^p} \le \gamma(\|H - H^n\|_{S^p} + \|J^n\|_{S^p}) \to 0.$$

This establishes the desired result when $H^n \to H$ in S^p.

If H and all H^n are in $\mathcal{H}^p_{\mathcal{S}}$, then it is clear that X and all X^n are semimartingales. As $f(X)$ and $f^n(X^n)$ are bounded, Émery's inequality immediately shows that X and all X^n are in $\mathcal{H}^p_{\mathcal{S}}$. If $H^n \to H$ in $\mathcal{H}^p_{\mathcal{S}}$, then our above representation of $X - X^n$ yields

$$\|X - X^n\|_{\mathcal{H}^p_{\mathcal{S}}} \le \|H - H^n\|_{\mathcal{H}^p_{\mathcal{S}}} + \|J^n\|_{\mathcal{H}^p_{\mathcal{S}}} + K\|X - X^n\|_{S^p}\|Y\|_{\mathcal{H}^\infty_{\mathcal{S}}}$$

and as we know $X^n \to X$ in S^p, we have $X^n \to X$ in $\mathcal{H}^p_{\mathcal{S}}$. □

Remark 16.4.2. If $Y^n = Y$ for all n, then a quick inspection of the proof shows that the above result holds without the assumption that $(f^n(Z))^* \leq c$ for all n and all $Z \in \mathcal{D}$.

Just as in the proof of Theorem 16.3.11, we can use a prelocalization argument to extend this to a more general setting.

Theorem 16.4.3. *Let Y be an m-dimensional semimartingale, H be a càdlàg adapted process (resp. a semimartingale) and $f \in \mathrm{Lip}(K)$. Let X be the unique solution to $X = H + f(X) \bullet Y$, and assume that $f(X)$ is prelocally bounded. Consider sequences $\{H^n\}_{n\in\mathbb{N}}$, $\{f^n\}_{n\in\mathbb{N}}$ and $\{Y^n\}_{n\in\mathbb{N}}$ such that:*

- *$H^n \to H$ in ucp (resp. in the semimartingale topology);*
- *$f^n \in \mathrm{Lip}(K)$ and $(f^n(X) - f(X))^* \to 0$ in ucp;*
- *$Y^n \to Y$ in the semimartingale topology.*

Let X^n be the solution to the equation $X^n = H^n + f^n(X^n) \bullet Y^n$. Then $X^n \to X$ in ucp (resp. in the semimartingale topology).

Proof. We present the case of convergence in ucp. The semimartingale case is identical, after corresponding changes of terminology. We also assume, for notational convenience, that our processes are scalar. The vector case follows with only minor changes.

By Lemma 12.4.8 (and Remark 12.4.9), it is enough to show that the convergence holds (pre)locally. By Lemma 16.2.7 and Theorem 16.3.9, by choosing a subsequence and prelocalizing, we can assume, for a given $p \geq 1$ that $H^n \to H$ in S^p, $Y^n \to Y$ in $\mathcal{H}^p_{\mathcal{S}}$, $Y \in \mathrm{Sl}((2Kc_p)^{-1})$ and $f(X)$ is bounded by a constant c.

The only condition lacking from a direct application of Lemma 16.4.1 is that we need $f^n(X^n)$ to be bounded. For notational convenience, we will write $B^k(x) := (-k) \vee x \wedge k$. We consider the new equation

$$Z^n = H^n + B^{K+c+1}(f^n(Z^n)) \bullet Y^n.$$

As $|f(X)| < c$, we know $f(X) = B^{K+c+1}(f(X))$. By Lemma 16.4.1, it follows that $Z^n \to X$ in ucp. Taking a subsequence, we can assume that $(Z^n - X)^*_t$ and $(f^n(X) - f(X))^*_t$ converge to zero almost surely, for all $t < \infty$. We now define

$$T_k = \inf\{t : |Z^n_t - X_t| + |f^n_t(X) - f_t(X)| \geq 1 \text{ for some } n \geq k\},$$

and note that the stopping times $T_n \uparrow \infty$. Prelocalizing with the sequence $\{T_k\}_{k\in\mathbb{N}}$, and ensuring that $n \geq k$, we can assume that $|Z^n - X|$ and $|f^n(X) - f(X)|$ are bounded by 1.

We can then write,

$$\begin{aligned}
|f^n(Z^n)| &\leq |f^n(Z^n) - f^n(X)| + |f^n(X) - f(X)| + |f(X)| \\
&\leq K|Z^n - X| + |f^n(X) - f(X)| + |f(X)| \leq K + 1 + c.
\end{aligned}$$

However, this implies that $B^{K+c+1}(f^n(Z^n)) = f^n(Z^n)$, and so (recalling that we have chosen a subsequence) we have $Z^n = X^n$ on the interval $[\![0, T_n[\![$. Therefore, we see that $X^n \to X$ prelocally in S^p, and so by Lemma 16.2.7, $X^n \to X$ in ucp.

Our only remaining concern is that we chose a subsequence in order to make our argument work, so we have only shown that a subsequence converges to X in ucp. However, applying this result to an arbitrary subsequence of X^n, we see that we always have a sub-subsequence which converges ucp, and always to the same limit X. Therefore, there exists no subsequence $\{X^{n'}\}_{n' \in \mathbb{N}}$ with $\|X^{n'} - X\|_{\mathrm{ucp}} > \epsilon$ for all n', that is, $X^n \to X$ in ucp. □

16.5 Approximation Schemes

As with ordinary differential equations, the above results do not provide any guidance as to how to construct an explicit solution to a concrete problem and, in general, closed form solutions are not available (see however Section 16.6). It is, therefore, important to have methods of approximating the solution to an SDE. We here give two classical methods for doing this.

Theorem 16.5.1 (Picard Iteration). *Let $H \in \mathcal{D}$, $f \in \mathrm{Lip}(K/m)$ for some constant $K > 0$ and Y be an m-dimensional semimartingale. Consider the equation $X = H + f(X) \bullet Y$. For any $X^0 \in \mathcal{D}$, the sequence X^n defined by*

$$X^{n+1} := H + f(X^n) \bullet Y$$

satisfies $(X^n - X) \to 0$ in the semimartingale topology (and hence ucp). (Note that we do not assume that X^n or X is a semimartingale, but their difference is always a semimartingale.)

Proof. By prelocalization, we can assume that X^0 is bounded and, for a fixed $\alpha > 0$, we have $Y \in \mathrm{Sl}(\alpha)$ with slicing sequence $\{T_i\}_{i=1}^k$. By Lemmata 12.4.8 and 16.2.7 (and Remark 12.4.9), it is enough to show that this prelocalized sequence converges in $\mathcal{H}^p_{\mathcal{S}}$ for some (arbitrary) $p \in [1, \infty[$.

Let c_p be the constant in Theorem 16.2.6. Assume, without loss of generality, that $c_p \geq 3$ and, for simplicity, we will simply write c for c_p. Take $\alpha = (c^3 K)^{-1}$ and write $\beta = \|Y\|_{\mathcal{H}^\infty_{\mathcal{S}}} \vee c\alpha$. The processes $V^n := X - X^n$ satisfy the equation

$$V^n = (X^{n+1} - X^n) + g^n(V^n) \bullet Y,$$

where $g^n(\cdot) := f\big((\cdot) + X^n\big) - f(X^n) \in \mathrm{Lip}(K/m)$. Then Lemma 16.3.10 gives

$$\|V^n\|_{S^p} \leq \gamma \|X^{n+1} - X^n\|_{S^p}$$

for a constant γ independent of n. It is therefore sufficient to show that $\|X^{n+1} - X^n\|_{S^p} \to 0$, as this will imply

$$\begin{aligned}\|V^{n+1}\|_{\mathcal{H}^p_{\mathcal{S}}} &= \|(f(X^n) - f(X))\bullet Y\|_{\mathcal{H}^p_{\mathcal{S}}} \\ &\le K\|X^n - X\|_{S^p}\|Y\|_{\mathcal{H}^\infty_{\mathcal{S}}} = K\|V^n\|_{S^p}\|Y\|_{\mathcal{H}^\infty_{\mathcal{S}}} \\ &\le K\gamma\|X^{n+1} - X^n\|_{S^p}\|Y\|_{\mathcal{H}^\infty_{\mathcal{S}}} \\ &\to 0.\end{aligned} \tag{16.8}$$

Let $Z^n := X^{n+1} - X^n$, and write $z^n_i = \|(Z^n)^{T_i-}\|_{S^p}$. Then we have

$$z^{n+1}_{i+1} \le z^{n+1}_i + \|\Delta Z^{n+1}_{T_i}\|_{L^p} + \|(Z^{n+1} - (Z^{n+1})^{T_i})^{T_{i+1}-}\|_{S^p}.$$

At the same time, Z^n satisfies the recurrence relation $Z^{n+1} = g^n(Z^n)\bullet Y$, so $\Delta Z^{n+1}_{T_i} = g^n_{T_i}(Z^n)\Delta Y_{T_i}$ and

$$(Z^{n+1} - (Z^{n+1})^{T_i})^{T_{i+1}-} = g^n(Z^n)\bullet(Y - Y^{T_i})^{T_{i+1}-}.$$

As $g^n \in \mathrm{Lip}(K/m)$ and $g^n(0) = 0$, we see that

$$z^{n+1}_{i+1} \le z^{n+1}_i + K\beta z^n_i + cK\alpha z^n_{i+1}. \tag{16.9}$$

It now remains to show that $z^n_k = \|(Z^n)^{T_k-}\|_{S^p} \to 0$ as $n \to \infty$, for all k. By (16.8), this will imply $V^{n+1} \to 0$ prelocally in $\mathcal{H}^p_{\mathcal{S}}$, and so in the semimartingale topology (Lemma 16.2.7). Write $a = K\beta$ and $b = cK\alpha = c^{-2} \le a$. Then define

$$v^n_i := c^{n+2i}a^i b^{n-i}.$$

As $c \ge 3$, the doubly indexed sequence $\{v^i_n\}_{i,n\in\mathbb{N}}$ satisfies the recurrence relation

$$v^{n+1}_i + av^n_i + bv^n_{i+1} = (cb + a + c^2a)v^n_i < c^3av^n_i = v^{n+1}_{i+1}. \tag{16.10}$$

This is analogous to the recurrence (16.9) satisfied by z^n_i, but the inequality is in the other direction. Now let r be a constant such that for all $i \le k$, we have $z^0_i \le rv^0_i$. As $z^n_0 = 0$, we also have $z^n_0 \le rv^n_0$ for all n. Now suppose the relation $z^n_i \le rv^n_i$ holds for the pairs (n, i), $(n, i + 1)$ and $(n + 1, i)$. Then combining (16.9) and (16.10) we have

$$z^{n+1}_{i+1} \le z^{n+1}_i + az^n_i + b^n_{i+1} \le r(v^{n+1}_i + av^n_i + bv^n_{i+1}) \le rv^{n+1}_{i+1},$$

and so the relation holds for the pair $(n + 1, i + 1)$. Therefore, the relation holds for any pair (n, i) with $i \le k$, in particular

$$z^n_k \le rc^{n+2k}(K\beta)^k(c^{-2})^{n-k} = r(c^4K\beta)^k c^{-n}.$$

It follows that for any k, we have $z^n_k \to 0$ as $n \to \infty$, and the theorem is proven. □

Theorem 16.5.2 (Euler–Maruyama Approximation). *For each $n \in \mathbb{N}$, let $\pi_n = \{0 = T_1^n, T_2^n, \ldots, T_n^n\}$ be a nondecreasing sequence of stopping times (which may be deterministic). Assume that*

$$\|\pi_n\| = \sup_i\{|T_{i+1}^n - T_i^n|\} \to 0 \text{ and } T_n^n \to \infty \text{ a.s. as } n \to \infty.$$

Let $H \in \mathcal{D}$ (resp. $H \in \mathcal{S}$), $Y \in \mathcal{S}$ and $f \in \mathrm{Lip}(K)$ for some constant $K > 0$. Assume that $f(Z)$ is left-continuous, for any $Z \in \mathcal{D}$ (resp. $Z \in \mathcal{S}$). For any process G we define the left- and right-continuous approximations

$$\pi_n^-(G) = I_{[\![0]\!]}G_0 + \sum_{i=1}^n G_{T_i} I_{]\!]T_i, T_{i+1}]\!]},$$
$$\pi_n(G) = \sum_{i=1}^n G_{T_i} I_{[\![T_i, T_{i+1}[\![}.$$

The stochastic Euler–Maruyama approximation *is the solution of the equation*

$$X^n = H^{T_n} + f^n(X^n)\bullet Y^{T_n}$$

where

$$f^n(Z) = \pi_n^-(f(\pi_n(Z))).$$

Then $X^n \to X = H + f(X)\bullet Y$ in ucp (resp. in the semimartingale topology).

Proof. We consider the scalar case for simplicity. By Lemma 12.4.8, we know $H^{T_n} \to H$ in ucp (resp. in the semimartingale topology) and $Y^{T_n} \to Y$ in the semimartingale topology. By Theorem 16.4.3, it remains to show that $(f^n(X) - f(X)) \to 0$ in ucp.

As f and f^n are left-continuous, we know that, a.s. for all t,

$$\big(f^n(X) - f(\pi_n(X))\big)_t^* = \big(\pi_n^-(f(\pi_n(Z))) - (f(\pi_n(Z)))\big)_t^* \to 0.$$

As f is Lipschitz and X is càdlàg, we have, a.s. for all t,

$$\big(f(\pi_n(X)) - f(X)\big)_t^* \le K\big(\pi_n(X) - X\big)_{t-}^* \to 0.$$

Therefore, $f_t^n(X) \to f_t(X)$ a.s., uniformly in t. The result follows. □

Remark 16.5.3. In the above, when f is a function only of X_{t-} at time t, it easy to see that X^n can be written

$$X_{T_i^n}^n = H_{T_i^n} + \sum_{j=1}^{i-1} f_{T_j^n}^n(X_{T_j^n}^n)(Y_{T_{j+1}^n} - Y_{T_j^n})$$

which is a more common representation of the approximation.

Remark 16.5.4. For the purposes of numerical approximation, we are often concerned with the rate of convergence of the approximation. This depends heavily on the choice of norm, so ucp and semimartingale convergence are often not of primary interest. For the Euler–Maruyama scheme, in the Markovian case (which we consider in the coming chapter), for X scalar, H constant, $Y_t = (t, W_t)^\top$ for W a Brownian motion, and the partition $\pi_n = \{k/n\}_{k=1}^{n^2}$, one can show that for any deterministic T,

$$E\big[|X_T^n - X_T|\big] \leq O(1/\sqrt{n})$$

and, for any Lipschitz function $g : \mathbb{R} \to \mathbb{R}$,

$$\big|E[g(X_T^n)] - E[g(X_T)]\big| \leq O(1/n).$$

See Glasserman [90, Chapter 6] or Kloeden and Platen [118] for further details and extensions.

16.6 Linear Equations

As mentioned before, the above arguments do not generally give any guidance on how to find explicit solutions to concrete problems. In the linear case, the following result (which is an extension of a result due to Yor and Yoeurp, see [155, p.378]) gives an explicit solution. For simplicity, we write X_- for the process $\{X_{t-}\}_{t\geq 0}$, and assume by convention that $Y_{0-} = H_{0-} = 0$, so $Y_0 = \Delta Y_0$ and $(H \bullet Y)_0 = H_0 Y_0$.

Theorem 16.6.1. *Consider the scalar linear equation $X = H + X_- \bullet Y$.*

(i) For $H \in \mathcal{D}$ and $Y \in \mathcal{S}$ with $\Delta Y \neq -1$, we have the solution

$$X = H - \mathcal{E}(Y)(H_- \bullet (\mathcal{E}(Y)^{-1})).$$

(ii) For H, Y semimartingales such that $\Delta H \Delta Y = 0$, $\Delta Y \neq -1$, we can also write

$$X = \mathcal{E}(Y)\big((\mathcal{E}(Y)^{-1})_- \bullet (H - [H, Y])\big)$$

In particular, if $[H, Y] = 0$, then

$$X = \mathcal{E}(Y)\big((\mathcal{E}(Y)^{-1})_- \bullet H\big)$$

As the equation is Lipschitz, our solution is unique, and so the representations in (i) and (ii) must agree.

Proof. The proof is similar in both cases – we expand $X_- \bullet Y$, where X is the proposed solution, use the product rule, and show that appropriate terms cancel. We shall prove (i). In the setting of (ii), the only additional step is to notice that, as $\Delta H \Delta Y = 0$, we know $\big[Y, [H, Y]\big] = 0$.

To show (i), using the product rule and the fact $\mathcal{E}(Y)_{s-}dY_s = d(\mathcal{E}(Y))_s$ we have

$$\begin{aligned}(X_-\bullet Y)_t &= (H_-\bullet Y)_t - \int_{[0,t]} \mathcal{E}(Y)_{s-}(H_-\bullet(\mathcal{E}(Y)^{-1}))_{s-}dY_s \\ &= (H_-\bullet Y)_t - \mathcal{E}(Y)_t\big(H\bullet(\mathcal{E}(Y)^{-1})\big)_t \\ &\quad + \big((H_-\mathcal{E}(Y)_-)\bullet(\mathcal{E}(Y)^{-1})\big)_t + [\mathcal{E}(Y), H_-\bullet(\mathcal{E}(Y)^{-1})]_t.\end{aligned}$$

From Lemma 15.1.5, we know that

$$d(\mathcal{E}(Y)^{-1})_t = \mathcal{E}(Y)_{t-}^{-1}\Big(- dY_t + d\langle Y^c\rangle_t + \frac{(\Delta Y_t)^2}{1+\Delta Y_t}\Big),$$

and so, expanding and cancelling $\mathcal{E}(Y)_{s-}\mathcal{E}(Y)_{s-}^{-1}$ terms,

$$\begin{aligned}&(X_-\bullet Y)_t \\ &= (H_-\bullet Y)_t + X_t - H_t \\ &\quad + \int_{[0,t]} H_{s-}\big(- dY_s + d\langle Y^c\rangle_s\big) + \sum_{0\le s\le t} H_{s-}\frac{(\Delta Y_s)^2}{1+\Delta Y_s} \\ &\quad + \int_{[0,t]} H_{s-}(-d\langle Y^c\rangle_s) + \sum_{0\le s\le t} H_{s-}\Delta Y_s\Big(- \Delta Y_s + \frac{(\Delta Y_s)^2}{1+\Delta Y_s}\Big) \\ &= X_t - H_t.\end{aligned}$$

□

Remark 16.6.2. The requirement that $\Delta Y \neq -1$ is simply to ensure $\mathcal{E}(Y)^{-1}$ is well defined. If it is possible that $\Delta Y_t = -1$ for some t, then we can find an explicit solution to our equation by 'restarting' our solution every time $\Delta Y_t = -1$, as at these times T our solution satisfies $X_T = \Delta H_T$. As this happens only finitely many times on any compact set, these restarted solutions can be pasted together to give an explicit (if not particularly elegant) solution to the SDE.

Remark 16.6.3. A multivariate form of this result is possible. However, as matrix multiplication is not commutative, this requires the definition of the left and right stochastic matrix exponentials (that is, solutions to the equations $X_t = I_d + \int_{[0,t]} X_{t-}dY_t$ and $X_t = I_d + \int_{[0,t]}(dY_t\, X_{t-})$ for Y a $\mathbb{R}^{d\times d}$-valued-semimartingale, where I_d is the identity matrix). See [155, Chapter IX Exercise 2.6] for details in the continuous case.

Example 16.6.4. A famous special case of this type of SDE is the (one-dimensional) Langevin equation

$$dX_t = dW_t - \beta X_t dt$$

for W a Brownian motion, $\beta \in \mathbb{R}$ and $X_0 = W_0 = 0$. Using these results, we see that this has solution

$$X_t = W_t - \beta \int_{[0,t]} e^{-\beta(t-s)} W_s ds = \int_{[0,t]} e^{-\beta(t-s)} dW_s.$$

16.7 Explosion Times

We now consider what happens when we relax the Lipschitz assumption on f. We have already seen (Lemma 16.3.6) that we can assume that f is stochastically Lipschitz, which corresponds to assuming that f is 'locally' Lipschitz, where 'locally' is interpreted in terms of ω and t (through the use of stopping times). We will now allow f to be locally Lipschitz in X, and discuss the corresponding existence results.

Example 16.7.1. For Y a continuous semimartingale with $Y_0 = 1$, the equation

$$X_t = -\int_{[0,t]} e^{-2X_s} d\langle Y\rangle_s + \int_{[0,t]} e^{-X_s} dY_s$$

has solution $X_t = \log(Y_t)$ up to the time $T = \inf\{t : Y_t = 0\}$, which may be finite or infinite.

Example 16.7.2. For Y a continuous semimartingale with $Y_0 = 0$, the equation

$$X_t = 1 + \int_{[0,t]} 3X_s^2 d\langle Y\rangle_s - \int_{[0,t]} 2X_s^{3/2} dY_s$$

has solution $X_t = (1+Y_t)^{-2}$ up to the time $T = \inf\{t : Y_t = -1\}$, which may be finite or infinite.

Note that in these examples, our coefficients are locally Lipschitz (i.e. the Lipschitz condition is satisfied for bounded values of X) and we have a solution up to the first time that $X_t = \pm\infty$.

Definition 16.7.3. *We say a function $f : \Omega \times [0, \infty[\times \mathcal{D} \to \mathbb{R}^{d\times m}$ is locally Lipschitz if there exists some a.s. finite $\mathcal{F} \otimes \mathcal{B}(\mathbb{R})$-measurable function $K : \Omega \times \mathbb{R} \to \mathbb{R}$ and some sequence of stopping times $\{T_n\}_{n\in\mathbb{N}}$ with $T_n \uparrow \infty$ such that, for all processes $X, \tilde{X} \in \mathcal{D}$ and all $\alpha > 0$, we have*

$$(f(X)-f(\tilde{X}))^* \leq K(\omega,\alpha)(X-\tilde{X})^* \quad \text{on} \quad [\![0, T_n]\!] \cap \{(\omega,t) : X_t^* \vee \tilde{X}_t^* \vee t \leq \alpha\}.$$

Definition 16.7.4. *For an optional process X, the random variable $T = \inf\{t : X_t^* = \infty\}$ is called the* (first) explosion time *of X.*

Lemma 16.7.5. *The explosion time of a real càdlàg adapted process is a predictable stopping time.*

Proof. Let T be the explosion time. Take $T_n = \inf\{t : X_t^* \geq n\} \wedge n$. As X_{T_n} is real, we have $T_n < T$ and $T_n \uparrow T$, so $\{T_n\}_{n\in\mathbb{N}}$ announces T. Therefore, T is predictable. □

Theorem 16.7.6. *Let $H \in \mathcal{D}$, $Y \in \mathcal{S}$ and f be locally Lipschitz. Consider the SDE $X = H + f(X)\bullet Y$. Then*

(i) all solutions to the SDE have the same explosion time T and
(ii) there is a unique càdlàg adapted solution X to the SDE on the interval $[\![0,T[\![$, that is, to the equation

$$X = H^{T-} + f(X)\bullet Y^{T-}.$$

Proof. Let X and $\tilde{X}$ be two solutions, with explosion times T and $\tilde{T}$. Set $T_n = \inf\{t : X_t^* \vee \tilde{X}_t^* \geq n\} \wedge n$. As f is locally Lipschitz, on the set $[\![0,T_n]\!]$, from Theorem 16.3.11 and Lemma 16.3.6 we know $X = \tilde{X}$. However, this implies that T_n is the announcing sequence for the explosion times T and $\tilde{T}$, as constructed in Lemma 16.7.5, so $T = \tilde{T}$ and $X = \tilde{X}$ on $[\![0,T[\![$.

To show a solution exists, we define a truncation $B^n(X) = (-n) \vee X \wedge n$, which is applied componentwise when X is a vector. Then $f(B^n(\cdot))$ is stochastically Lipschitz, so the equation

$$X = H + f(B^n(X))\bullet Y$$

admits a solution $X^{(n)}$ and this solution is unique for each n. Defining $T_n = \inf\{t : (X^{(n)})^* \geq n\}$, we see that there is a càdlàg solution to

$$X = H + f(X)\bullet Y$$

on the interval $[\![0,T_n[\![$. By uniqueness, $X^{(n+1)} = X^{(n)}$ on the interval $[\![0,T_n[\![$, so $T_{n+1} \geq T_n$. By pasting, we can construct a solution to $X = H + f(X)\bullet Y$ on the set $\cup_n[\![0,T_n[\![= [\![0,T[\![$. □

Remark 16.7.7. By considering again Example 16.0.11, we see that we cannot in general remove the locally Lipschitz assumption without losing uniqueness of the solution.

Theorem 16.7.8. *Consider the scalar SDE*

$$X_t = X_0 + \int_{[0,t]} \mu(\omega, s, X_{s-})ds + \int_{[0,t]} \sigma(\omega, s, X_{s-})dW_s,$$

where W is a Brownian motion, μ and σ are locally Lipschitz and satisfy

$$\int_{[0,T]} E\big[|\mu_s(0)|^2 + |\sigma_s(0)|^2\big]ds < \infty$$

and the linear growth condition, for some $K \in \mathbb{R}$,

$$|\mu_s(x)| \leq |\mu_s(0)| + K|x|, \qquad |\sigma_s(x)| \leq |\sigma_s(0)| + K|x|$$

Then the SDE admits a unique solution for all time.

Proof. We know, from Theorem 16.7.6, that a solution exists up to the first explosion time of X. Let $T_n := \inf\{t : |X_t| \geq n\}$, so a solution certainly exists up to time T_n. Lemma 16.1.4 states that, for any deterministic time t, there exists C (independent of n) such that

$$E[(X_t^{T_n})^2] \leq C \int_{[0,t]} E\big[|\mu_s(0)|^2 + |\sigma_s(0)|^2\big] ds < \infty.$$

By Markov's inequality, $P(T_n \leq t) = P(|X_t^{T_n}| \geq n) \to 0$ as $n \to \infty$. In other words, $T_n \to \infty$ as $n \to \infty$ (a.s. for a subsequence) and we see that the explosion time of X is almost surely infinite. □

Remark 16.7.9. As noted in Remark 16.1.5, the above argument works equally well in multiple dimensions, and when μ and σ are nonanticipative functionals of the path of X.

16.8 Exercises

Exercise 16.8.1. Verify that, in each of the examples in this chapter, the stated solutions hold. Also determine in which settings we can prove uniqueness.

Exercise 16.8.2. Prove Theorem 16.6.1(ii), that is, for H, Y semimartingales such that $\Delta H \Delta Y = 0$, $\Delta Y \neq -1$, if $X = H + X_- \bullet Y$, we can write

$$X = \mathcal{E}(Y)\big((\mathcal{E}(Y)^{-1})_- \bullet (H - [H, Y])\big).$$

Exercise 16.8.3. Give sufficient conditions on a predictable nonanticipative function f such that the SDE

$$X = H + f(X) \bullet Y + \langle X \rangle$$

admits a unique continuous solution, where Y is a continuous semimartingale and H a continuous predictable process. Show that there are always multiple discontinuous solutions to this equation.

Exercise 16.8.4. Consider the generalized Ornstein–Uhlenbeck equation with jumps, defined by

$$dX_t = -\kappa(X_t - \mu_t)dt + dW_t + d\tilde{N}_t; \qquad X_0 = x_0,$$

where W is a Brownian motion, $\tilde{N}$ is a compensated Poisson process, $\kappa > 0$ is a constant and μ is an integrable progressive process. Using the results of Section 16.6, give two explicit formulae for X in terms of κ, W, $\tilde{N}$ and μ. Hence or otherwise, show that if μ is a constant, we have

$$T^{-1} \int_{[0,T]} X_t dt \to \mu \qquad \text{in probability as } T \to \infty.$$

Exercise 16.8.5. Consider the SDE

$$dX_t = \Big(-\frac{5}{2}X_t + X_t^3 e^{4t}\Big) dt + X_t \tan(W_t) dW_t$$

with initial value $X_0 = 1$, where W is a Brownian motion. Show that this equation has a unique solution up to the stopping time $\inf\{t : |W_t| = \frac{\pi}{2}\}$, and give an expression for the solution of the form $X_t = f(W_t)g(t)$.

Exercise 16.8.6. Show that the scalar SDE

$$X_t = 1 + \int_{[0,t]} f\big(s, X_s, (X \bullet Y)_s\big) dY_s,$$

where Y is a semimartingale and f is a uniformly Lipschitz function, admits a unique solution for all time. In the case where Y is a Brownian motion and $f(s, x, z) = ze^{-s}$, show that $E[X_t^2] \leq e^{t/2}$.

Exercise 16.8.7. Suppose X^1 and X^2 are solutions to the SDEs,

$$dX_t^i = f^i(X_t^i)dt + g(X_t^i)dW_t$$

where W is a Brownian motion and f^1, f^2 and g are Lipschitz coefficients of SDEs. Suppose $f^1 \geq f^2$ and $X_0^1 \geq X_0^2$. By applying Itô's lemma to $((X^2 - X^1)^+)^3$ and Grönwall's inequality, show that $X^1 \geq X^2$ up to evanescence.

17

Markov Properties of SDEs

In the previous chapter, we have considered SDEs where the integral is with respect to a general semimartingale. In this chapter, we focus our attention on a much more specialized setting, where the integral is taken with respect to time (i.e. Lebesgue measure), a Brownian motion and a compensated Poisson random measure. Working in this setting allows the Markovian properties of the Brownian motion and the Poisson process to be inherited by the SDE solution. A full treatment of this topic would require consideration of general Markov processes. For this, see Ethier and Kurtz [77], or the more specialized treatments in Karatzas and Shreve [117] or Revuz and Yor [155]. We shall instead present only a selection of these issues.

Suppose we have a filtered probability space, in which we have

- a sequence of $N \leq \infty$ independent Brownian motions $\{W^1, W^2, \ldots\}$,
- a random measure $\mu \in \tilde{\mathcal{A}}^1_\sigma$ on $\mathcal{Z} \times [0, \infty]$ for some Blackwell space $(\mathcal{Z}, \mathfrak{Z})$, with deterministic (hence predictable) compensator $\mu_p(d\zeta, dt) = \nu(d\zeta)dt$. We write $\tilde{\mu} = \mu - \mu_p$,

We allow for $N = \infty$ for the sake of generality, as this does not make our calculations any more complex[1]. Whether N is finite or infinite, for a fixed finite m, we write $\mathbb{R}^{m\times N}$ for the space of $\mathbb{R}^m$-valued sequences of length N, arranged as a matrix, so $\mathbb{R}^N = \mathbb{R}^{N\times 1}$, and $\|z\|^2 = \sum_i \|z^i\|^2$ for $z = (z^1, z^2, \ldots, z^N) \in \mathbb{R}^{m\times N}$.

[1]When thinking about these equations, the key issues are typically present for the simple case $N = 1$. Conceptually, this is also a little easier to work with, so the reader should feel free to assume this for the sake of simplicity. However, the general case of $N \in \mathbb{N} \cup \{\infty\}$, $m \in \mathbb{N}$ follows with no notational changes. For $N = \infty$, we identify $\mathbb{R}^N$ with ℓ_2, so $\mathcal{B}(\mathbb{R}^N)$ refers to the Borel topology inherited from the ℓ_2 metric.

S.N. Cohen, R.J. Elliott, *Stochastic Calculus and Applications*,
Probability and Its Applications, DOI 10.1007/978-1-4939-2867-5_17

Writing W for the vector $(W^1, W^2, ...)^\top \in \mathbb{R}^N$, for a predictable process Z taking values in $\mathbb{R}^{m\times N}$ we have the $\mathbb{R}^m$-valued continuous local martingale

$$(Z^\top \bullet W)_t = \int_{[0,t]} Z_s^\top dW_s = \sum_{i=1}^N \int_{[0,t]} Z_s^i dW_s^i.$$

As the components of the vector W are uncorrelated, there is no difference between the vector integral of Section 12.5 and the sum of the integrals component by component. For $N = \infty$, provided $[Z^\top \bullet W] = \int_{[0,\cdot]} \|Z_s\|^2 ds$ is locally integrable, this sum is well defined[2] as a limit in $\mathcal{H}^2_{\mathrm{loc}}$.

If ν is the measure on $\mathcal{Z}$ such that $\mu_p(dz, dt) = \nu(dz)dt$, then we write $L^2(\nu)$ for the space of (equivalence classes of) functions $g : \mathcal{Z} \to \mathbb{R}^m$ such that $\int_{\mathcal{Z}} \|g(\zeta)\|^2 \nu(d\zeta) < \infty$. This space has norm

$$\|g\|_\nu = \Big(\int_{\mathcal{Z}} \|g(\zeta)\|^2 \nu(d\zeta)\Big)^{1/2}. \tag{17.1}$$

We do not assume that $\nu(\mathcal{Z}) < \infty$, so our random measure may have infinitely many jumps on any interval $]s,t]$.

In this chapter, we consider SDEs of the following type.

$$\begin{cases} dX_t = f(t, X_t)dt + \sigma(t, X_t)dW_t + \displaystyle\int_{\zeta\in\mathcal{Z}} g(\zeta, t, X_{t-})\tilde{\mu}(d\zeta, dt), \\ X_s = x \in \mathbb{R}^d \end{cases} \tag{17.2}$$

where $f : [0,\infty[\times\mathbb{R}^d \to \mathbb{R}^d$, $\sigma : [0,\infty[\times\mathbb{R}^d \to \mathbb{R}^{d\times N}$ and $g : \mathcal{Z}\times[0,\infty[\times\mathbb{R}^d \to \mathbb{R}^d$ are Borel measurable in the appropriate product σ-algebras, and satisfy the growth bound

$$\|f(t,x)\|^2 + \|\sigma(t,x)\|^2 + \|g(\cdot, t, x)\|_\nu^2 \leq K(1 + \|x\|^2), \tag{17.3}$$

for some $K \in \mathbb{R}$. As usual, the boundedness conditions stated above can be weakened, by localizing prior to any explosion time.

Remark 17.0.1. A classic example of our random measure is that associated with a Lévy process, where $\mathcal{Z} = \mathbb{R}^N$, for $N < \infty$. In this case, one often encounters a bound of the form $\|g(\zeta, s, x)\|^2 \leq K(\|\zeta\|^2 \wedge 1)(1 + \|x\|^2)$, which implies $\|g(\cdot, s, x)\|_\nu^2 \leq K(1 + \|x\|^2)$, as we know $\int_{\mathbb{R}^N} (1 \wedge \|\zeta\|^2)\nu(d\zeta) < \infty$ (Theorem 13.5.6).

Remark 17.0.2. We could readily extend our setting to allow the compensator μ_p to depend more fully on time, that is, by taking $\mu_p(d\zeta)dt = \nu(t, d\zeta)dt$, where $\nu(t,\cdot)$ is a measure on $(\mathcal{Z}, \mathfrak{Z})$ for each t, and $\nu(\cdot, A)$ is measurable in time, for each $A \in \mathfrak{Z}$. The only change that this will require is notational, as the norm $\|\cdot\|_\nu$ will then need to depend on time.

[2] Throughout this chapter, we simply write $\mathcal{H}^2$ for the space of $\mathbb{R}^m$-valued processes with components in $\mathcal{H}^2$, and similarly for $\mathcal{H}^2_{\mathrm{loc}}$.

Note that, as $\mu \in \tilde{\mathcal{A}}^1_\sigma$, our process has jumps given by

$$\Delta X_t = \int_{\zeta \in \mathcal{Z}} g(\zeta, t, X_{t-})\tilde{\mu}(d\zeta, \{t\}) = g(z_t, t, X_{t-}) I_{\{\Delta X_t \neq 0\}},$$

where z_t is the (unique) value such that $\mu(\{(z_t, t)\}) = 1$, if one exists, and is an arbitrary value otherwise.

The key idea in this chapter is that in (17.2) we are 'starting' our equation at time s in state x, with the aim of studying the relationship between the starting values on the behaviour of the solution. For notational clarity, we write $X^{(s,x)}$ for the solution process started at x at time s, and either $X^{(s,x)}_t$ or $X^{(s,x)}(\omega, t)$ for its value at time t.

It will be useful to define the filtration $\{\mathcal{F}_t\}_{t \geq 0}$ to be the (completed, right-continuous) filtration generated by W and $\tilde{\mu}$, possibly augmented to include an independent random variable x_0 at time zero. Furthermore, we define, for any $s \leq t$, the σ-algebra $\mathcal{F}_{s,t}$ to be that generated by $\{W_u - W_s\}_{u \in]s,t]}$ and $\{\tilde{\mu}(A) : A \subset]s,t] \times \mathbb{R}^d\}$. That is, $\mathcal{F}_{s,t}$ will contain the information that can be obtained by observing $W - W_s$ and $\tilde{\mu}$ during the interval $]s,t]$. Due to the independence of increments of W and $\tilde{\mu}$, we know that for any $0 \leq s \leq t$, the σ-algebras $\mathcal{F}_{s,t}$ and $\mathcal{F}_s$ are independent, in the sense of Definition 2.1.10.

Remark 17.0.3. The equation (17.2) does not describe the most general class of Markov processes which we might wish to study. In particular, we could add a term of the form $\int_{\mathcal{Z}} k(\zeta, t, X_t)\mu(d\zeta, dt)$ to the dynamics of X, assuming that $k(\cdot, t, x)$ is $|d\mu|$-integrable. However, if we suppose that $[X]^{1/2}$ is locally integrable (equivalently, that X is a special semimartingale), then one can show that this adds no generality to the form stated in (17.2). For this reason, we will give our attention to the slightly more restrictive form in (17.2).

17.1 Dependence on Initial Data

Lemma 17.1.1. *Suppose f, σ and g satisfy the Lipschitz conditions*

$$\begin{aligned} \|f(t,x) - f(t,x')\|^2 &\leq K\|x - x'\|^2, \\ \|\sigma(t,x) - \sigma(t,x')\|^2 &\leq K\|x - x'\|^2, \\ \|g(\cdot,t,x) - g(\cdot,t,x')\|^2_\nu &\leq K\|x - x'\|^2, \end{aligned} \tag{17.4}$$

for some $K \in \mathbb{R}$, as well as the bounds (17.3). Then equation (17.2) admits a unique solution for each s and x, denoted $X^{(s,x)} = \{X^{(s,x)}_t\}_{t \in [s,T]}$, and this solution is in S^2. For a given s and x, the process $X^{(s,x)}$ is independent of $\mathcal{F}_s$.

If X and $\tilde{X}$ are solutions to SDEs with parameters f, σ, g and $\tilde{f}, \tilde{\sigma}, \tilde{g}$, then there is a constant C, which depends only on K, ν and T, such that for $x, y \in \mathbb{R}^d$ and $s \in [0,T]$,

$$
\begin{aligned}
&E\Big[\sup_{s\leq t\leq T}\|X_t^{(s,x)}-\tilde{X}_t^{(s,y)}\|^2\Big]\\
&\leq C\Big(\|x-y\|^2+\int_{[0,T]}E\big[\|f(t,X_t^{(s,x)})-\tilde{f}(t,X_t^{(s,x)})\|^2\\
&\qquad\qquad\qquad\qquad+\|\sigma(t,X_t^{(s,x)})-\tilde{\sigma}(t,X_t^{(s,y)})\|^2\\
&\qquad\qquad\qquad\qquad+\|g(\cdot,t,X_t^{(s,x)})-\tilde{g}(\cdot,t,X_t^{(s,x)})\|_\nu^2\big]dt\Big).
\end{aligned}
$$

Furthermore, for $p\geq 2$, if

$$
E\Big[\sup_t\Big(\int_{\mathcal{Z}}\|g(\zeta,t,X_t^{(s,x)})\|\mu(d\zeta,\{t\})\Big)^p\Big]<\infty,
$$

(for example, if g is bounded) then the solution is in S^p.

Proof. As $\{W_u-W_s\}_{u\geq s}$ and $\{\tilde{\mu}(]s,u]\times A)\}_{u\geq s,A\in\mathfrak{Z}}$ are independent of $\mathcal{F}_s$ and as x is deterministic, we can perform a time change and use the filtration $\{\tilde{\mathcal{F}}_t\}_{t\geq 0}=\{\mathcal{F}_{s,(s+t)}\}_{t\geq 0}$. A solution adapted to $\{\tilde{\mathcal{F}}_t\}_{t\geq 0}$ will then be independent of $\mathcal{F}_s$, as required, and we can proceed on the assumption that $s=0$.

In the absence of the random measure term, (17.2) would be a multivariate version of the equation considered in Section 16.1, and the existence of a solution and the stated inequality would follow from an (easy) multivariate extension of Lemma 16.1.4 (applied to the difference $X^{(s,x)}-X^{(s,y)}$), Lemma 16.1.6 and Theorem 16.1.2.

Given the presence of the random measure term, the main difficulty is establishing a version of Lemma 16.1.4. Using the BDG and Jensen inequalities, we have, for some constant C depending on T, K and p, which may vary from line to line,

$$
\begin{aligned}
&E[(\|X^{(s,x)}-\tilde{X}^{(s,y)}\|_t^*)^p]\\
&\leq CE\big[\|x-y\|^p\big]+C\int_{[0,T]}E\big[\|f_t(X_t^{(s,x)})-\tilde{f}_t(\tilde{X}_t^{(s,y)})\|^p\big]ds\\
&\quad+CE\Big[\Big\|\Big(\int_{[0,\cdot]}(\sigma_t(X_t^{(s,x)})-\tilde{\sigma}_t(\tilde{X}_t^{(s,y)}))dW_t\Big)_T^*\Big\|^p\Big]\\
&\quad+CE\Big[\Big\|\Big(\int_{\mathcal{Z}\times[0,\cdot]}g_t(\zeta,X_s^{(s,x)})-\tilde{g}_t(\zeta,\tilde{X}_t^{(s,y)})\tilde{\mu}(d\zeta,dt)\Big)_T^*\Big\|^p\Big]\\
&\leq CE\big[\|x-y\|^p\big]+C\int_{[0,T]}E\big[\|f_t(X_t^{(s,x)})-\tilde{f}_t(\tilde{X}_t^{(s,y)})\|^p\big]ds\\
&\quad+CE\Big[\Big(\int_{[0,T]}\|\sigma_t(X_t^{(s,x)})-\tilde{\sigma}_t(\tilde{X}_t^{(s,y)})\|^2dt\Big)^{p/2}\Big]\\
&\quad+CE\Big[\Big(\int_{\mathcal{Z}\times[0,T]}\|g_t(\zeta,X_s^{(s,x)})-\tilde{g}_t(\zeta,\tilde{X}_t^{(s,y)})\|^2\mu(d\zeta,dt)\Big)^{p/2}\Big].
\end{aligned}
$$

We now separate the cases $p = 2$ and $p > 2$. In the case $p = 2$, we observe that, as the expectations of optional and predictable quadratic variations agree (Lemma 11.3.4), we have

$$\begin{aligned}
&E\Big[\Big(\int_{\mathcal{Z}\times[0,T]} \|g_t(\zeta, X_s^{(s,x)}) - \tilde{g}_t(\zeta, \tilde{X}_t^{(s,y)})\|^2 \mu(d\zeta, dt)\Big)\Big] \\
&= E\Big[\Big(\int_{[0,T]} \|g_t(\zeta, X_s^{(s,x)}) - \tilde{g}_t(\zeta, \tilde{X}_t^{(s,y)})\|_\nu^2 dt\Big)\Big].
\end{aligned}$$

We can then write, for some constant C,

$$\|f_t(X_t^{(s,x)}) - \tilde{f}_t(\tilde{X}_t^{(s,y)})\|^2 \leq C\big(\|f_t(X_t^{(s,x)}) - \tilde{f}_t(X_t^{(s,x)})\|^2 + \|X_t^{(s,x)} - \tilde{X}_t^{(s,y)}\|^2\big),$$

and similarly for the σ and g terms. Applying Grönwall's inequality, as in the proof of Lemma 16.1.4, yields the desired inequality of the theorem. In the same way, one can verify that the proofs of Lemma 16.1.6 and Theorem 16.1.2 continue to hold, with the addition of the relevant $\tilde{\mu}$ terms, and so a solution in S^2 exists. The details are left to the reader.

If $p > 2$, then we assume $\tilde{f}$, $\tilde{\sigma}$, $\tilde{g}$ and y are all zero, so $\tilde{X}^{(s,y)} = 0$. As the jump of $g * \mu$ at time t is $\int_{\mathcal{Z}} g(\zeta)\mu(d\zeta, \{t\})$, we then apply Theorem 8.2.20 (in particular, see Remark 11.5.8) to obtain, for some C,

$$\begin{aligned}
&E\Big[\Big(\int_{\mathcal{Z}\times[0,T]} \|g_t(\zeta, X_s^{(s,x)})\|^2 \mu(d\zeta, dt)\Big)^{p/2}\Big] \\
&= CE\Big[\Big(\int_{[0,T]} \|g_t(\zeta, X_s^{(s,x)})\|_\nu^2 dt\Big)^{p/2} + \sup_t \Big(\int_{\mathcal{Z}} \|g(\zeta, t, X_t^{(s,x)})\|\mu(d\zeta, \{t\})\Big)^p\Big].
\end{aligned}$$

We have assumed the last term is integrable, so the result follows as before by expanding each term and applying Grönwall's inequality. □

Remark 17.1.2. For the remainder of the chapter, we assume that we have sufficient conditions for the result of Lemma 17.1.1 to hold, that is, that we have a unique solution to (17.2) for each s and x, and we have the bound stated in the lemma. However, we do not explicitly make use of the Lipschitz assumption beyond having these results.

Remark 17.1.3. We know, from Corollary 13.3.17, that the condition

$$E\Big[\sup_t \Big(\int_{\mathcal{Z}} \|g(\zeta, t, X_t^{(s,x)})\|\mu(d\zeta, \{t\})\Big)^p\Big] < \infty,$$

is implied by the (typically more easily verified) statement

$$E\Big[\int_{\mathcal{Z}\times[0,\infty[} \|g(\zeta, t, X_t^{(s,x)})\|^p \,\nu(d\zeta)dt\Big] < \infty.$$

Theorem 17.1.4. *For each $s \in [0,T]$, there is an $\mathbb{R}^d$ valued map*

$$X^{(s)}(\omega,t,x) : \Omega \times [s,T] \times \mathbb{R}^d \to \mathbb{R}^d,$$

such that

(i) for each $x \in \mathbb{R}^d$, $X^{(s)}(\omega,t,x)$ is a solution of the stochastic differential equation (17.2), (that is, $X^{(s)}(\omega,t,x) = X_t^{(s,x)}(\omega)$ a.s.), and
(ii) for each $t \in [s,T]$ the restriction of $X^{(s)}(\omega,u,x)$ to $\Omega \times [s,t] \times \mathbb{R}^d$ is $\mathcal{F}_{s,t} \otimes \mathcal{B}([s,t]) \otimes \mathcal{B}(\mathbb{R}^d)$ measurable.

Proof. The key difficulty is to establish measurability with respect to x. Consider a point with dyadic rational coordinates

$$\alpha_m^k = (k_1 2^{-m}, \ldots, k_d 2^{-m}) \in \mathbb{R}^d,$$

and consider a process $\{X^{(s,\alpha_m^k)}(\omega,t)\}_{t\in[s,t]}$ which is a solution of (17.2), with initial condition α_m^k at $t = s$, and which is continuous for all ω. This solution is progressively measurable, and, for fixed s and α_m^k, the map

$$(\omega,u) \mapsto X^{(s,\alpha_m^k)}(\omega,u) \qquad \text{restricted to } \Omega \times [s,t]$$

is $\mathcal{F}_{s,t} \otimes \mathcal{B}([s,t])$ measurable.

We now approximate an arbitrary point x. For $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ let $\alpha_m^k(x)$ be such that

$$(\alpha_m^k(x))_j = k_j 2^{-m} \leq x_j \leq (k_j+1)2^{-m} \text{ for all } 1 \leq j \leq d$$

and write

$$X^{(s;m)}(\omega,t,x) := X^{(s,\alpha_m^k(x))}(\omega,t).$$

For each x and m, we know that $X^{(s;m)}(\omega,t,x)$ is a.s. right-continuous in t and the function

$$(\omega,u,x) \mapsto X^{(s;m)}(\omega,u,x) \qquad \text{restricted to } \Omega \times [s,t]$$

is $\mathcal{F}_{s,t} \otimes \mathcal{B}([s,t]) \otimes \mathcal{B}(\mathbb{R}^d)$-measurable. Define

$$X^{(s)}(x,t,\omega) = \limsup_{m\to\infty} X^{(s;m)}(\omega,t,x). \tag{17.5}$$

We see that the process $X^{(s)}(\omega,t,x)$ has the measurability property of statement (ii).

It remains to show that $X^{(s)}(\omega,t,x)$ solves (17.2) for each x. Let $X^{(s,x)}(t,\omega)$ be the unique continuous solution of (17.2) with initial condition $x \in \mathbb{R}^d$ at $t = s$. Then, from Lemma 17.1.1,

$$E\Big[\sup_{s\leq t\leq T} \|X^{(s,x)}(t) - X^{(s,\alpha_m^k(x))}(t)\|^2\Big] \leq C\|x - \alpha_m^k(x)\|^2 \leq Cd2^{-m}.$$

By the Borel–Cantelli lemma (Theorem 2.1.13),

$$P\Big(\sup_{s\le t\le T}\|X^{(s,x)}(\omega,t)-X^{(s,\alpha_m^k(x))}(\omega,t)\|>1/k \\ \text{for infinitely many } k\in\mathbb{N}\Big)=0.$$

Therefore, for each $x\in\mathbb{R}^d$,

$$P\Big(X^{(s;m)}(\omega,\cdot,x)\to X^{(s,x)}(\omega,\cdot)\quad\text{uniformly on } [s,T]\Big)=1.$$

Consequently, up to indistinguishability,

$$X^{(s)}(\omega,t,x)=X^{(s,x)}(\omega,t),$$

and so, for every $x\in\mathbb{R}^d$, $X^{(s)}(\omega,t,x)$ is a solution of (17.2) with initial condition x at $t=s$, which is càdlàg for almost all ω. □

Corollary 17.1.5. *The stochastic differential equation* (17.2) *has a unique solution* $\{X_t^{(s,x)}\}_{t\ge s}$ *which is* $\mathcal{F}_{s,t}\otimes\mathcal{B}(\mathbb{R}^d)$ *measurable as a function of* (ω,x).

Lemma 17.1.6. *Suppose* $\phi:\mathbb{R}^d\to\mathbb{R}$ *is a bounded Borel measurable function. Then, for any* $t\in[s,T]$*, the map* $f(\omega,x)=\phi\big(X^{(s,x)}(\omega,t)\big)$ *is* $\mathcal{F}_{s,t}\otimes\mathcal{B}(\mathbb{R}^d)$ *measurable.*

Proof. Suppose first that $\phi=I_A$ for some set $A\in\mathcal{B}(\mathbb{R}^d)$. Then

$$f(\omega,x)=I_A\big(X^{(s,x)}(\omega,t)\big),$$

and so

$$\{(\omega,x):I_A(X)=1\}=\{(\omega,x):X_{s,x}(\omega,t)\in A\}\in\mathcal{F}_{s,t}\otimes\mathcal{B}(\mathbb{R}^d).$$

Therefore, the lemma is true for all $A\in\mathcal{B}(\mathbb{R}^d)$. The result follows for general ϕ by approximation with simple functions. □

17.2 Transition Probabilities

We now seek to understand the law of the solution of (17.2), that is, 'What is the probability that, at time t, $X^{(s,x)}$ will take values in a given set A?'. This is possible mainly because we can show that X is a Markov process. We give a definition here, however, we have already seen this property in relation to Poisson processes and Brownian motion, in Theorem 5.5.23 and Exercise 14.7.7.

Definition 17.2.1. *We say that a stochastic process X is a* Markov process *with respect to a filtration $\{\mathcal{F}_t\}_{t\geq 0}$ if, for every $s \leq t$ and every bounded Borel measurable function $\phi : \mathbb{R}^d \to \mathbb{R}$,*

$$E[\phi(X_t)|\mathcal{F}_s] = E[\phi(X_t)|X_s]$$

or equivalently, for any $A \in \mathcal{B}(\mathbb{R}^d)$, $P(X_t \in A|\mathcal{F}_s) = P(X_t \in A|X_s)$.

We say that the process X is a strong Markov process *if the above holds for s and t replaced with stopping times.*

It is important to note that the property of being a Markov process is not preserved under changes of filtration or probability.

Definition 17.2.2. *For fixed x, s and t, with $s \leq t$, write*

$$P(s, x; t, A) = P\big(X^{(s,x)}(\omega, t) \in A\big), \qquad \text{for } A \in \mathcal{B}(\mathbb{R}^d).$$

We call this the transition probability function *of the process X. As*

$$P(s, x; t, A) = \int_\Omega I_A\big(X^{(s,x)}(\omega, t)\big) dP = E\big[I_A\big(X^{(s,x)}(\omega, t)\big)\big],$$

it follows from Lemma 17.1.6 that, for fixed s, t and A, the map $x \mapsto P(s, x; t, A)$ is $\mathcal{B}(\mathbb{R}^d)$ measurable. For fixed x, s and t, by dominated convergence, we also know that $P(s, x; t, \cdot)$ is a probability measure on $\mathcal{B}(\mathbb{R}^d)$.

The following theorem is stated using the assumption of Lipschitz coefficients; however, as observed in Remark 17.1.2, this is only needed to guarantee existence and stability of solutions.

Theorem 17.2.3. *Consider the stochastic differential equation*

$$\begin{cases} dX_t = f(t, X_t)dt + \sigma(t, X_t)dW_t + \displaystyle\int_{\mathcal{Z}} g(\zeta, t, X_{t-})\tilde{\mu}(d\zeta, dt), \\ X_0 = x_0 \in \mathbb{R}^d, \end{cases} \tag{17.6}$$

where x_0 is independent of $\mathcal{F}_{0,\infty}$ and the coefficients satisfy the Lipschitz continuity and boundedness conditions at the start of the chapter ((17.4) and (17.3)). This equation has a unique solution X which is a Markov process relative to $\{\mathcal{F}_t\}_{t\geq 0}$. If $g \equiv 0$, then X is a continuous Markov process.

Proof. Let $X^{(s,x)}(\omega, t)$ be the unique solution of (17.6) and write

$$f(\omega, x) = \phi\big(X^{(s,x)}(\omega, t)\big)$$

where, as in Lemma 17.1.6, $\phi : \mathbb{R}^d \to \mathbb{R}$ is any bounded Borel measurable function. For each $x \in \mathbb{R}^d$, we know $\omega \mapsto f(\omega, t)$ is $\mathcal{F}_{s,t}$-measurable, and so is independent of $\mathcal{F}_s$.

Consequently, if Z is a simple $\mathcal{F}_s$-measurable random variable,

$$E\big[\phi\big(X^{(s,Z)}(\omega,t)\big)\big|\mathcal{F}_s\big] = g(Z),$$

where

$$g(x) = E[\phi(X_t^{(s,x)})].$$

An approximation argument and L^2-continuity of $X_t^{(s,x)}$ with respect to x shows this holds for any $\mathcal{F}_s$-measurable Z. Write

$$X_s = X^{(0,x_0)}(\omega,s),$$

so X_s is $\mathcal{F}_s$-measurable, and take $Z = X_s$ in the above. By the uniqueness of the solution, we have that

$$X^{(s,X_s)}(t,\omega) = X^{(0,x_0)}(t,\omega) = X_t,$$

so

$$E[\phi(X_t)|\mathcal{F}_s] = g(X_s) \quad \text{a.s.}$$

Therefore, for any $s \leq t$, we have

$$E[\phi(X_t)|X_s] = E\big[E[\phi(X_t)|\mathcal{F}_s]\big|X_s\big] = E[g(X_s)|X_s] = g(X_s) = E[\phi(X_t)|\mathcal{F}_s],$$

and it follows that X is a Markov process with respect to $\{\mathcal{F}_t\}_{t\geq 0}$. □

Remark 17.2.4. One can also show that, if f, σ and g are independent of time, X is a strong Markov process. The simplest argument for this is to first note that W and $\tilde{\mu}$ are strong Markov (for $\tilde{\mu}$, this follows from combining Theorem 5.5.23 and Lemma 13.5.11), and so one can perform a time change $\tilde{\mathcal{F}}_t = \mathcal{F}_{S+t}$ for S a stopping time. As W and $\tilde{\mu}$ are strong Markov, we see that X_S is independent of $\tilde{\mathcal{F}}_{0,t}$ for any t. Taking a stopping time T in the place of t in the proof of Theorem 17.2.3 then yields the result.

Theorem 17.2.5 (Chapman–Komogorov equation). *For any $s \leq u \leq t$, any $x \in \mathbb{R}^d$ and $A \in \mathcal{B}(\mathbb{R}^d)$, the transition probability function satisfies the* Chapman–Kolmogorov *equation*

$$P(s,x;t,A) = \int_{\mathbb{R}^d} P(u,y;t,A)P(s,x;u,dy).$$

Proof. From the definition of the transition probability,

$$E\big[\phi\big(X_t^{(s,x)}\big)\big] = \int_{\mathbb{R}^d} \phi(y)P(s,x;t,dy). \tag{17.7}$$

Taking $\phi = I_A$ for some $A \in \mathcal{B}(\mathbb{R}^d)$, we have, for any $s \leq u$,

$$P(X_u^{(s,x)} \in A|\mathcal{F}_s) = P(X_u^{(s,x)} \in A|X_s^{(s,x)}) = P(s,x;u,A) \quad \text{a.s.}$$

Now take $\phi(y) = P(u, y; t, A)$, where $u \le t$ and $A \in \mathcal{B}(\mathbb{R}^d)$ is arbitrary. Then

$$\phi(X_t^{(s,x)}) = P\big(u, X_u^{(s,x)}; t, A\big)$$

and so, from (17.7),

$$\begin{aligned}
\int_{\mathbb{R}^d} P(u, y; t, A) P(s, x; u, dy) &= E[\phi(X_t^{(s,x)})] = E\big[P\big(u, X_u^{(s,x)}; t, A\big)\big] \\
&= E\big[P(X_t^{(s,x)} \in A | X_u^{(s,x)})\big] = E\big[E[I_A\big(X_t^{(s,x)}\big) | X_u^{(s,x)}]\big] \\
&= E\big[I_A\big(X_t^{(s,x)}\big)\big] = P(X_t^{(s,x)} \in A) \\
&= P(s, x; t, A).
\end{aligned}$$

□

17.3 Feller Processes

Definition 17.3.1. *Suppose* $0 \le s < t \le T$, $x \in \mathbb{R}^d$, *and that* $X^{(s,x)}(t, \omega)$ *is the unique solution of* (17.2). *If* $B_0(\mathbb{R}^d) = B_0$ *is the space of all real valued, bounded, Borel measurable functions on* $\mathbb{R}^d$, *then we can define a family of linear operators* $\{T_s^t\}_{0 \le s \le t}$ *on* B_0 *by*

$$T_s^s = I, \ \textit{the identity operator,}$$
$$T_s^t v(x) = \int_{\mathbb{R}^d} v(y) P(s, x; t, dy),$$

where $P(s, x; t, dy)$ *is the transition probability function of the process* X. *The family* $\{T_s^t\}_{0 \le s \le t}$ *is called the transition semigroup of the process* X.

Note that
$$T_s^t v(X_s) = E[v(X_t) | \mathcal{F}_s] = E[v(X_t) | X_s].$$

Remark 17.3.2. If $0 \le s \le t \le u \le T$, it is a consequence of the Chapman–Kolmogorov equation that, for $v \in B_0$,

$$T_s^u v = T_s^t T_t^u v.$$

That is, the family of operators is a semigroup, which justifies the name.

Remark 17.3.3. From the definition, it is clear that the operators T_s^t are *positive*, that is if $v(x) \ge 0$ for all x, then $T_s^t v(x) \ge 0$ for all x. They also satisfy the inequality $\sup_x |T_t^s v(x)| \le \sup_y |v(y)|$ so, using Notation 17.3.4, $\|T_s^t\| \le 1$.

Of particular importance are those cases where the transition probability does not depend directly on time, in particular, when T_s^t depends only on $(t - s)$. This occurs when the dynamics of X are autonomous, i.e. f, σ and g do not depend on t. In this case, we often simplify and write T_{t-s} for T_s^t. This leads to the following important special case.

Notation 17.3.4. Write C_b for the space of bounded real continuous functions on $\mathbb{R}^d$, and $C_b^2 = C_b^2(\mathbb{R}^d)$ for the subspace of functions in C_b which also have first and second derivatives in C_b.

Write C_0 for the space of continuous functions which vanish at infinity. Note that $C_0 \subset C_b$. For a function $v \in C_0$ we write $\|v\| = \sup_x |v(x)|$. For a linear operator T, as usual we write

$$\|T\| = \sup_{\{v:\|v\|\leq 1\}} \|Tv\| = \sup_v \frac{\|Tv\|}{\|v\|}.$$

Definition 17.3.5. *A family of positive linear operators $\{T_t\}_{t\geq 0}$ on C_0 is called a* Feller semigroup on C_0 *if, for every $v \in C_0$,*

(i) $T_0 = I$ *and* $\|T\| \leq 1$,
(ii) $T_t T_s = T_{s+t}$,
(iii) $\lim_{t\downarrow 0} \|T_t f - f\| = 0$ *for every* $f \in C_0$.

A process whose transition semigroup is a Feller semigroup will be called a Feller process.

The 'strong continuity' of point (iii) is often more easily verified using the following result (the presentation of which is adapted from Revuz and Yor [155]).

Theorem 17.3.6. *A transition semigroup satisfies point (iii) of Definition 17.3.5 if and only if $T_t C_0 \subseteq C_0$ and $\lim_{t\downarrow 0} T_t v(x) = v(x)$ for all $x \in \mathbb{R}^d$ and all $v \in C_0$.*

Proof. The necessity of these statements is clear from the definition of a Feller semigroup (property (ii) implies that $T_t C_0 \subseteq C_0$, otherwise the semigroup is not well defined). To show sufficiency, suppose $v \in C_0$ and (i) and (ii) hold. Then $T_t v \in C_0$ by (i), so $\lim_{h\to 0} T_{t+h} v(x) = T_t v(x)$ by (ii). It follows that the map $(t, x) \mapsto T_t v(x)$ is right-continuous in t, continuous in x, and hence measurable on $[0,\infty[\, \times \mathbb{R}^d$. Therefore, for each $p > 0$, the function

$$x \mapsto U_p v(x) := \int_{[0,\infty[} e^{-pt} T_t v(x) dt$$

is measurable and, by (ii), $\lim_{p\to\infty} pU_p v(x) = v(x)$. Furthermore, we can verify that $U_p v \in C_0$, and that U_p satisfies the 'resolvent equation'

$$U_p v - U_q v = (q-p)U_p U_q v = (q-p)U_q U_p v.$$

Therefore, the image $D = U_p(C_0)$ does not depend on $p > 0$. By direct calculation, $\|pU_p v\| \leq \|v\|$.

By the dominated convergence theorem, if γ is any bounded measure on C_0 (with its Borel σ-algebra) such that $\gamma(D) = 0$, then

$$\int_{C_0} v\, d\gamma = \lim_{p\to\infty} \int_{C_0} pU_p v\, d\gamma = 0.$$

Therefore, $\gamma = 0$, which implies D is dense in C_0. Using Fubini's theorem,

$$T_t U_p v(x) = e^{pt} \int_{[t,\infty[} e^{-ps} T_s v(x) ds,$$

and so

$$\|T_t U_p v - U_p v\| \leq (e^{pt} - 1)\|U_p v\| + t\|v\|.$$

It follows that $\lim_{t\downarrow 0} \|T_t v - v\| = 0$ for all $v \in D$, and the result follows as D is dense. □

Corollary 17.3.7. *If X is a solution to (17.6) with bounded Lipschitz continuous coefficients f, σ and g which do not depend on t, then X is a Feller process.*

Proof. First notice that for any fixed $t > 0$, we know that X is a.s. continuous at $t = 0$ (as μ_p is continuous, X is quasi left-continuous, so the probability of a jump at any predictable time is zero). As $v \in C_0$ is bounded and continuous, by dominated convergence we have

$$T_t v(x) = E[v(X_t)|X_0 = x] \to E[v(X_0)|X_0 = x] = v(x) \quad \text{as } t \to 0.$$

By Lemma 17.1.1, if K is a bound for the coefficients, then we have the estimate

$$\begin{aligned}
&E\big[\sup_{0\leq s\leq t} \|X_s^{(0,x)} - x\|^2\big] \\
&\leq C\Big(\int_{[0,t]} E\big[\|f(X_s^{(0,x)})\|^2 + \|\sigma(X_s^{(0,x)})\|^2 + \|g(\cdot, X_s^{(0,x)})\|_\nu^2\big] ds\Big) \\
&\leq 3CK^2 t.
\end{aligned}$$

If K is also a bound on v, for any $\alpha > 0$, we also have

$$\begin{aligned}
T_t v(x) &= E[v(X_t)|X_0 = x] \\
&\leq KP(\|X_t^{0,x} - x\| > \alpha\|x\|^{1/2}) + \sup_{\{y:\|y-x\|\leq\alpha\|x\|^{1/2}\}} v(y) \\
&\leq K\frac{E\big[\|X_t^{0,x} - x\|^2\big]}{\alpha^2\|x\|} + \sup_{\{y:\|y-x\|\leq\alpha\|x\|^{1/2}\}} v(y) \\
&\leq \frac{3CK^3 t}{\alpha^2\|x\|} + \sup_{\{y:\|y-x\|\leq\alpha\|x\|^{1/2}\}} v(y).
\end{aligned}$$

Taking $\|x\| \to \infty$, we see that $\sup_{\{y:\|y-x\|\leq\alpha\|x\|^{1/2}\}} v(y) \to 0$ (as $v \in C_0$) and hence $T_t v(x) \to 0$ as $\|x\| \to \infty$. Similarly, we can verify that $T_t v$ is continuous, and it follows that $T_t C_0 \subseteq C_0$.

By Theorem 17.3.6, we conclude that X is a Feller process. □

This setting also allows us to give a simple result on right-continuity of filtrations, which we alluded to earlier.

Theorem 17.3.8. *Let X be a Feller process in a probability space $(\Omega, \mathcal{F}, P)$, and consider the σ-algebras $\mathcal{G}_t^0 = \sigma(X_s; s \leq t)$. If $\mathcal{N}$ denotes all subsets of P-null sets in $\mathcal{G}_\infty^0$, define $\mathcal{G}_t = \mathcal{G}_t^0 \vee \mathcal{N}$. Then the filtration $\{\mathcal{G}_t\}_{t\geq 0}$ is right-continuous and complete.*

Proof. Completeness of $\{\mathcal{G}_t\}_{t\geq 0}$ is trivial. As $\mathcal{G}_t$ and $\mathcal{G}_{t+} = \cap_{s>t}\mathcal{G}_s$ are both P-complete, it is enough to prove that for any $\mathcal{G}_\infty^0$-measurable nonnegative random variable Z,

$$E[Z|\mathcal{G}_t] = E[Z|\mathcal{G}_{t+}] \qquad P-\text{a.s.}$$

By the monotone class theorem, it is enough to prove this equality when $Z = \prod_{i=1}^n z_i(X_{t_i})$ for some $z_i \in C_0$ and $t_1 < t_2 < \ldots < t_n$.

We know that $E[Z|\mathcal{G}_t] = E[Z|\mathcal{G}_t^0]$ P-a.s. for each t. Fixing t, we know that there is an integer k such that $t_{k-1} \leq t < t_k$, and for $h < t_k - t$,

$$\begin{aligned} E[Z|\mathcal{G}_{t+h}] &= \Big(\prod_{i=1}^{k-1} z_i(X_{t_i})\Big) E\Big[\prod_{i=k}^{n} z_i(X_{t_i})\Big|\mathcal{G}_{t+h}\Big] \\ &= \Big(\prod_{i=1}^{k-1} z_i(X_{t_i})\Big) y_h(X_{t+h}) \end{aligned}$$

where y_h is the continuous function

$$y_h(x) := \int \Big(\int \Big(\prod_{i=k}^{n} z_i(x_i)\Big) \bigotimes_{i=k+1}^{n} P(t_{i-1}, x_{i-1}; t_i, dx_i)\Big) P(t+h, x; t_k, dx_k),$$

the integrals being taken over copies of $\mathbb{R}^d$. If we take $h \downarrow 0$, y_h converges uniformly on $\mathbb{R}^d$ to

$$\begin{aligned} y(x) &:= \int \Big(\int \Big(\prod_{i=k}^{n} z_i(x_i)\Big) \bigotimes_{i=k+1}^{n} P(t_{i-1}, x_{i-1}; t_i, dx_i)\Big) P(t, x; t_k, dx_k) \\ &= E\Big[\prod_{i=k}^{n} z_i(X_{t_i})\Big| X_t = x\Big]. \end{aligned}$$

By right continuity of paths, $X_{t+h} \to X_t$ as $h \downarrow 0$ and, by Lemma 5.1.6,

$$E[Z|\mathcal{G}_{t+}] = \lim_{h\downarrow 0} E[Z|\mathcal{G}_{t+h}] = \Big(\prod_{i=1}^{k-1} z_i(X_{t_i})\Big) y(X_t) = E[Z|\mathcal{G}_t].$$

The result follows. □

Remark 17.3.9. The theory of Feller processes contains many other elegant results, which we shall not consider further, as this would require us to build up the underlying analytic theory in much greater detail. In particular, one can show that any Feller process admits a càdlàg modification, and that this modification is a quasi-left continuous process. The details can be found in Revuz and Yor [155], Ethier and Kurtz [77] or Kallenberg [113] among many other references.

17.4 Links to Partial (Integro-)Differential Equations

We now see how the transition semigroup of a (not-necessarily Feller) solution to (17.6) is generated by a certain integro-differential operator.

Definition 17.4.1. *Write $a = (a^{ij})$ for the $d \times d$ matrix $\sigma\sigma^\top$, where $\sigma^\top$ denotes the transpose of σ. Write $X = X^{(0,x_0)}$ for the solution of (17.2) with initial value $X_0 = x_0$.*

Let $\{\mathcal{L}_s\}_{s\geq 0}$ be the time-dependent second-order integro-differential operator defined by

$$\mathcal{L}_s v(x) = \sum_i f^i(s,x)\frac{\partial v}{\partial x^i}(x) + \frac{1}{2}\sum_{i,j} a^{ij}(s,x)\ \frac{\partial^2 v}{\partial x^i \partial x^j}(x)$$
$$+ \int_{\mathcal{Z}} \Big(v(x+g(\zeta,s,x)) - v(x) - \sum_i \frac{\partial v}{\partial x^i}(x) g^i(\zeta,s,x)\Big)\nu(d\zeta)\,.$$

We call $\mathcal{L}$ the infinitesimal generator *of the process X.*

Remark 17.4.2. If our process is continuous, so $\nu(dz) \equiv 0$, then we see that the integral term disappears and we are left with a second-order differential operator.

To justify this terminology, we observe the following result.

Theorem 17.4.3. *(i) For each $v \in C_b^2$, the map $(u,y) \mapsto \mathcal{L}_u v(y)$ is jointly measurable in u and y.*

(ii) For each $v \in C_b^2$, for every $x \in \mathbb{R}^d$ and $0 \leq s < u \leq T$,

$$\int_{\mathbb{R}^d} |(\mathcal{L}_u v)(y)| P(s,x;u,dy) < \infty.$$

(iii) For each $v \in C_b^2$, for every $x \in \mathbb{R}^d$ and $0 \leq s < t \leq T$,

$$T_s^t v(x) = v(x) + \int_{]s,t]} T_s^u \mathcal{L}_u v(x) du = v(x) + \int_{]s,t]} \mathcal{L}_u T_s^u v(x) du.$$

(iv) If f, σ and g are continuous in s and x, then, for each $v \in C_b^2$, for every $x \in \mathbb{R}^d$ and $0 \leq s < T$,

$$\lim_{h\downarrow 0} \frac{T_s^{s+h} v(x) - v(x)}{h} = \mathcal{L}_s v(x).$$

Proof. The measurability properties of part (i) are immediate from the definitions.

As $v \in C_b^2$, we know that

$$\Big\| v\big(x + g(\zeta, s, x)\big) - v(x) - \sum_i \frac{\partial v}{\partial x^i}(x) g^i(\zeta, s, x) \Big\| \leq K \|g(\zeta, s, x)\|^2$$

for some K, and as $g \in L^2(\nu)$, the integral term in $\mathcal{L}_t v$ is well defined.

For fixed (s, x), suppose X is the solution to (17.2), that is

$$X_t = x + \int_{]s,t]} f(u, X_t) du + \int_{]s,t]} \sigma(u, X_u) dW_u + \int_{\mathcal{Z}\times]s,t]} g(\zeta, u, X_{u-}) \tilde{\mu}(d\zeta, du).$$

Applying Itô's rule with $v \in C_b^2$, we obtain

$$\begin{aligned} dv(X_t^{(s,x)}) &= \mathcal{L}_u v(X_u) du + \sum_i \frac{\partial v}{\partial x_i}(X_{u-}) \Big(\sum_j \sigma^{i,j}(s, X_s) dW_u^j \Big) \\ &\quad + \int_{\mathcal{Z}} \Big(v\big(X_{u-} + g(\zeta, u, X_{u-})\big) - v(X_{u-}) \Big) \tilde{\mu}(d\zeta, du). \end{aligned}$$

(Note that when simplifying to obtain $\mathcal{L}_u v$, the left limit in X in the $\nu(d\zeta)dt$ integral can be omitted.) From Lemma 17.1.1 and the growth bounds (17.3), we see that, for any $v \in C_b^2$, the final two terms in this equation integrate to martingales. Therefore, integrating and taking an expectation,

$$E\big[v\big(X_t^{(s,x)}\big)\big] = v(x) + \int_{]s,t]} E\big[\mathcal{L}_u v\big(X_u^{(s,x)}\big)\big] du.$$

The final integral exists if $v \in C_b^2$, using Lemma 17.1.1 and the growth bounds (17.3). This proves the first equality in part (iii) of the theorem. The boundedness of v and its derivatives, along with the growth conditions (17.3), also imply that $|\mathcal{L}_u v(X_t^{(s,x)})| \leq K(1 + \|X_t\|^2)$ for some $K > 0$, and (ii) then follows from the definition of $P(s, x; t, dy)$. We know

$$T_s^u \mathcal{L}_u v(x) = E[\mathcal{L}_u v(X_u^{(s,x)})] = \mathcal{L}_u E[v(X_u^{(s,x)})] = \mathcal{L}_u T_s^u v(x),$$

proving the second equality in (iii), where by dominated convergence, we can exchange the order of differentiation and expectation and, with Fubini's theorem for the integral term, we see that $\mathcal{L}_u$ commutes with the expectation.

Finally, when f, σ and g are continuous, by dominated convergence we see that

$$T_s^u \mathcal{L}_u v(x) = E\big[\mathcal{L}_u v\big(X_u^{(s,x)}\big)\big]$$

is a continuous function with respect to $u \in [s, T]$, and with (iii) this implies (iv), by the fundamental theorem of calculus.

□

Remark 17.4.4. When X is a Feller process, then its infinitesimal generator is independent of time, that is, $\mathcal{L}_t = A$, for A an integro-differential operator on C_b^2.

Remark 17.4.5. By taking $s = 0$ and $X_0 = x$ in Theorem 17.4.3(iii), we obtain 'Dynkin's formula',

$$E[v(X_t)] = v(x) + \int_{[0,t]} E[\mathcal{L}_u v(X_u)]du.$$

We now obtain the classical definition of the infinitesimal generator.

Definition 17.4.6. *For X a Markov process, we say that a function $v \in C_0$ is in the domain $\mathcal{D}_\mathcal{L}$ of the infinitesimal generator if the limit*

$$\tilde{\mathcal{L}}_s v(x) := \lim_{h \downarrow 0} \frac{T_s^{s+h} v(x) - v(x)}{h}$$

exists in C_0 (that is, the limit is uniform in x). We say that a function $v \in C_0$ is in the domain $\mathbb{D}_\mathcal{L}$ of the extended *infinitesimal generator if there exists a Borel measurable function γ such that $\int_{[0,t]} |\gamma(s, X_s)|ds < \infty$ a.s. for every t and*

$$Y_t = v(X_t^{(0,x)}) - v(x) - \int_{[0,t]} \gamma(s, X_s^{(0,x)})ds$$

defines an $\{\mathcal{F}_t\}_{t\geq 0}$-right continuous local martingale for every x. Clearly $\mathcal{D}_\mathcal{L} \subset \mathbb{D}_\mathcal{L}$, and the operator $\tilde{\mathcal{L}}_t v$ extends to $\mathbb{D}_\mathcal{L}$ with the definition

$$\tilde{\mathcal{L}}_t v = \gamma.$$

Remark 17.4.7. Note that γ is only defined up to appropriate sets of measure zero, so the map $v \mapsto \tilde{\mathcal{L}}_t v$ is only 'almost' linear.

Remark 17.4.8. From our above results, it is easy to deduce that $C_b^2 \subset \mathbb{D}_\mathcal{L}$, and we have the equivalence $\tilde{\mathcal{L}} = \mathcal{L}$ on C_b^2 (cf. Corollary 17.4.11). If the jumps of X are bounded (i.e. g is bounded), by localization, this easily extends to demonstrate $C^2 \subseteq \mathbb{D}_\mathcal{L}$. We have also seen that if f, σ and g are continuous in s and x, then $C_b^2 \subset \mathcal{D}_\mathcal{L}$. Conversely, if we permitted the extension to (17.2) described in Remark 17.0.3 (with appropriate modifications to the generator), then it would not necessarily be the case that $C_b^2 \subset \mathbb{D}_\mathcal{L}$. See Ethier and Kurtz [77] for details.

Definition 17.4.9. *To deal with the case where the jumps are not bounded, we define the space*

$$C_\nu^2 = \Big\{ v \in C^2 : \xi \mapsto \Big(v(x + g(\zeta, s, x)) - v(x) - \sum_i \frac{\partial v}{\partial x^i}(x) g^i(\zeta, s, x) \Big) \in L^1(\nu) \\ \text{for all } x \in \mathbb{R}^d, s \in [0, \infty] \Big\}.$$

We have seen that $C_b^2 \subset C_\nu^2 \subseteq C^2$ and, if g is bounded, $C_\nu^2 = C^2$. By localization, we can see that in general $C_\nu^2 \subseteq \mathbb{D}_\mathcal{L}$.

Recall that we say a function $v : [0, T[\times \mathbb{R}^d \to \mathbb{R}^d$ is $C^{1,2}$ if it is once continuously differentiable in its first argument and twice continuously differentiable in its second. Extending this, we say a function is $C^{1,2}_\nu$ if it is also in C^2_ν with respect to its second argument.

We now give a fundamental result linking solutions to SDEs with solutions to certain partial integro-differential equations. Our presentation of this result is guided by Karatzas and Shreve [117].

Theorem 17.4.10 (Feynman–Kac Theorem). *Let $r : [0, t] \times \mathbb{R}^d \to [0, \infty[$ be Borel measurable, $b : [0, T] \times \mathbb{R}^d \to \mathbb{R}$ be Borel measurable and satisfy, for some $K > 0, m \geq 1$,*

$$|b(t, x)| \leq K(1 + \|x\|^{2m}) \qquad \text{or} \qquad b(t, x) \geq 0, \qquad \text{for all } (t, x) \in [0, T] \times \mathbb{R}^d,$$

and $\xi : \mathbb{R}^d \to \mathbb{R}$ be Borel measurable and satisfy

$$|\xi(x)| \leq K(1 + \|x\|^{2m}) \qquad \text{or} \qquad \xi(x) \geq 0, \qquad \text{for all } x \in \mathbb{R}^d.$$

Suppose g is such that

$$E\Big[\int_{\mathcal{Z}\times[0,T]} \|g(\zeta, t, X_t^{(s,x)})\|^p \nu(d\zeta)dt\Big] < \infty,$$

for some $p > 2m$.

Consider a function $v : [0, T] \times \mathbb{R}^d \to \mathbb{R}$ which is $C^{1,2}_\nu$ on $[0, T[\times \mathbb{R}^d$, and satisfies the Cauchy problem

$$\begin{aligned} -\frac{\partial v}{\partial t} + rv &= \mathcal{L}_t v + b &&\text{on } [0, T[\times \mathbb{R}^d, \\ v(T, x) &= \xi(x) &&\text{for } x \in \mathbb{R}^d, \end{aligned}$$

as well as the growth condition,

$$\max_{t\in[0,T]} |v(t, x)| \leq K(1 + \|x\|^{2m}) \qquad \text{for } x \in \mathbb{R}^d.$$

Then v admits the representation

$$\begin{aligned} v(t, x) = E\Big[&\xi(X_T^{(t,x)}) \exp\Big(-\int_{]t,T]} r(s, X_s^{(t,x)})ds\Big) \\ &+ \int_{]t,T]} b(u, X_u^{(t,x)}) \exp\Big(-\int_{[t,u]} r(s, X_s^{(t,x)})ds\Big)du\Big], \end{aligned}$$

where $X^{(t,x)}$ is the solution to (17.2).

Proof. For simplicity, we omit (t, x) when writing $X^{(t,x)}$. Consider the process $v(u, X_u) \exp\Big(-\int_{]t,u]} r(s, X_s)ds\Big)$. Let $T_n = \inf\{s \geq t : \|X_s\| \geq n\} \wedge T$.

Applying Itô's rule and taking an expectation, as v satisfies the Cauchy problem, we have

$$\begin{aligned}v(t,x) = E\Big[\int_{]t,T_n]} b(u,X_u)\exp\Big(-\int_{]t,u]} r(s,X_s)ds\Big)du \\ + v(T_n,X_{T_n})\exp\Big(-\int_{]t,T_n]} r(s,X_s)ds\Big) I_{\{T_n<T\}} \\ + \xi(X_T)\exp\Big(-\int_{]t,T]} r(s,X_s)ds\Big) I_{\{T_n\geq T\}}\Big].\end{aligned}$$

By Lemma 17.1.1, we know that, for our choice of p,

$$E\big[\sup_{u\in[t,T]} \|X_u^{(t,x)}\|^p\big] \leq K(1+\|x\|^p)$$

for some constant K. By Markov's inequality,

$$P(T_n < T) \leq K(1+\|x\|^p)/n^p.$$

From the growth bounds or nonnegativity of b, and the nonnegativity of r, the dominated convergence theorem or monotone convergence theorem implies

$$\begin{aligned}&E\Big[\int_{]t,T_n]} b(u,X_u)e^{-\int_{]t,u]} r(s,X_s)ds}du\Big] \\ &\to E\Big[\int_{]t,T]} b(u,X_u)e^{-\int_{]t,u]} r(s,X_s)ds}du\Big].\end{aligned}$$

We know that

$$\begin{aligned}|v(T_n,X_{T_n})|I_{\{T_n<T\}} &\leq K(1+\|X_{T_n}\|^{2m})I_{\{T_n<T\}} \\ &\leq 2^{2m-1}K(1+n^{2m}+\|\Delta X_{T_n}\|^{2m})I_{\{T_n<T\}}.\end{aligned}$$

As $\|\Delta X_{T_n}\|^2 \leq 2^{2m-1}\sup_{u\in[t,T]}\|X_u^{(t,x)}\|^{2m}$, which is integrable (Lemma 17.1.1), by dominated convergence we know $E[\|\Delta X_{T_n}\|^{2m}I_{\{T_n<T\}}] \to 0$. So, as $p > 2m$, from our bound on $P(T_n < T)$ we see that, sending $n\to\infty$,

$$\begin{aligned}&E\Big[|v(T_n,X_{T_n})|e^{-\int_{]t,u]} r(s,X_s)ds} I_{\{T_n<T\}}\Big] \\ &\leq 2^{2m-1}KE\Big[(1+n^{2m}+\|\Delta X_{T_n}\|^{2m})I_{\{T_n<T\}}\Big] \\ &\leq 2^{2m-1}K^2(1+\|x\|^p)\frac{1+n^{2m}}{n^p} + 2^{2m-1}KE\Big[\|\Delta X_{T_n}\|^{2m}I_{\{T_n<T\}}\Big] \\ &\to 0.\end{aligned}$$

Similarly, by either dominated or monotone convergence, we know

$$E\Big[\xi(X_T)e^{-\int_{]t,u]} r(s,X_s)ds} I_{\{T_n\geq T\}}\Big] \to E\Big[\xi(X_T)e^{-\int_{]t,u]} r(s,X_s)ds}\Big],$$

which gives the desired representation. □

Corollary 17.4.11. *Let v be a $C^{1,2}_\nu$ function satisfying the polynomial growth condition: for some $K > 0, m \geq 1$,*

$$\max_{t\in[0,T]} \Big\{|v(t,x)| + \Big|\mathcal{L}_t v(t,x) + \frac{\partial v}{\partial t}(t,x)\Big|\Big\} \leq K(1+\|x\|^{2m}) \qquad \text{for } x \in \mathbb{R}^d,$$

and suppose g is such that

$$E\Big[\int_{\mathcal{Z}\times[0,T]} \|g(\zeta,t,X_t^{(s,x)})\|^p \nu(d\zeta)dt\Big] < \infty$$

for some $p > 2m$. If X is the solution to our SDE (17.6), then the process

$$M_t^v = v(t,X_t) - v(0,x) - \int_{]0,t]} \Big(\mathcal{L}_t v(u,X_u) + \frac{\partial v}{\partial t}(u,X_u)\Big) du$$

is a martingale. In particular, if v does not depend on t, then

$$M_t^v = v(X_t) - v(x) - \int_{]0,t]} \mathcal{L}_t v(X_u) du$$

is a martingale.

Proof. Taking $b(t,x) = -\mathcal{L}_t v(t,x) - \frac{\partial v}{\partial t}(t,x)$ and $r \equiv 0$ in the Theorem 17.4.10, we have

$$v(t,X_t) = E\Big[v(T,X_T) + \int_{]t,T]} b(u,X_u)du\Big|X_t\Big].$$

By addition and the Markov property,

$$M_t^v = E\Big[v(T,X_T) + \int_{]0,T]} b(u,X_u)du\Big|\mathcal{F}_t\Big],$$

which is a classical example of a martingale, as the growth conditions and Lemma 17.1.1 ensure the terms inside the expectation are integrable. □

Remark 17.4.12. The above result is the basis for the *martingale problem*, which provides an alternative approach to constructing solutions to the SDE (17.6). We consider this in more detail in the coming chapter.

As a converse, we have the following result. In dimension one, a stronger result can be obtained using the theory of viscosity solutions of PDEs. See Theorem 19.5.3.

Theorem 17.4.13. *Let X be the solution to (17.6), and assume that X_t has support $\mathbb{R}^d$ for all $t > 0$. Suppose v is a $C^{1,2}_\nu$ function such that $v(t,X_t)$ is a local martingale. Then v satisfies the PIDE*

$$0 = \mathcal{L}_t v + \frac{\partial v}{\partial t} \qquad \text{on} \quad]0,T[\times\mathbb{R}^d.$$

Proof. Applying Itô's rule to $v(t, X_t)$, as above we obtain

$$v(t, X_t) = v(0, x) + \int_{]0,t]} \Big(\mathcal{L}_t v(u, X_u) + \frac{\partial v}{\partial t}(u, X_u)\Big) du + M_t^v$$

for M_t^v a local martingale. Therefore, we know that $\mathcal{L}_t v(u, X_u) + \frac{\partial v}{\partial t}(u, X_u) = 0$ $dt \times dP$-a.e. As X_t has support $\mathbb{R}^d$ for all $t > 0$ and v is continuous, it follows that v satisfies the stated PIDE for all $(t, x) \in]0, T[\times \mathbb{R}^d$. □

As a consequence of this, we derive the following classical equations for the density of X, assuming that it exists.

Theorem 17.4.14 (Kolmogorov Equations). *Suppose that, for all* $t \in]0, T[$*, the law of* $X_t^{(s,x)}$ *has a density with respect to Lebesgue measure, that is,* $P(s, x; t, dy) = p(s, x; t, y)dy$*, for some* p*. Suppose* g *is such that*

$$E\Big[\int_{\mathcal{Z}\times[0,T]} \|g(\zeta, t, X_t^{(s,x)})\|^q \nu(d\zeta)dt\Big] < \infty,$$

for some $q > 2$*. Provided* $(s, x) \mapsto p(s, x; t, y)$ *is* $C_\nu^{1,2}$*, the density satisfies*

$$-\frac{\partial p}{\partial s} = \mathcal{L}_s p(s, \cdot; t, y)$$

(the Kolmogorov Backward Equation*).*

Conversely, given f*,* $a = \sigma\sigma^\top$ *and* g *are also* $C^{1,2}$*, and for all* t, y*, in* (s, x)*, if the map* $(t, y) \mapsto p(s, x; t, y)$ *is* $C^{1,2}$*,and for all* t, y*,*

$$\zeta \mapsto \Big(p(t, y - g((\zeta, t, y)) - p(t, y) + \sum_i \frac{\partial[g^i(\zeta, \cdot, \cdot)p]}{\partial y^i}(t, y)\Big) \in L^1(\nu),$$

then we have

$$\frac{\partial p}{\partial t} = \mathcal{L}_t^* p(s, x; t, \cdot)$$

(the Kolmogorov Forward Equation *or* Fokker–Plank Equation*), where* $\mathcal{L}_t^*$ *is the adjoint of* $\mathcal{L}_t$*, and is given by, for* $p \in C^{1,2}$*,*

$$\begin{aligned}\mathcal{L}_t^* p(t, y) = &\sum_i \frac{\partial[f^i p]}{\partial y^i}(t, y) + \frac{1}{2}\sum_{i,j} \frac{\partial^2[a^{ij}p]}{\partial y^i \partial y^j}(t, y)\\ &+ \int_{\mathcal{Z}} \Big(p(t, y - g(\zeta, t, y)) - p(t, y) + \sum_i \frac{\partial[g^i(\zeta, \cdot, \cdot)p]}{\partial y^i}(t, y)\Big)\nu(d\zeta).\end{aligned}$$

Proof. For notational simplicity, we write $\partial_t v$ for $\partial v/\partial t$, and similarly for p. We first derive the backward equation. For any $A \in \mathcal{B}(\mathbb{R}^d)$, we know that

$$P(X_T^{(s,x)} \in A) = \int_{\mathbb{R}^d} I_{\{y\in A\}} p(s, x; T, y)dy = E\big[I_{\{X_t^{(s,x)}\in A\}}\big] = P(s, x; t, A).$$

The Chapman–Kolmogorov equation implies that, for any $s < t < T$,

$$\int_{\mathbb{R}^d} I_{\{y\in A\}} p(s,x;T,y)dy = \int_{\mathbb{R}^d} \Big(\int_{\mathbb{R}^d} I_{\{y\in A\}} p(t,x';T,y)dy \Big) p(s,x;t,x')dx'.$$

As A was arbitrary,

$$p(s,x;T,y) = \int_{\mathbb{R}^d} p(t,x';T,y)p(s,x;t,x')dx' = E[p(t,X_t^{(s,x)};T,y)] \quad dy\text{-a.e.}$$

For $X = X^{(0,x_0)}$, the solution to (17.6), we see that $p(s,X_s;T,y)$ is a martingale. As we assume the density is $C^{1,2}$ with respect to (s,x), Theorem 17.4.13 implies that $0 = \partial_s p + \mathcal{L}_s p$, as desired.

We now derive the forward equation. Let (s,x) be fixed and, for simplicity, we omit to write (s,x) as an argument of p. Let v be an arbitrary $C_b^{1,2}$-function such that $v(s,x) = 0$ and $v(t,y) \to 0$ as $t \to T$, uniformly in $y \in \mathbb{R}^d$. By Corollary 17.4.11, we have

$$\begin{aligned} 0 = E[v(T,X_T)] - v(s,x) &= E\Big[\int_{]s,T]} \partial_t v(t,X_t) + \mathcal{L}_t v(t,X_t)\Big] \\ &= \int_{]s,T]}\int_{\mathbb{R}^d} \Big(\partial_t v(t,y) + \mathcal{L}_t v(t,y)\Big) p(t,y)dy\,du. \end{aligned} \tag{17.8}$$

Using integration by parts, from our assumptions on v, we know that

$$\int_{]s,T]} (\partial_t v(t,y))p(t,y)du = -\int_{]s,T]} v(t,y)(\partial_u p(t,y))du.$$

As p is a probability density, we know $(1+\|y\|)p(t,y) \to 0$, dy-a.e., as $\|y\| \to \infty$, for all t. Again using integration by parts,

$$\int_{\mathbb{R}^d} f^i(t,y)(\partial_{y_i} v(t,y))p(t,y)dy = -\int_{\mathbb{R}^d} v(t,y)\partial_{y_i}[f^i p](t,y)dy,$$
$$\int_{\mathbb{R}^d} a^{ij}(t,y)(\partial^2_{y_i y_j} v(t,y))p(t,y)dy = \int_{\mathbb{R}^d} v(t,y)\partial^2_{y_i y_j}[a^{ij}p](t,y)dy,$$

and, for any ζ, by change of variables,

$$\int_{\mathbb{R}^d} v(t,y+g(\zeta,t,y))p(t,y)dy = \int_{\mathbb{R}^d} v(t,y)p(t,y-g(\zeta,t,y))dy.$$

Therefore, combining these equalities and using Fubini's theorem, we have the identity

$$\int_{\mathbb{R}^d} \mathcal{L}_t v(t,y)p(t,y)dy = \int_{\mathbb{R}^d} v(t,y)\mathcal{L}_u^* p(t,y)dy.$$

As v was an arbitrary function in a dense set (in $L^1([s,T] \times \mathbb{R}^d)$), this implies that $\mathcal{L}_t^*$ is the adjoint of $\mathcal{L}_t$ (in the sense of Lemma 1.5.10). From (17.8), we see that

$$0 = \int_{]s,T]} \int_{\mathbb{R}^d} v(t,y)\Big(- \partial_t p(t,y) + \mathcal{L}_u^* p(t,y)\Big) dy\, dt$$

and, as v was arbitrary, we have the forward equation for p. □

Example 17.4.15. The density $p(s,x;t,y)$ of a Brownian motion X satisfies the heat equations

$$-\frac{\partial p}{\partial s} = \frac{1}{2}\frac{\partial^2 p}{\partial x^2}, \qquad \frac{\partial p}{\partial t} = \frac{1}{2}\frac{\partial^2 p}{\partial y^2}.$$

(Note that s runs backwards in time, so the negative sign is natural.)

Remark 17.4.16. When the existence of a sufficiently differentiable density is not guaranteed, then, for a fixed x, Theorem 17.4.3(iii) allows us to formally write $\frac{d}{dt} T_s^t = \mathcal{L}_t^* T_s^t$. However, for each x, T_s^t is a linear operator on $C_b^{1,2}$, and so $\mathcal{L}_t^*$ should be thought of as the adjoint of $\mathcal{L}_t$ in a wider sense, based on the inherent duality between functions and measures. This provides a very general, if not particularly concrete, approach to the dynamics of the law of a general Markov process.

Remark 17.4.17. From the perspective of P(I)DE theory, the density of X corresponds to the fundamental solution of the PIDE. This follows because, for any function ξ satisfying the requirements of Theorem 17.4.10, we can write the solution of the Cauchy problem $\partial_t v + \mathcal{L}_t v = 0$, with boundary value $v(T,\cdot) = \xi(\cdot)$ as

$$v(s,x) = \int_{\mathbb{R}^d} \xi(y) p(s,x;T,y) dy.$$

Remark 17.4.18. In the continuous case ($\nu \equiv 0$), a sufficient condition for the existence of a smooth density p is, for example, that the coefficient functions f and σ in (17.2) have continuous derivatives up to the third order, which satisfy a growth condition. (See Gihman and Skorohod [88, p.99].)

More generally, the study of the question of the existence of a smooth density was one of the motivating factors leading to the development of 'Malliavin calculus' for stochastic processes. This theory is presented in Nualart [140] in the continuous case, among many other works. Bichteler, Gavereaux and Jacod [15] and Di Nunno, Øksendal and Proske [59] give versions of this theory for Lévy processes.

17.5 Exercises

Exercise 17.5.1. Let X be the Ornstein–Uhlenbeck process, with dynamics

$$dX_t = \kappa(\alpha - X_t)dt + dW_t$$

where W is a Brownian motion and κ and α are constants. If $\kappa > 0$, find constants μ and σ such that, if $X_0 \sim N(\mu, \sigma^2)$, then $X_t \sim N(\mu, \sigma^2)$ (that is, the distribution $N(\mu, \sigma)$ is a stationary distribution for X). Write down the generator of X and its adjoint, and verify that $\mathcal{L}^*\phi = 0$, where ϕ is the density of the stationary distribution of X.

Exercise 17.5.2. Let W be a P-Brownian motion, so that $W_1 \sim N(0,1)$. Consider an equivalent measure Q, given by

$$\frac{dQ}{dP} = f(W_1),$$

where f is a strictly positive C_b^2 function with $\int_{\mathbb{R}} f(x)\frac{1}{\sqrt{2\pi}}\exp(-x^2/2)dx = 1$.

(i) Show that $E^P[f(W_1)|\mathcal{F}_t] = g(t, W_t)$ for some $C^{1,2}$ function g.
(ii) Derive a PDE satisfied by g.
(iii) Write down the drift of W under Q, in terms of the function g.
(iv) Show that there is a drift process $\mu(t,x)$, locally Lipschitz in x, such that the process X defined by

$$dX_t = \mu(t, X_t)dt + dB_t, \qquad X_0 = 0$$

for B a Brownian motion, is a Markov process and X_1 has density $f(x)\frac{1}{\sqrt{2\pi}}\exp(-x^2/2)$ with respect to Lebesgue measure.

Exercise 17.5.3. For every $s \in [0,T]$, and $x \in \mathbb{R}^d$, let $\{X_t^{(s,x)}\}_{t\in[s,T]}$ be the solution to (17.2). Define $X_t^{(s,x)} = x$ for $t < s$. If f, σ and g are uniformly Lipschitz functions, then show that there exists C such that

$$E\Big[\sup_{s\in[0,T]} \|X_s^{(t,x)} - X_s^{(t',x')}\|^2\Big] \leq C(1 + \|x\|^2)(\|x - x'\|^2 + |t - t'|).$$

Exercise 17.5.4. Consider the PDE for $v : [0,T] \times [0,1] \to \mathbb{R}$

$$-\frac{dv}{dt} = f(t,x)\frac{dv}{dx} + \frac{\sigma^2(t,x)}{2}\frac{\partial^2 v}{\partial x^2}$$

with boundary conditions

$$v(T,x) = \phi(x), \qquad v(t,0) = \psi^0(t), \qquad v(t,1) = \psi^1(x).$$

Here f and σ are Lipschitz continuous in x and Borel measurable in t, and ϕ, ψ^0 and ψ^1 are bounded Borel measurable functions with $\phi(0) = \psi^0(0)$ and $\phi(1) = \psi^1(0)$. Supposing the solution to this equation exists and is unique, give a representation of the solution in terms of expected values of a stopped process.

Exercise 17.5.5. For N a Poisson count process with rate λ, find the infinitesimal generator $\mathcal{L}$ of $X_t := (N_t - \lambda t)/\sqrt{\lambda t}$.

Exercise 17.5.6. Let X be the solution to an equation of the form (17.2). Let $\alpha < \beta$ be constants. Suppose the PDE

$$\begin{aligned} \mathcal{L}v + \frac{\partial v}{\partial t} &= -1, && \text{on }]0, \infty[\times]\alpha, \beta[, \\ v(t, x) &= 0, && \text{for all } x \notin [\alpha, \beta], \end{aligned}$$

has a unique, bounded C^2 solution for all time. Show that

$$v(0, x) = E\big[\inf\{t : X_t \notin [\alpha, \beta]\}\big|X_0 = x\big].$$

18

Weak Solutions of SDEs

So far, we have focussed on solutions of SDEs where we are simply given a filtration, and with it the Brownian motion W and the random measure μ. We then construct the solution to our equation (17.2). In essence, we have used no properties of the filtration except the fact that W and μ are adapted. As we shall see, there are occasions where this approach is insufficient, and we require that the filtration is slightly richer.

Example 18.0.1 (Tanaka's Equation). Consider the following simple SDE,

$$dX_t = \mathrm{sign}(X_t)dW_t; \qquad X_0 = 0,$$

where W is a Brownian motion, and $\mathrm{sign}(x) := x/|x|$ with the convention $0/0 := 1$. If we have a solution X to this equation, then it is easy to see that $-X$ is also a solution, so there is no hope that solutions are unique. We can also see, from Lévy's characterization of Brownian motion, that a solution X must be a Brownian motion, as its quadratic variation is $\langle X\rangle_t = \int_{[0,t]} \mathrm{sign}(X_t)^2 dt = t$. Furthermore, solutions do exist; if we have a Brownian motion B in its natural filtration, and we take W to be the Brownian motion defined by $W = \mathrm{sign}(B)\bullet B$, then $X = B$ is a solution to the equation.

Suppose we have a solution to this equation, in the space with filtration $\{\mathcal{F}_t\}_{t\geq 0}$. Applying the Tanaka–Meyer–Itô formula[1] to X, we obtain $d|X_t| = dW_t + dL_t$, so $W = |X| - L$ for L the local time of X at zero. As L is adapted to the subfiltration $\{\mathcal{F}_t^{|X|}\}_{t\geq 0}$ generated by $|X|$ (Lemma 14.3.5), we see that W

[1] We do not need to be concerned by the convention $0/0 := 1$ in this setting, even though this conflicts with the convention given in the Tanaka–Meyer–Itô formula. This is because the set $\{X_t = 0\}$ is almost surely Lebesgue-null, and hence null in $L^2(W)$, so the value assigned to sign(0) does not change the stochastic integral. However, using this convention prevents $X_0 = 0$ from being a trivial solution to the SDE.

S.N. Cohen, R.J. Elliott, *Stochastic Calculus and Applications*,
Probability and Its Applications, DOI 10.1007/978-1-4939-2867-5_18

is also adapted to $\{\mathcal{F}_t^{|X|}\}_{t\geq 0}$. Consequently, the filtration $\{\mathcal{F}_t^W\}_{t\geq 0}$ generated by W is a subfiltration of $\{\mathcal{F}_t^{|X|}\}_{t\geq 0}$, and we observe that the sign of X cannot be adapted to $\{\mathcal{F}_t^W\}_{t\geq 0}$.

It follows that, given a Brownian motion W *in its natural filtration*, there exists no adapted solution to the SDE. Therefore, we see that the existence of a solution to this equation depends in a rather delicate way on the choice of filtration.

To give a more precise understanding of these differences, we define two different notions of a solution to an SDE of a similar type to that considered in Chapter 17:

$$\begin{cases} dX_t = f(\omega, t, X)dt + \sigma(\omega, t, X)dW_t + \displaystyle\int_{\mathbb{R}^d} \zeta\, \tilde{\mu}(d\zeta, dt), \\ X_s = x \in \mathbb{R}^d \end{cases} \tag{18.1}$$

where we also require $\tilde{\mu}$ to be the martingale random measure $\mu - \mu_p$ with

$$\mu_p(d\zeta, dt) = \nu(\omega, t, X; d\zeta)dt, \tag{18.2}$$

so the compensator of the random measure (or 'rate' of the jumps) depends on the state X. The Brownian motion and martingale random measure are to be constructed as part of the solution. We shall see that we are looking to construct a semimartingale X, in some probability space, with characteristics given by (B, C, μ_p), where μ_p is as in (18.2),

$$dB_t = f(\omega, t, X)dt \qquad \text{and} \qquad dC_t = \sigma\sigma^\top(\omega, t, X)dt.$$

Here f and σ are sufficiently integrable and measurable processes that this equation has a meaning. To ensure that X is a semimartingale with integrable jumps (in particular, that its small jumps are square summable), we assume throughout this chapter that ν is such that, for any t, X,

$$\int_{\mathbb{R}^d} (\|\zeta\|^2 \wedge \|\zeta\|)\nu(\omega, t, X; d\zeta) < \infty.$$

For simplicity, we will make an abuse of notation and write

$$(\zeta * \tilde{\mu})_t = \int_{[0,t]\times\mathcal{Z}} \zeta\, \tilde{\mu}(d\zeta, ds).$$

Remark 18.0.2. There is a close connection between the type of equation considered here and in Chapter 17. Comparing with (17.2), we no longer have a term $g(\zeta, t, X) * \tilde{\mu}$, in our equation, or rather, we assume that $g(\zeta) = \zeta$ and $\mathcal{Z} = \mathbb{R}^d$. This corresponds to taking $\mu = \mu^X$, in the sense of Example 13.4.1, so that $\Delta X_t = \int_{\mathbb{R}^d} \zeta\, \mu(d\zeta, \{t\})$. Clearly, if we assume that X is continuous (so $g \equiv 0$ in (17.2), and $\nu \equiv 0$ in (18.1)), they are the same equation. More generally, the difference is how jumps are incorporated.

Suppose we have $X = H + g * \tilde{\mu}$, where H is continuous and g is a $\tilde{\mu}$-stochastically integrable process. Define μ^X as in Example 13.4.1, so $\mu^X(dt \times \{x\}) = 1$ if and only if $\int_{\mathcal{Z}} g_s(\zeta)\mu(dt, d\zeta) = x$. Then

$$\mu^X(A \times dt) = \int_{\mathcal{Z}} I_{\{g_s(\zeta)\in A\}}\mu(d\zeta, dt)$$

and hence

$$\mu_p^X(A \times dt) = \int_{\mathcal{Z}} I_{\{g_s(\zeta)\in A\}}\mu_p(d\zeta, dt).$$

In particular,

$$\int_{\mathbb{R}^d} \zeta\, \tilde{\mu}^X(d\zeta, dt) = \int_{\mathcal{Z}} g_s(\zeta)\tilde{\mu}(d\zeta, dt).$$

If μ_p is of the form $\mu_p(d\zeta, dt) = \nu'(d\zeta)dt$, we have an equivalence between equations with a deterministic integrand and a stochastic compensator $\nu(\omega, t, X; d\zeta)$ (i.e. equations with a term $\zeta * \tilde{\mu}^X$) and those with a stochastic integrand and deterministic compensator $\nu'(d\zeta)$ (i.e. equations of the form $g(\cdot, X) * \tilde{\mu}$) whenever

$$\nu(\omega, t, X; A) := \int_{\mathcal{Z}} I_{\{g(t,X;\zeta)\in A\}}\nu'(d\zeta) \text{ for any } A \in \mathcal{B}(\mathbb{R}^d). \tag{18.3}$$

The key difference between these approaches comes down to what continuity we assume. In the '$g(X) * \tilde{\mu}$' setting, the assumption we needed was in terms of the continuity of g with respect to X. In the '$\zeta * \tilde{\mu}^X$' setting we consider here, we do not need continuity of the compensator with respect to X, but we *shall* require that all the compensators are absolutely continuous with respect to a reference measure on $\mathbb{R}^d$. It is easy to check that neither of these cases implies the other in general.

In the Markovian setting, when f, σ and ν depend on ω and X only through the value $X_t(\omega)$, solutions to the SDE (18.1) are often constructed to be Markovian, in which case they have an infinitesimal generator given by the integro-differential operator

$$\begin{aligned}\mathcal{L}_s v(x) = &\sum_i f^i(s,x)\frac{\partial v}{\partial x^i}(x) + \frac{1}{2}\sum_{i,j} a^{ij}(s,x)\,\frac{\partial^2 v}{\partial x^i \partial x^j}(s,x)\\ &+ \int_{\mathbb{R}^d}\Big(v(x+\zeta) - v(x) - \sum_i \frac{\partial v}{\partial x^i}(x)\zeta^i\Big)\nu(s,x;d\zeta)\,,\end{aligned} \tag{18.4}$$

where $a = \sigma\sigma^\top$ and ζ^i is the ith component of ζ. For $v \in C_b^2$, this can be checked in the same way as in Theorem 17.4.3 (namely, by applying Itô's formula). By simple change of variables arguments, this agrees with Theorem 17.4.3 whenever (18.3) holds.

Definition 18.0.3. *We say that a process X is a* strong solution *to the equation (18.1) if it solves (18.1) and is adapted to the (completed, right-continuous) filtration $\{\mathcal{F}_t^{W,\mu}\}_{t\geq 0}$ generated by the Brownian motion W and the random measure μ.*

The importance of the definition of a strong solution is that the filtration generated by W and μ is, in some sense, the minimal filtration in which the SDE can be defined. Consequently, if there exists a solution in this filtration, then there must exist a solution in any other filtration for which we consider our SDE.

Remark 18.0.4. The results of Chapters 16 and 17 do not rely on the choice of filtration. Therefore, it is clear that the solutions constructed are in a 'strong' sense. Consequently, we know that if b, σ and g are Lipschitz continuous and (18.3) holds for ν' some deterministic measure on $\mathcal{Z}$, then the equation admits a (unique) strong solution.

Definition 18.0.5. *We say that $(X, W, \mu, P, \{\mathcal{F}_t\}_{t\geq 0})$ is a* weak solution *to the SDE (18.1) if X satisfies (18.1), W is a P-Brownian motion, μ is a random measure with P-compensator $\mu_p(dt, d\zeta) = \nu(t, X; d\zeta)dt$, and X, W and μ are $\{\mathcal{F}_t\}_{t\geq 0}$ optional.*

If we are in the continuous case (i.e. $\nu \equiv 0$), we naturally omit μ from the definition of a weak solution. The point here is that W, P, μ and $\{\mathcal{F}_t\}_{t\geq 0}$ are constructed as part of the solution, rather than being prescribed in advance.

18.1 Modifying the Drift and Jumps

A key technique in the study of weak solutions is the use of Girsanov's theorem to transform the probability measure. This allows one to introduce drifts to the equation and to modify the rates of different jumps in μ. This technique allows us to directly construct weak solutions to many SDEs.

Consider a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$ satisfying the usual conditions, such that W is an N-dimensional Brownian motion ($N \leq \infty$) and μ a random measure with compensator $\nu(d\zeta)dt$, adapted to the filtration $\{\mathcal{F}_t\}_{t\geq 0}$. For σ, g predictable and of appropriate dimension, let

$$X_t = X_0 + \int_{[0,t]} \sigma_s dW_s + \int_{\mathcal{Z}\times[0,t]} g_s(\zeta)\tilde{\mu}(d\zeta, ds).$$

Suppose α, β are predictable and of appropriate dimension, $\beta > 0$ and that $\mathcal{E}(\alpha \bullet W + (\beta - 1) * \tilde{\mu})$ is a uniformly integrable martingale. Then, under the measure Q defined by $dQ/dP = \mathcal{E}(\alpha \bullet W + (\beta - 1) * \tilde{\mu})_\infty$, we know that X can be written

$$X_t = X_0 + \int_{[0,t]} \Big(\sigma_s \alpha_s + \int_{\mathcal{Z}} g_s(\zeta)\beta_s(\zeta)\nu(d\zeta)\Big) ds + \int_{[0,t]} \sigma_s dW_s^Q$$
$$+ \int_{\mathcal{Z}\times[0,t]} g_s(\zeta)\tilde{\mu}^Q(d\zeta, ds)$$

where

$$dW_s^Q = dW_s - \alpha_s ds,$$
$$\tilde{\mu}^Q(ds, d\zeta) = \mu(ds, d\zeta) - \mu_p^Q(ds, d\zeta) = \mu(ds, d\zeta) - \beta_s(\zeta)\nu(d\zeta)ds.$$

Applying Girsanov's theorem (in the form of Corollaries 15.3.4 and 15.3.7), we see that W^Q is a Q-Brownian motion and μ_p^Q is the compensator of μ under Q.

If we need only to determine the drift and jump measure of our process, then this discussion immediately implies the following general result.

Theorem 18.1.1. *Let σ, $\{\mathcal{F}\}_{t\geq 0}$, P, W, μ and ν be as described above, with the additional assumption $\mathcal{Z} = \mathbb{R}^d$ and $\|\zeta\| \wedge 1 \in L^2(\nu)$. For $x \in \mathbb{R}^d$, define*

$$X = x + \sigma \bullet W + \zeta * \tilde{\mu}.$$

Let $f : [0, \infty[\, \times \mathcal{D} \to \mathbb{R}^d$ and $\hat{\nu} : [0, \infty[\, \times \mathcal{D} \times \mathfrak{Z} \to \mathbb{R}^+$ (countably additive in $\mathfrak{Z}$) be such that there exist predictable α, β of appropriate dimension with

$$f(\omega, t, X) = \sigma_s \alpha_s + \int_{\mathcal{Z}} \beta_s(\zeta)\zeta\, \nu(d\zeta),$$
$$\hat{\nu}(\omega, t, X, d\zeta) = \beta_s(\zeta)\nu(d\zeta).$$

Suppose $\mathcal{E}(\alpha \bullet W + (\beta - 1) * \tilde{\mu})$ is a uniformly integrable martingale. Then $(X, W^Q, \mu, Q, \{\mathcal{F}_t\}_{t\geq 0})$ is a weak solution to the SDE

$$dX_t = f(t, X)dt + \sigma_t dW_t + \int_{\mathbb{R}^d} \zeta\, \tilde{\mu}^X(d\zeta, dt)$$

with $\mu_p^X(d\zeta, dt) = \hat{\nu}(\omega, t, X, d\zeta)dt$, and $X_0 = x$.

Remark 18.1.2. It is important to see that a change of measure only allows us to modify the drift and the compensator of the jump measure, and leaves the volatility σ unaltered.

Remark 18.1.3. If we take σ and g to be predictable Lipschitz functions of X, then applying this argument to $X = \sigma(X) \bullet W + g * \tilde{\mu}$ (which we know has solutions) gives us a solution (in an appropriate weak sense) to the more general equation

$$dX_t = f(t, X)dt + \sigma(t, X)dW_t + \int_{\mathcal{Z}} g(t, X, \zeta)\tilde{\mu}(d\zeta, dt)$$

where

$$\tilde{\mu}(d\zeta, dt) = \mu(d\zeta, dt) - \hat{\nu}(\omega, t, X; d\zeta)dt.$$

This approach to solving SDEs begs the question: 'Under what conditions can we be sure that $\mathcal{E}(\alpha\bullet W+(\beta-1)*\tilde{\mu})$ is a uniformly integrable martingale?' Of course, this is a question we addressed in some detail in Chapter 15. We can, therefore, obtain the following corollaries. The following is an extension of an argument due to Beneš [11], for the setting without jumps.

Corollary 18.1.4. *Consider the equation*

$$dX_t = f(t,X)dt + \sigma(t,X)dW_t \tag{18.5}$$

with $X_0 = x \in \mathbb{R}^d$*, where* W *is an* $N \leq \infty$ *dimensional Brownian motion. Suppose*

- *for some* $K > 0$*, we know* $|f(\omega,t,X)| \leq K(1+X_t^*)$*,*
- $\sigma(t,X)$ *has a right inverse* $\sigma(t,X)^{-1}$ *(as a matrix, which implies* $N \geq d$*), which is measurable in* (t,X)*, and both* σ *and* σ^{-1} *are uniformly bounded in* X *and on compacts in* t*, that is, for each* $T > 0$ *there exists* $K > 0$ *such that*
$$\sup_{t\in[0,T]}\sup_{X\in\mathcal{D}}\left\{\|\sigma(t,X)\| + \|\sigma(t,X)^{-1}\|\right\} \leq K,$$
- $\sigma(t,X)$ *is Lipschitz continuous and non-anticipative in* X*, that is, for some* $K > 0$*,*
$$\|\sigma(t,X) - \sigma(t,X')\| \leq K(X-X')_t^*.$$

Then (18.5) *admits a weak solution on* $[0,T]$*, for any finite time* T*.*

Proof. Let X be the solution to the SDE $X = x + \sigma(t,X)\bullet W$ in the filtration generated by an N-dimensional Brownian motion W. This exists as σ is Lipschitz continuous[2]. By Example 15.5.6 and Lemma 15.5.7, we know $\mathcal{E}\big((\sigma(t,X)^{-1}f(t,X))\bullet W\big)$ is a uniformly integrable martingale. Therefore, we can define a probability measure Q by $dQ/dP = \mathcal{E}\big((\sigma(t,X)^{-1}f(t,X))\bullet W\big)_T$. As $W_t^Q = W_t - \int_{[0,t]}\sigma(s,X)^{-1}f(s,X)ds$ is a Q-Brownian motion, we then see that X satisfies the equation

$$dX = f(t,X)dt + \sigma(t,X)dW^Q, \qquad X_0 = x$$

as desired. Therefore, $(X, W^Q, Q, \{\mathcal{F}_t^W\}_{t\geq 0})$ is a weak solution to (18.5). □

For equations with bounded jumps and volatility, we can obtain the following, more general, result.

[2]While Theorem 16.3.11 is in a finite dimensional context, if $N=\infty$ then the extension to infinitely many Brownian motions is relatively straightforward given a sufficiently strong integrability assumption, as we saw in Lemma 17.1.1. For example, it is sufficient that $\|\sigma(t,X)\| + \|\sigma(t,X)^{-1}\|$ is bounded, where $\|\sigma\|^2 = \mathrm{Tr}(\sigma\sigma^\top) = \sum_{i=1}^d\sum_{j\in\mathbb{N}}\sigma_{ij}^2$.

Corollary 18.1.5. *Consider the equation*

$$dX_t = f(t,X)dt + \sigma(t,X)dW_t + \int_{\mathbb{R}^d} \zeta\, \tilde{\mu}^X(d\zeta, dt)$$

with $X_0 = x \in \mathbb{R}^d$, where W is an N dimensional Brownian motion ($N \leq \infty$) and $\mu_p^X(d\zeta, dt) = \nu(t, X; d\zeta)dt$. Suppose that, for some $K > 0$,

- *$f(\omega, t, X) \leq K(1 + (X_t^*)^{1/2})$,*
- *for some compactly supported deterministic measure ν' on $\mathbb{R}^d$,*
  - *for any (t, X), we know $\nu(t, X; \cdot)$ and ν' are equivalent measures, and*
  - *writing $\beta_t = d\nu(t, X; \cdot)/d\nu'$ we have, for any X, dt-a.e.,*

$$\int_{\mathbb{R}^d} \big(\|\zeta\|^2 \beta_t^2(\zeta) + (\beta_t(\zeta) - 1)^2\big)\nu'(d\zeta) \leq K(1 + (X_t^*)^{1/2}),$$

- *$\sigma(t, X)$ has a right inverse $\sigma(t, X)^{-1}$ (as a matrix, which implies $N \geq d$), which is measurable in (t, X), and both σ and σ^{-1} are uniformly bounded in X and on compacts in t, and*
- *$\sigma(t, X)$ is Lipschitz continuous and non-anticipative in X, that is, for some $K > 0$, for all paths X and X',*

$$\|\sigma(t, X) - \sigma(t, X')\| \leq K(X - X')_t^*.$$

Then this equation admits a weak solution on $[0, T]$, for any finite time T.

Proof. Essentially, this follows in the same way as Corollary 18.1.4. Begin with a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$ where μ is a random measure on $\mathbb{R}^d$ with compensator $\nu'(d\zeta)dt$ and W is an N-dimensional Brownian motion. Define

$$X_t = x + \sigma(t, X)\bullet W_t + \int_{\mathbb{R}^d} \zeta\, \tilde{\mu}(d\zeta, dt), \tag{18.6}$$

which exists as σ is Lipschitz continuous and non-anticipative.

As σ is bounded and ν is compactly supported, by Lemma 15.5.12 we know $E[e^{a\|X_t\|}] < \infty$ for every $a > 0$ and $t > 0$.

We seek to change the measure so that X has the desired dynamics. Let β be a predictable version of the Radon–Nikodym derivative $d\nu(t, X; \cdot)/d\nu'$. Define

$$\alpha_t = \sigma(t, X)^{-1}\Big(f(t, X) - \int_{\mathbb{R}^d} \zeta\beta_t(\zeta)\nu'(d\zeta)\Big).$$

We then observe that, for some $K > 0$,

$$\|\alpha_t\|^2 + \int_{\mathbb{R}^d} (\beta_t(\zeta) - 1)^2 \nu'(d\zeta) \leq K(1 + (X_t^*)^{1/2}).$$

Therefore, by Example 15.5.9, we know that $\mathcal{E}(\alpha\bullet W + (\beta - 1) * \tilde{\mu})$ is a true martingale. The result follows by Girsanov's theorem, as in the previous theorem. □

Remark 18.1.6. Corollary 18.1.5 does not include the case where f and ν are of linear growth. However, a Lipschitz continuous linear growth term can be added without difficulty, by including it in (18.6).

Example 18.1.7 (Tsirel'son's SDE). Fix a decreasing sequence $t_0 = 1 > t_1 > t_2 > \ldots$ such that $\inf_n t_n = 0$. For x a continuous path, define

$$f(t,x) = \begin{cases} \frac{x_{t_j} - x_{j+1}}{t_j - t_{j+1}} \bmod 1 & \text{if } t \in \,]t_j, t_{j-1}] \\ 0 & \text{if } t = 0 \end{cases}$$

where $y \bmod 1$ denotes the fractional part of y. Note that f is then a nonanticipative, bounded, measurable function.

Let $X = W$ be a P-Brownian motion in its natural filtration $\{\mathcal{F}_t\}_{t\geq 0}$ (satisfying the usual conditions), and using Corollary 18.1.4, define the measure Q under which X satisfies the SDE

$$dX = f(t,X)dt + dW$$

for W a Q-Brownian motion.

One can then show (see Tsirel'son [178] or the presentation in Rogers and Williams [159, V.18] or Revuz and Yor [155, p.392]) the surprising result that for any solution of the equation $dX = f(t,X)dt + dW$, the drift $f(t,X)$ is independent of the filtration generated by W. In particular, W does not generate the same filtration as X. Therefore, like Tanaka's equation, this only admits weak solutions.

18.2 Determining the Volatility

We have seen how it is possible to modify the drift and jump compensator characteristics of a process X using change of measure techniques.

The key remaining term which causes us difficulty is the volatility σ. If σ is a bounded Lipschitz function of X (but may depend on the whole path of X in a nonanticipative way), then we know that it is possible to solve (18.1), by Corollary 18.1.5. If σ is locally Lipschitz, then the result is again possible, up to the first explosion time of the solution. We also would like to relax the conditions on the jump measure required by Corollary 18.1.5.

It is possible to construct solutions to these equations under fairly weak conditions. The key general result in this area is due to Stroock and Varadhan (see [174]) in the finite dimensional diffusion case (i.e. when the jump term is zero). See also the presentation in Rogers and Williams [159]. Lepeltier and Marchal [123] and Jacod [107] study weak solutions of equations of the form (17.2), and we shall here outline the results of Stroock [173].

Definition 18.2.1. *Let $\mathcal{L}$ be the time-dependent integro-differential operator defined in (18.4). The* martingale problem *starting from (t,x) associated with $\mathcal{L}$ asks: "Can we find a measure P on the space of càdlàg paths $\mathcal{D}([0,\infty[,\mathbb{R}^d) = \Omega$ with canonical element $X_t(\omega) = \omega_t$, such that $P(X_t = x) = 1$ and*

$$v(X_s) - \int_{]t,s]} \mathcal{L}_u v(X_u)du$$

is a P-martingale, for all $v \in C_0^\infty(\mathbb{R}^d)$ (where $C_0^\infty(\mathbb{R}^d)$ denotes the smooth functions which vanish at infinity)?" A solution to the martingale problem is a measure with this property.

The idea is that we can solve the martingale problem using purely analytic techniques, rather than Itô calculus. Given a solution to the martingale problem, we can then extract a Brownian motion and a compensated jump measure such that X solves the SDE (18.1). This extraction is the purpose of the next lemma.

Lemma 18.2.2. *Let P be a solution to the martingale problem, and X is the canonical element of $\mathcal{D}([0,\infty[,\mathbb{R}^d) = \Omega$ (that is, $X_t(\omega) = \omega_t$). Suppose σ is a square matrix and invertible. Then there exists a P-Brownian motion W and a jump measure μ^X (with the desired compensator) such that*

$$dX_t = \sigma dW_t + \int_{\mathbb{R}^d} \zeta \tilde{\mu}^X(d\zeta, dt).$$

Proof. We give a sketch of the proof. For any $v \in C_0^\infty(\mathbb{R}^d)$, in particular for $v(x) = x_i e^{-\epsilon x_i^2}$, we know $v(X_s) - \int_{]t,s]} \mathcal{L}_u v(X_u)du$ is a martingale. Expanding, taking the limit $\epsilon \to 0$ and using dominated convergence (given the bounds on ν), we see that

$$X_t - \int_{]0,t]} f(u, X_u)du$$

defines a vector valued local martingale (where f is the 'drift' term appearing in the generator $\mathcal{L}$), so X is a special semimartingale. Similarly,

$$(X_i X_j)_t - \frac{1}{2}\int_{]0,t]} a^{ij}(u, X_u)du + \int_{]0,t]\times\mathbb{R}^d} \zeta^i \zeta^j \nu(u, X_u; d\zeta)du$$

defines a local martingale, for each i, j, and so, recalling $a = \sigma\sigma^\top$, we know

$$d\langle X\rangle_s/ds = \sigma(s, X_s)\sigma(s, X_s)^\top + \int_{\mathbb{R}^d} \zeta\zeta^\top \nu(s, X_s; d\zeta).$$

Again using a similar argument, for any x and any compact set $A \subset \mathbb{R}^d$ with $x \notin A$, as $P(X_t = x) = 1$ we know

$$(X_i X_j)_t I_{\{X\in A\}} - \int_{]0,t]\times\mathbb{R}^d} \zeta^i \zeta^j I_{\{\zeta\in A\}} \nu(u, X_u; d\zeta)$$

defines a martingale on $[\![0, \tau]\!]$, where $\tau = \inf\{s > t : \|X_s - a\| < \epsilon$ for some $a \in A\}$. On $[\![0, \tau]\!]$, this process can only change by jumps, so simple calculations show that

$$d\langle X^d \rangle_s / ds = \int_{\mathbb{R}^d} \zeta\zeta^\top \nu(s, X_s; d\zeta),$$

that is, $\nu(s, X_s; \cdot)ds$ is the compensator of μ^X, as defined by Example 13.4.1. Consequently, $d\langle X^c \rangle_s / ds = \sigma(s, X_s)\sigma(s, X_s)^\top$. Taking

$$W_s = (\sigma(s, X))^{-1} \bullet \Big(X_s - \int_{]0,t]} f(u, X_u)du - \int_{]0,t] \times \mathbb{R}^d} \zeta \tilde{\mu}^X(d\zeta, dt) \Big),$$

we can see that W is a continuous martingale with quadratic variation equal to the identity matrix. By Lévy's characterization, we see that W is a Brownian motion. □

Stroock [173] gives the following result, which we state without proof.

Theorem 18.2.3. *Suppose*

- $a : [0, \infty[\times \mathbb{R}^d \to \mathbb{R}^{d \times d}$ *is continuous, bounded, and* $a(t, x)$ *is strictly positive definite for each* (t, x),
- $b : [0, \infty[\times \mathbb{R}^d \to \mathbb{R}^d$ *is bounded and measurable,*
- $\int_A \frac{\zeta}{1+\|\zeta\|^2} \nu(s, X; d\zeta)$ *is bounded and continuous for all* $A \in \mathcal{B}(\mathbb{R}^d \setminus \{0\})$.

Then the martingale problem for $\mathcal{L}$ *is well posed, that is, for each* (t, x) *there is exactly one measure* P *which satisfies the martingale problem.*

Remark 18.2.4. We can see that, if there is a unique solution P to the martingale problem, then this is the unique measure such that the processes W and $\tilde{\mu}(\cdot, A)$ (for $A \in \mathcal{B}(\mathbb{R}^d \setminus \{0\})$) are martingales, where W and $\tilde{\mu}$ are as constructed in Lemma 18.2.2. The significance of this will become apparent in the coming section.

18.3 The Jacod–Yor Theorem

In the previous sections, we have been constructing a Brownian motion and a jump measure as part of the solution of an SDE, and we have seen that they may not generate the filtration to which our SDE solution is adapted. This then raises the question, given a construction of this type, whether it is possible to show that we have a martingale representation theorem with respect to these processes. This leads us naturally to the following general result, due to Jacod and Yor [109] (see also Jacod [107] and Protter [152]), which gives necessary and sufficient conditions under which a martingale representation theorem will hold.

Example 18.3.1. Consider the solution to Tanaka's SDE (Example 18.0.1), $dX_t = \mathrm{sign}(X_t)dW$. Given that the process X is a Brownian motion generating the filtration $\{\mathcal{F}_t\}_{t\geq 0}$, we know that there is a martingale representation theorem with respect to X (Theorem 14.5.1). Therefore, for any $\{\mathcal{F}_t\}_{t\geq 0}$-local martingale M, there exists a predictable H such that $M = M_0 + H \bullet X$. Taking $Z_t = H_t\mathrm{sign}(X_t)$, we see that $M = M_0 + Z \bullet W$, and so we have a martingale representation with respect to W, even though W does *not* generate the filtration $\{\mathcal{F}_t\}_{t\geq 0}$.

Definition 18.3.2. *Let $\mathcal{N} \subset \mathcal{H}^1_{0,\mathrm{loc}}$. We define the p-stable subspace generated by $\mathcal{N}$, denoted $\mathcal{I}^p(\mathcal{N})$, to be the smallest closed subspace of $\mathcal{H}^p$ which contains the constants and the stochastic integrals $H \bullet M$ for all (finite-dimensional vector) processes M with components in $\mathcal{N}$ and H predictable such that $E[(H^2 \bullet [M])^{p/2}] < \infty$.*

Note that these spaces always depend on the choice of measure P, both through the integrability requirements, and because they are defined as equivalence classes under equality P-a.s. For this reason, it may be preferable to write $\mathcal{I}^p(\mathcal{N}, P)$ and $\mathcal{H}^p(P)$ to avoid confusion.

It is easy to see that $\mathcal{I}^p(\mathcal{N})$ is stable under stopping, so $\mathcal{I}^2(\mathcal{N})$ is a stable subspace in the sense of Definition 10.1.18. Extending this definition to $\mathcal{H}^p$ in the natural way, we see that $\mathcal{I}^p(\mathcal{N})$ is a stable subspace of $\mathcal{H}^p$. This naturally leads us to consider the space of orthogonal martingales as a subspace of the dual of $\mathcal{H}^p$. For $p \in]1, \infty[$, we showed (Remark 10.1.12) that this dual space is equivalent to $\mathcal{H}^q$, for $p^{-1} + q^{-1} = 1$. The case $p = 1$ was left unconsidered, but is treated in Appendix A.8, in particular Theorem A.8.14. The dual of $\mathcal{H}^1$ is the space of 'BMO' martingales $\mathcal{H}^{\mathrm{BMO}}$. One property of this space we shall require is that all processes in $\mathcal{H}^{\mathrm{BMO}}$ are locally bounded (Lemma A.8.7).

Definition 18.3.3. *For $p \in [1, \infty[$ and a stable subspace $\mathcal{K}$ of $\mathcal{H}^p_0$, we define $\mathcal{K}^\perp$ to be the space of martingales $N \in \mathcal{H}^q_0$ such that $E[M_\infty N_\infty] = 0$ for all $M \in \mathcal{K}$, where $p^{-1} + q^{-1} = 1$ if $p > 1$, and $q = \mathrm{BMO}$ if $p = 1$.*

The key question is whether $\mathcal{I}^1(\mathcal{N}, P) = \mathcal{H}^1(P)$, which would indicate that any $\mathcal{H}^1(P)$ martingale can be represented by a stochastic integral with respect to processes in $\mathcal{N}$. The condition under which this is possible is given in terms of the measure P.

Definition 18.3.4. *Let $\mathfrak{P}(\mathcal{N})$ denote the set of all measures on $(\Omega, \mathcal{F})$ such that all elements of $\mathcal{N}$ are local martingales. A measure $P \in \mathfrak{P}(\mathcal{N})$ is* extremal *if, for any $Q, Q' \in \mathfrak{P}(\mathcal{N})$ and $\lambda \in]0, 1[$ such that $\lambda Q + (1-\lambda)Q' = P$, we have $Q = Q' = P$.*

Remark 18.3.5. We now see that the martingale problem (Definition 18.2.1) consists of showing that $\mathfrak{P}(\mathcal{N})$ is nonempty, where

$$\mathcal{N} = \Big\{ \Big\{ f(X_t) - \int_{[0,t]} \mathcal{L}_s f(X_s) ds \Big\}_{t\geq 0}, f \in C_0^\infty(\mathbb{R}^d) \Big\}.$$

Theorem 18.3.6 (Jacod–Yor Theorem). *Let $\mathcal{N}$ be a subset of $\mathcal{H}_0^1(P)$, and suppose $\mathcal{F} = \mathcal{F}_{\infty-}$. The following are equivalent.*

(i) $\mathcal{I}^1(\mathcal{N}, P) = \mathcal{H}^1(P)$ *and* $\mathcal{F}_0$ *is* P*-trivial*
(ii) $(\mathcal{I}^1(\mathcal{N}, P))^\perp$ *contains only the zero process (up to equality* P*-a.s.),*
(iii) P *is an extremal point of* $\mathfrak{P}(\mathcal{N})$*,*

Proof. That (i) is equivalent to (ii) is the result of the Hahn–Banach theorem, in the form of Corollary 1.5.14.

To show (iii) implies (ii), first observe that if $A \in \mathcal{F}_0$ is not P-trivial, then the probabilities $Q = P(A \cap \cdot)/P(A)$ and $Q' = P(A^c \cap \cdot)/P(A^c)$ are in $\mathfrak{P}(\mathcal{N})$, and $P = P(A)Q + (1 - P(A))Q'$, so P cannot be extremal. Therefore $\mathcal{F}_0$ must be P-trivial.

Let $M \in (\mathcal{I}^1(\mathcal{N}, P))^\perp \subset \mathcal{H}^{\mathrm{BMO}}$, so M is locally a bounded P-martingale with a localizing sequence $\{\tau_n\}_{n\in\mathbb{N}}$, such that $M_0 = 0$ and MN is a local P-martingale for all $N \in \mathcal{N}$. It follows that $M^{\tau_n}N$ is a local P-martingale, and, if k_n is a uniform bound on M^{τ_n}, then we can define measures Q^n, Q^{-n} equivalent to P by

$$\frac{dQ^n}{dP} = 1 + \frac{M^{\tau_n}}{2k_n}, \qquad \frac{dQ^{-n}}{dP} = 1 - \frac{M^{\tau_n}}{2k_n}.$$

As $M^{\tau_n}N$ is a local martingale, $\langle M^{\tau_n}, N\rangle = 0$, so by Girsanov's theorem, N is both a Q^n and Q^{-n} local martingale, for all $N \in \mathcal{N}$. Therefore, $Q^n, Q^{-n} \in \mathfrak{P}(\mathcal{N})$. However, $P = (Q^n + Q^{-n})/2$, so (iii) implies $Q^n = Q^{-n} = P$, that is, $M \equiv 0$ P-a.s. Therefore, $(\mathcal{I}^p(\mathcal{N}, P))^\perp$ contains only the zero process (up to equivalence P-a.s.).

To show (ii) implies (iii), let $P = \lambda Q + (1-\lambda)Q'$, for some $Q, Q' \in \mathfrak{P}(\mathcal{N})$ and some $\lambda \in \,]0,1[$. Without loss of generality, assume $\lambda \leq 1/2$, so we can write

$$\begin{aligned} P &= \alpha\Big(\Big(1 - \frac{\lambda}{2}\Big)Q + \frac{\lambda}{2}Q'\Big) + (1-\alpha)\Big(\frac{\lambda}{2}Q + \Big(1 - \frac{\lambda}{2}\Big)Q'\Big) \\ &=: \alpha Q_\lambda + (1-\alpha)Q'_\lambda \end{aligned} \tag{18.7}$$

where $\alpha = \lambda/(2-2\lambda) \in \,]0, 1/2]$. Easy calculation shows that the measures Q_λ and Q'_λ are both in $\mathfrak{P}(\mathcal{N})$, and are equivalent to P. Therefore, we can write

$$1 = \alpha\frac{dQ_\lambda}{dP} + (1-\alpha)\frac{dQ'_\lambda}{dP} \geq \alpha\frac{dQ_\lambda}{dP}.$$

It follows that the martingale Λ defined by $\Lambda_t = E[dQ_\lambda/dP|\mathcal{F}_t]$ is bounded above by $1/\alpha$ and below by 0. However, for any $N \in \mathcal{N}$, as Λ is the density with respect to P of a measure in $\mathfrak{P}(\mathcal{N})$, we know $(\Lambda - 1)N$ is a local P-martingale. Therefore $\Lambda - 1 \in (\mathcal{I}^p(\mathcal{N}, P))^\perp$, which by (ii) implies $\Lambda = 1$, that is, $Q_\lambda = Q'_\lambda = P$ and hence $P = Q = Q'$. □

The following extension holds in a general setting; we restrict ourselves to the case where $\mathcal{N}$ is finite for simplicity.

Corollary 18.3.7. *If $\mathcal{N}$ is a finite collection of processes, then the statements of Theorem 18.3.6 are equivalent to the statement*

(iv) if $Q \in \mathfrak{P}(\mathcal{N})$ is absolutely continuous with respect to P, then $Q = P$.

Proof. If $\mathcal{N}$ is finite, then we recall that the space of vector stochastic integrals is closed in $\mathcal{S}$ (Theorem 12.5.16) and hence in the stronger topology of $\mathcal{H}^1$ (alternatively, this is easy to prove directly). Therefore, every element of $I(\mathcal{N}, P)$ can be written $M = M_0 + H^\top \bullet N$, for N the vector whose components are the elements of $\mathcal{N}$. By Theorem 18.3.6(i), we see that, for any $A \in \mathcal{F}_{\infty-}$, there exists H such that $I_A = P(A) + (H^\top \bullet N)_\infty$. Under Q, the process $H^\top \bullet N$ is well defined (as Q is absolutely continuous with respect to P) and is a bounded Q-σ-martingale. All bounded σ-martingales are local martingales of class (D) (Corollary 12.3.20), and hence true martingales (Lemma 5.6.6). It follows that

$$Q(A) = E^Q[I_A] = E^Q[P(A) + (H^\top \bullet N)_\infty] = P(A).$$

As A was arbitrary, we conclude that $Q = P$.

Conversely, if P is not extremal in $\mathfrak{P}(\mathcal{N})$ (so the conditions of Theorem 18.3.6 do not hold), then $P = \lambda Q + (1-\lambda)Q'$. Taking Q_λ as in (18.7), we see that Q_λ is equivalent to P and is an element of $\mathfrak{P}(\mathcal{N})$, so (iv) cannot hold. □

Remark 18.3.8. This statement is not typically stated as part of the Jacod–Yor theorem; however, the result arises commonly in mathematical finance, where it is closely related to the 'Second Fundamental Theorem of Asset Pricing', see Delbaen and Schachermayer [51]. In the setting where $\mathcal{N}$ is infinite, similar results are possible, but our proof would require an appropriate infinite-dimensional vector stochastic integral, see Mikulevicius and Rozovskii [137] for a construction, and De Donno, Guasoni and Pratelli [50] for further discussion of related issues in mathematical finance.

The following corollary covers the case considered in Theorem 18.2.3.

Corollary 18.3.9. *If the solution to the martingale problem for $\mathcal{L}$ is unique (where $\mathcal{L}$ is as in (18.4)) and the conditions of Lemma 18.2.2 hold, then the Brownian motion W and the random measure $\tilde{\mu}$ constructed in Lemma 18.2.2 satisfy $\mathcal{I}^1(\{W, \tilde{\mu}(\cdot, A) : A \in \mathcal{B}(\mathbb{R}^d \setminus \{0\}\}) = \mathcal{H}^1$. Hence the pair $(W, \tilde{\mu})$ has the predictable representation property in $(\{\mathcal{F}_t\}_{t\geq 0}, P)$.*

Proof. Suppose there are two measures P and Q, which make the processes $\mathcal{N} = \{W, \tilde{\mu}(\cdot, A) : A \in \mathcal{B}(\mathbb{R}^d \setminus \{0\}\}$ into local martingales. Then by applying Itô's lemma to the process $v(X)$, where X satisfies (18.1) and $v \in C_0^\infty$, we observe that these are both solutions to the martingale problem for $\mathcal{L}$. By uniqueness $P = Q$, so there is a unique measure in $\mathfrak{P}(\mathcal{N})$, and the martingale representation theorem follows. To convert stochastic integrals with respect to the processes $\tilde{\mu}(\cdot, A)$ into a stochastic integral with respect to the random measure $\tilde{\mu}$, first use the fact that $\mu \in \tilde{\mathcal{A}}_\sigma^1$ so sums of integrals with respect to the processes $\tilde{\mu}(\cdot, A)$ correspond to integrals of simple integrands with respect to $\tilde{\mu}$. The result follows from density of the simple integrands. □

18.4 Exercises

Exercise 18.4.1. A *scale function* of a process X is a C^2 function such that $s(X)$ is a local martingale.

(i) If $dX_t = \mu(X_t)dt + \sigma(X_t)dW_t$, for W a Brownian motion, show that a scale function s should satisfy the differential equation

$$\mu(x)\frac{\partial s}{\partial x} + \frac{\sigma^2(x)}{2}\frac{\partial^2 s}{\partial x^2} = 0$$

and hence is given by

$$s(x) = \int_{]c,x]} \exp\Big(-2\int_{]c,y]} \frac{\mu(z)}{\sigma^2(z)}dz\Big)dy$$

for any $c \in \mathbb{R}$.

(ii) Conversely, suppose μ and σ are such that s is well defined, invertible and differentiable. Define $g(y) = s'(s^{-1}(y))\sigma(s^{-1}(y))$ and $\gamma(t) = \int_{]0,t]} g^{-2}(B_u)du$ where B is a Brownian motion, and suppose γ is invertible. Then show that $X_t = s^{-1}(B_{\gamma^{-1}(t)})$ defines a weak solution to the equation $dX_t = \mu(X_t)dt + \sigma(X_t)dW_t$.

Exercise 18.4.2. Consider Tanaka's equation, in a given filtration $\{\mathcal{F}_t\}_{t\geq 0}$ generated by a Brownian motion B. Describe all (weak) solutions to the equation which are adapted to this filtration, that is, all pairs (X, W) adapted to the filtration, with $dX = \operatorname{sign}(X_{t-})dW$, such that W is a Brownian motion.

Exercise 18.4.3. Using a measure change argument, construct a jump process N with jumps of size one, such that jumps occur at a rate λ satisfying the SDE

$$d\lambda_t = -\alpha\lambda_t dt + \beta dN_t.$$

(This is a special case of a "Hawkes' process".)

Exercise 18.4.4. Consider the process $X_t = e^{W_t} \bullet B_t$, where W and B are independent Brownian motions. Show that both W and B are adapted to the filtration generated by X, but that the space of integrals satisfies $\{H \bullet X\}_{H\in L^1(X)} = \{H \bullet B\}_{H\in L^1(B)}$, and hence X does not have the martingale representation property. Verify directly that the statement of Corollary 18.3.7 does not hold.

Exercise 18.4.5. Let $\mathcal{Z} = \{1/n\}_{n\in\mathbb{N}}$. Let W be a Brownian motion in its natural filtration and μ be a random measure in $\tilde{\mathcal{A}}^1_\sigma$ such that, in the (right-continuous, complete) filtration generated by W and μ,

$$\mu_p(dt \times \{1/n\}) = ne^{W_t}dt.$$

Now consider the natural filtration $\{\mathcal{F}^\mu_t\}_{t\geq 0}$ of μ and let ν be the compensator of μ in this filtration.

(i) Show that W is adapted to the filtration generated by μ, so $\nu = \mu_p$
(ii) By considering the continuous and purely discontinuous martingales, show that not all $\{\mathcal{F}^\mu_t\}_{t\geq 0}$-martingales in this space can be written as stochastic integrals with respect to μ.
(iii) Show directly that the statement of Corollary 18.3.7 does not hold.

19

Backward Stochastic Differential Equations

In this chapter, we consider a different type of stochastic differential equation. In the setting of Chapter 17, we specified a solution process X through its dynamics and its initial value, as in (17.6). In this chapter, we specify a solution process Y through its dynamics and its *terminal* value, at a fixed, deterministic time $T \in]0, \infty[$. The difficulty with this is that the terminal value is allowed to be a random variable, but we look for a solution which is adapted to a given filtration.

Example 19.0.1. To see why this is problematic, consider the filtration generated by a Brownian motion W, a Borel function $\phi : \mathbb{R} \to \mathbb{R}$, and consider the equation on $[0, T]$ given by

$$dY_t = 0, \qquad Y_T = \phi(W_T).$$

Clearly, the only possible solution to the equation is to set $Y_t = \phi(W_T)$ for all t; however, this is not generally an adapted process.

On the other hand, we have seen, in the martingale representation theorem (Theorem 14.5.1), that if $E[\phi(W_T)^2] < \infty$, then there exists a predictable process Z such that

$$\phi(W_T) = E[\phi(W_T)] + (Z \bullet W)_T,$$

and so there exists a pair of processes (Y, Z) satisfying the equation

$$dY_t = Z_t dW_t, \qquad Y_T = \phi(W_T),$$

namely $Y_t = E[\phi(W_T)|\mathcal{F}_t] = E[\phi(W_T)] + (Z \bullet W)_t$. If we require that Y must be square integrable (that is, $Y \in S^2$), then we also observe that Y is unique (as its dynamics imply that it must be a local martingale, hence in $\mathcal{H}^2$, and so

S.N. Cohen, R.J. Elliott, *Stochastic Calculus and Applications*,
Probability and Its Applications, DOI 10.1007/978-1-4939-2867-5_19

must equal the conditional expectation of its terminal value). Conceptually, the process Z allows us to control the randomness of Y in such a way as to ensure that we hit the stochastic 'target' ξ.

In this chapter, we extend this example to include jumps, and to introduce a drift, which may depend in a nonlinear way on the values of Y and Z. We shall see that this also gives a nonlinear version of the Feynman–Kac theorem (Theorem 17.4.10), which connects solutions of BSDEs with semilinear PIDEs.

In the light of these remarks, we restrict our attention in this chapter to a setting where the martingale representation theorem holds. Based on Theorem 14.5.7, as in Chapter 17 we suppose we have

- a sequence of $N \leq \infty$ independent Brownian motions $W = \{W^1, W^2, \ldots\}$,
- a random measure $\mu \in \tilde{\mathcal{A}}^1_\sigma$ on a Blackwell space $(\mathcal{Z}, \mathfrak{Z})$, with deterministic compensator $\mu_p(d\zeta, dt) = \nu(d\zeta)dt$; we write $\tilde{\mu} = \mu - \mu_p$, and
- $\{\mathcal{F}_t\}_{t\geq 0}$ a complete, right-continuous filtration such that the pair $(W, \tilde{\mu})$ has the predictable representation property in $(\{\mathcal{F}_t\}_{t\geq 0}, P)$ (for example, the filtration generated by W and μ).

We use the same notation as in Chapter 17, in particular, we write $L^2(\nu)$ for the space of functions $\theta : \mathcal{Z} \to \mathbb{R}^m$ such that $\int_{\mathcal{Z}} \theta^2(\zeta)\nu(d\zeta) < \infty$ and $\|\theta\|^2_\nu = \int_{\mathcal{Z}} \theta^2(\zeta)\nu(d\zeta)$ and, if $N = \infty$, we identify $\mathbb{R}^N$ with ℓ_2 for notational convenience.

Remark 19.0.2. The fact that the compensator of μ is deterministic is unnecessary, but it simplifies notation. The important fact, as we shall see, is that the martingale representation theorem holds (i.e. W, μ have the predictable representation property), and the compensator of μ is a.s. absolutely continuous with respect to dt.

Definition 19.0.3. *Let $m < \infty$. Consider a function*

$$f : \Omega \times]0, T] \times \mathbb{R}^m \times \mathbb{R}^{m\times N} \times L^2(\nu) \to \mathbb{R}^m$$

such that $f(\omega, t, y, z, \theta)$ is progressively measurable in (ω, t) and Borel measurable in (y, z, θ).

For such a function f and an $\mathcal{F}_T$-measurable $\mathbb{R}^m$-valued random variable ξ, a Backward Stochastic Differential Equation (BSDE) is the equation

$$\begin{cases} dY_t = -f(\omega, t, Y_t, Z_t, \Theta_t)dt + Z_t dW_t + \displaystyle\int_{\zeta\in\mathcal{Z}} \Theta_t(\zeta)\tilde{\mu}(d\zeta, dt), \\ Y_T = \xi, \end{cases} \tag{19.1}$$

or, by integrating on $]t, T]$ and rearranging, we obtain the integrated form

$$Y_t + \int_{]t,T]} Z_s dW_s + \int_{\mathcal{Z}\times]t,T]} \Theta_s(\zeta)\tilde{\mu}(d\zeta, ds) = \xi + \int_{]t,T]} f(\omega, s, Y_s, Z_s, \Theta_s)ds. \tag{19.2}$$

We call f the driver *of the BSDE and the pair (ξ, f) the* data *of the BSDE.*

A solution to a BSDE is a triple of processes (Y, Z, Θ) *satisfying the above equation, such that* Y *is* $\mathbb{R}^m$*-valued, càdlàg and adapted,* Z *is* $\mathbb{R}^{m\times N}$*-valued and predictable and* Θ *is a* $\tilde{\mu}$*-stochastically integrable (and hence predictable) process taking values in* $L^2(\nu)$*.*

Remark 19.0.4. Note that a BSDE solution has a natural decomposition, into a predictable, finite variation, 'drift' part $f(\omega, t, Y_t, Z_t, \Theta_t)dt$, a continuous martingale part $Z_t dW_t$ and a pure-jump martingale part $\int_{\zeta\in\mathcal{Z}} \Theta_t(\zeta)\tilde{\mu}(d\zeta, dt)$.

Remark 19.0.5. The appearance in (19.1) of the martingale term $Z_t dW_t + \int_{\zeta\in\mathcal{Z}} \Theta_t(\zeta)\tilde{\mu}(d\zeta, dt)$ may seem complicated at first; however it is simply the term appearing in the martingale representation theorem. The processes Z and Θ represent the martingale part of Y in a sufficiently concrete way that they can appear in the drift term. Conceptually, we allow the drift at time t to depend on the 'amount of randomness' needed at t in order to ensure we hit our target ξ at time T.

If we consider the filtration generated by a single Brownian motion, the martingale terms would reduce to $Z_t dW_t$. Similarly, if we consider the filtration generated by a pure jump process, then the Brownian term disappears.

Remark 19.0.6. For future reference, we note that the optional quadratic variation of $\int_{[0,t]} Z_s dW_s + \int_{\mathcal{Z}\times[0,t]} \Theta_s(\zeta)\tilde{\mu}(d\zeta, ds)$ is given by

$$\int_{[0,t]} \|Z_t\|^2 dt + \int_{\mathcal{Z}\times[0,t]} \|\Theta_s(\zeta)\|^2 \mu(d\zeta, ds),$$

and the predictable quadratic variation by

$$\int_{[0,t]} \|Z_t\|^2 dt + \int_{\mathcal{Z}\times[0,t]} \|\Theta_s(\zeta)\|^2 \mu_p(d\zeta, ds) = \int_{[0,t]} (\|Z_t\|^2 + \|\Theta_s\|_\nu^2) dt,$$

where, as in (17.1), $\|\theta\|_\nu^2 = \int_{\mathcal{Z}} \theta(\zeta)^2 d\nu$.

19.1 Lipschitz BSDEs

In a similar way to how we approached the existence of solutions to SDEs in Section 16.1, we now consider solutions to BSDEs when the driver f satisfies a Lipschitz continuity assumption. In addition to the use of the martingale representation result, a key difference between this setting and those of the previous chapters is that we can no longer rely on localization arguments to assume that all our terms are sufficiently bounded – given the *terminal* value of the BSDE is what is prescribed, we need to work on the fixed interval $[0, T]$, rather than up to a stopping time.

This type of equation was first considered by Pardoux and Peng [147] in the continuous setting with a finite-dimensional Brownian motion. The continuous case with infinite dimensional noise was considered by Fuhrman and Hu [83]. With jumps, but still a finite dimensional Brownian motion, Tang and Li [175], Situ [168] and Barles, Buckdahn and Pardoux [3] considered a certain subclass of these equations, see also Royer [162]. An infinite dimensional approach with jumps was presented by the authors in [34], for a general filtration. The interested reader should also consult the review by El Karoui, Peng and Quenez [65], and the books of Delong [55] and Crépey [38].

Definition 19.1.1. *For the case of a Lipschitz driver, the following spaces naturally arise. Recall that $\mathcal{D}$ denotes the space of càdlàg adapted processes, and we defined, in Section 16.2, the space*

$$S^p = \{Y \in \mathcal{D} : \|Y\|_{S^p} := \|Y^*_\infty\|_{L^p} < \infty\},$$

where $Y^* = \max_i \sup_{s\le t} |Y^i_s|$, for $Y = (Y^1, Y^2, \dots, Y^m)$. We now also define

$$\begin{aligned} L^2(P\times t;\mathbb{R}^m) &= \Big\{Y\in\mathcal{D} : E\Big[\int_{[0,T]}\|Y_t\|^2dt\Big]<\infty\Big\},\\ L^2(\langle W\rangle;\mathbb{R}^{m\times N}) &= \Big\{Z \textit{ predictable} : E[\langle Z\bullet W\rangle_T] = E\Big[\int_{[0,T]}\|Z_t\|^2dt\Big]<\infty\Big\},\\ L^2(\langle\tilde\mu\rangle;L^2(\nu)) &= \Big\{\Theta : \Theta \textit{ is } \tilde\mu\textit{-stochastically integrable},\\ &\qquad E[\langle\Theta * \tilde\mu\rangle_T] = E\Big[\int_{[0,T]}\|\Theta_t\|^2_\nu dt\Big]<\infty\Big\}. \end{aligned}$$

These spaces are simply L^2 spaces under some measures; however, it is convenient to have a fixed notation. The term after the semicolon refers to the space in which the processes take values and will typically be omitted. However, we should be clear that a process in $L^2(\langle\tilde\mu\rangle)$ is a process taking values in $L^2(\nu)$. Using our earlier notation, we could equally write $L^2(W)$ for $L^2(\langle W\rangle)$, however this could be ambiguous when we write $L^2(\langle\tilde\mu\rangle)$, as we are here thinking about the *stochastic* integrals, rather than the space of pathwise integrals.

Remark 19.1.2. It is easy to see that

$$Z\bullet W + \Theta * \tilde\mu \in \mathcal{H}^2 \quad \text{if and only if} \quad (Z,\Theta)\in L^2(\langle W\rangle)\times L^2(\langle\tilde\mu\rangle)$$

(where $\mathcal{H}^2$ is the space of square integrable martingales).

Definition 19.1.3. *We say that the data (ξ, f) are* standard *if $E[\|\xi\|^2]<\infty$ and $E[\int_{[0,T]}\|f(\omega,t,0,0,0)\|^2dt]<\infty$, where 0 represents either a zero vector or function as appropriate. We say data are* standard Lipschitz *if, in addition, there exists a constant K such that, $dP\times dt$-a.e.,*

$$\|f(\omega,t,y,z,\theta) - f(\omega,t,y',z',\theta')\|^2 \le K\Big(\|y-y'\|^2 + \|z-z'\|^2 + \|\theta-\theta'\|^2_\nu\Big),$$

for all $y, y'\in\mathbb{R}^m$, $z,z'\in\mathbb{R}^{m\times N}$ and $\theta,\theta'\in L^2(\nu)$.

Lemma 19.1.4. *Let (Y, Z, Θ) be a solution of (19.1) with standard Lipschitz data (ξ, f). Suppose $Z \in L^2(\langle W \rangle)$ and $\Theta \in L^2(\langle \tilde{\mu} \rangle)$. Then $Y \in S^2$ if and only if $Y \in L^2(P \times t)$.*

Proof. If $Y \in S^p$ then

$$E\Big[\int_{[0,T]} \|Y_t\|^2 dt\Big] \leq TE[\sup_t \|Y_t\|^2] \leq Tm\|Y\|_{S^2} < \infty,$$

so $Y \in L^2(P \times t)$. Conversely, suppose $Y \in L^2(P \times t)$. Expanding the integrated form of (19.1) and using the Itô isometry, we see

$$\begin{aligned}
&\sup_t \|Y_t\|^2 \\
&\leq 4\|\xi\|^2 + 4\int_{[0,T]} \|f(\omega, s, Y_s, Z_s, \Theta_s)\|^2 ds \\
&\qquad + 4\sup_t \Big\{ \int_{]t,T]} Z_s dW_s \Big\} + 4 \sup_t \Big\{ \int_{\mathcal{Z}\times]t,T]} \Theta_s(\zeta)\tilde{\mu}(d\zeta, ds) \Big\} \\
&\leq 4\|\xi\|^2 + 4\int_{[0,T]} \|f(\omega, s, 0, 0, 0)\|^2 ds \\
&\qquad + 4K\int_{[0,T]} \Big(\|Y_t\|^2 + \|Z_t\|^2 + \int_{\mathcal{Z}} \|\Theta_t(\zeta)\|^2 \nu(d\zeta)\Big) dt \\
&\qquad + 4\sup_t \Big\{ \Big(\int_{]t,T]} Z_s dW_s \Big)^2 \Big\} + 4 \sup_t \Big\{ \Big(\int_{\mathcal{Z}\times]t,T]} \Theta_s(\zeta)\tilde{\mu}(d\zeta, ds) \Big)^2 \Big\}.
\end{aligned}$$

Taking an expectation and using Doob's L^2-inequality, we have

$$\begin{aligned}
E[\sup_t \|Y_t\|^2] &\leq E\Big[4\|\xi\|^2 + 4\int_{[0,T]} \|f(\omega, s, 0, 0, 0)\|^2 ds \Big] \\
&\qquad + 4(K+4)\int_{[0,T]} E\Big[\|Y_t\|^2 + \|Z_t\|^2 + \|\Theta_t\|_\nu^2 \Big] dt.
\end{aligned}$$

As $\|Y\|_{S^2} = E[\max_i \sup_{t\geq 0} |Y_t^i|^2]^{1/2} \leq E[\sup_{t\geq 0} \|Y_t\|^2]^{1/2}$, the result follows from the assumptions. □

Using the martingale representation theorem, we obtain the following simple existence result.

Lemma 19.1.5. *Let (ξ, f) be standard data, and suppose that $f(\omega, t, y, z, \theta)$ is independent of (y, z, θ). Then the BSDE (19.1) admits a unique solution $(Y, Z, \Theta) \in S^2 \times L^2(\langle W \rangle) \times L^2(\langle \tilde{\mu} \rangle)$.*

Proof. As we require $(Z,\Theta) \in L^2(\langle W\rangle) \times L^2(\langle\tilde{\mu}\rangle)$, we can take a conditional expectation in (19.2) to see that Y should be the càdlàg process uniquely defined by

$$Y_t = E\Big[\xi + \int_{]t,T]} f(\omega, s, 0, 0, 0)ds\Big|\mathcal{F}_t\Big].$$

Using the martingale representation theorem (Theorem 14.5.7), we can find unique processes $Z \in L^2(\langle W\rangle)$ and $\Theta \in L^2(\langle\tilde{\mu}\rangle)$ such that, for all t, we have

$$\begin{aligned} Y_t + \int_{[0,t]} f(\omega, s, 0, 0, 0)ds &= E\Big[\xi + \int_{[0,T]} f(\omega, s, 0, 0, 0)ds\Big|\mathcal{F}_t\Big] \\ &= E\Big[\xi + \int_{[0,T]} f(\omega, s, 0, 0, 0)ds\Big] \\ &\quad + \int_{[0,t]} Z_s dW_s + \int_{\mathcal{Z}\times[0,t]} \Theta_s(\zeta)\tilde{\mu}(d\zeta, ds). \end{aligned}$$

As $E\big[\|\xi\|^2 + \int_{[0,T]} \|f(\omega, s, 0, 0, 0)\|^2 ds\big] < \infty$, it is easy to show that $Y \in L^2(P \times t)$. Formally differentiating this equation, we obtain the required equality

$$dY_t + f(\omega, t, 0, 0, 0)dt = Z_t dW_t + \int_{\mathcal{Z}} \Theta_t(\zeta)\tilde{\mu}(d\zeta, dt)$$

and $Y_T = E[\xi|\mathcal{F}_T] = \xi$. By Lemma 19.1.4, we conclude that $Y \in S^2$, and the result is proven. $\square$

As for SDEs in Section 16.1, we now prove a useful stability estimate for solutions of BSDEs. Following a variation of the argument of El Karoui, Peng and Quenez [65], together with the simple existence lemma above, this allows us to construct solutions using a contraction mapping method.

Theorem 19.1.6. *Let (ξ^1, f^1) and (ξ^2, f^2) be standard Lipschitz data for two BSDEs, whose solutions are (Y^1, Z^1, Θ^1) and (Y^2, Z^2, Θ^2), both in $S^2 \times L^2(\langle W\rangle) \times L^2(\langle\tilde{\mu}\rangle)$. Let K be a Lipschitz constant for f^1, and write*

$$\begin{aligned} &\delta Y = Y^1 - Y^2, \quad \delta Z = Z^1 - Z^2, \quad \delta\Theta = \Theta^1 - \Theta^2 \\ \textit{and} \quad &\delta_2 f_t = f^1(\omega, t, Y_t^2, Z_t^2, \Theta_t^2) - f^2(\omega, t, Y_t^2, Z_t^2, \Theta_t^2). \end{aligned}$$

Then, for any $\beta \geq 4K + 1/2$,

$$\begin{aligned} e^{\beta s}E\big[\|\delta Y_s\|^2\big] &+ \frac{1}{2}\int_{]s,T]} e^{\beta t}E\big[\|\delta Z_t\|^2 + \|\delta\Theta_t\|_\nu^2\big]dt \\ &\leq E\Big[\|\xi^1 - \xi^2\|^2 + \frac{4}{2\beta - 1}\int_{]s,T]} e^{\beta t}\|\delta_2 f_t\|^2 dt\Big]. \end{aligned}$$

Proof. We first apply Itô's formula to the semimartingale $e^{\beta t}\|\delta Y_t\|^2$, for a given $\beta > 0$. As Y takes values in $\mathbb{R}^d$, we can write $\|\delta Y_t\|^2 = \delta Y_t^\top \delta Y_t$, with '$\top$' denoting vector transposition. In a differential form, we obtain

$$\begin{aligned}
&d(e^{\beta t}\|\delta Y_t\|^2) \\
&= \beta e^{\beta t}\|\delta Y_t\|^2 dt + 2e^{\beta t}\delta Y_t^\top d(\delta Y)_t + e^{\beta t}\operatorname{Tr}(d[\delta Y_t]) \\
&= \beta e^{\beta t}\|\delta Y_t\|^2 dt - 2e^{\beta t}\delta Y_t^\top (f(\omega,t,Y_t,Z_t,\Theta_t) - \tilde f(\omega,t,\tilde Y_t,\tilde Z_t,\tilde\Theta_t))dt \\
&\qquad + 2e^{\beta t}\delta Y_t^\top \delta Z_t dW_t + 2\int_{\mathcal{Z}} e^{\beta t}\delta Y_t^\top \delta\Theta_t(\zeta)\tilde\mu(d\zeta,dt) \\
&\qquad + e^{\beta t}\|\delta Z_t\|^2 dt + e^{\beta t}\int_{\mathcal{Z}} \|\delta\Theta_t(\zeta)\|^2 \mu(d\zeta,dt).
\end{aligned} \tag{19.3}$$

The next step is to integrate this equation on the interval $]s,T]$ and take an expectation. However, for the sake of clarity, we consider the terms of this equation separately. We know

$$\begin{aligned}
\Big(\int_{[0,T]} e^{2\beta t}\big\|\delta Y_t^\top \delta Z_t\|^2 dt\Big)^{1/2} &\le m e^{\beta T}(\max_i \sup_t |\delta Y_t^i|^2)^{1/2}\Big(\int_{[0,T]}\big\|\delta Z_t\|^2 dt\Big)^{1/2} \\
&\le \frac{m e^{\beta T}}{2}\Big(\max_i \sup_t |\delta Y_t^i|^2 + \int_{[0,T]}\big\|\delta Z_t\|^2 dt\Big),
\end{aligned} \tag{19.4}$$

so, as $\delta Y \in S^2$ and $\delta Z \in L^2(\langle W\rangle)$, it follows from the BDG inequality that $\big\{\int_{]0,s]} e^{\beta t}\delta Y_t^\top \delta Z_t dW_t\big\}_{s\ge 0} \in \mathcal{H}^1$, in particular,

$$E\Big[\int_{]s,T]} e^{\beta t}\delta Y_t^\top \delta Z_t dW_t\Big|\mathcal{F}_s\Big] = 0.$$

Using the same argument, we observe that

$$E\Big[\int_{\mathcal{Z}\times]s,T]} e^{\beta t}\delta Y_t^\top \delta\Theta_t(\zeta)\tilde\mu(d\zeta,dt)\Big|\mathcal{F}_s\Big] = 0$$

and

$$E\Big[\int_{\mathcal{Z}\times]s,T]} e^{\beta t}\|\delta\Theta_t(\zeta)\|^2\mu(d\zeta,dt)\Big|\mathcal{F}_s\Big] = E\Big[\int_{\mathcal{Z}\times]s,T]} e^{\beta t}\|\delta\Theta_t(\zeta)\|_\nu^2 dt\Big|\mathcal{F}_s\Big].$$

For any $\lambda > 0$, using the inequality $2a^\top b \le \lambda\|a\|^2 + \lambda^{-1}\|b\|^2$, we know

$$\begin{aligned}
&2\delta Y_t^\top (f^1(\omega,t,Y_t^1,Z_t^1,\Theta_t^1) - f^2(\omega,t,Y_t^2,Z_t^2,\Theta_t^2)) \\
&\le \lambda\|\delta Y_t\|^2 + \lambda^{-1}\|f^1(\omega,t,Y_t^1,Z_t^1,\Theta_t^1) - f^2(\omega,t,Y_t^2,Z_t^2,\Theta_t^2)\|^2 \\
&\le \lambda\|\delta Y_t\|^2 + 2\lambda^{-1}\|f^1(\omega,t,Y_t^1,Z_t^1,\Theta_t^1) - f^1(\omega,t,Y_t^2,Z_t^2,\Theta_t^2)\|^2 + 2\lambda^{-1}\|\delta_2 f\|^2 \\
&\le \lambda\|\delta Y_t\|^2 + 2K\lambda^{-1}\Big(\|\delta Y_t\|^2 + \|\delta Z\|^2 + \|\delta\Theta\|_\nu^2\Big) + 2\lambda^{-1}\|\delta_2 f\|^2,
\end{aligned}$$

and hence

$$
\begin{aligned}
&-2\int_{]s,T]} e^{\beta t} E[\delta Y_t^\top (f^1(\omega,t,Y_t^1,Z_t^1,\Theta_t^1) - f^2(\omega,t,Y_t^2,Z_t^2,\Theta_t^2))]dt \\
&\geq -\int_{]s,T]} e^{\beta t} E\Big[\Big(\lambda + \frac{2K}{\lambda}\Big)\|\delta Y_t\|^2 + \frac{2K}{\lambda}\Big(\|\delta Z_t\|^2 + \|\delta\Theta_t\|_\nu^2\Big) + \frac{2}{\lambda}\|\delta_2 f_t\|^2\Big]dt.
\end{aligned}
$$

Therefore, integrating (19.3) over $]s,T]$ and taking an expectation, we obtain, for any $\beta > 0$,

$$
\begin{aligned}
&e^{\beta T} E[\|\xi^1 - \xi^2\|^2] - e^{\beta s} E[\|\delta Y_s\|^2] \\
&= \int_{]s,T]} \beta e^{\beta t} E[\|\delta Y_t\|^2]dt + \int_{]s,T]} e^{\beta t} E[\|\delta Z_t\|^2 + \|\delta\Theta_t\|_\nu^2]dt \\
&\qquad - 2\int_{]s,T]} e^{\beta t} E[\delta Y_t^\top (f^1(\omega,t,Y_t^1,Z_t^1,\Theta_t^1) - f^2(\omega,t,Y_t^2,Z_t^2,\Theta_t^2))]dt \\
&\geq -\frac{2}{\lambda}\int_{]s,T]} e^{\beta t} E[\|\delta_2 f_t\|^2]dt + \Big(\beta - \lambda - \frac{2K}{\lambda}\Big)\int_{]s,T]} e^{\beta t} E[\|\delta Y_t\|^2]dt \\
&\qquad + \Big(1 - \frac{2K}{\lambda}\Big)\Big(\int_{]s,T]} e^{\beta t} E[\|\delta Z_t\|^2 + \|\delta\Theta_t\|_\nu^2]dt\Big).
\end{aligned}
$$

Taking $\lambda = 4K$, as $\beta - \lambda - 2K/\lambda = \beta - 4K - 1/2 > 0$ we have

$$
\begin{aligned}
&e^{\beta s} E\Big[\|\delta Y_s\|^2\Big] + \frac{1}{2}\int_{]s,T]} e^{\beta t} E\Big[\|\delta Z_t\|^2 + \|\delta\Theta_t\|_\nu^2\Big]dt \\
&\leq E\Big[\|\xi^1 - \xi^2\|^2 + \frac{1}{2K}\int_{]s,T]} e^{\beta t}\|\delta_2 f_t\|^2 dt\Big].
\end{aligned}
$$

Finally, as we can always take the Lipschitz constant to be larger than its minimal value, we can assume without loss of generality that $4K + 1/2 = \beta$. Then $1/(2K) = 4/(2\beta - 1)$, and the result holds as stated. $\square$

Theorem 19.1.7. *Let (ξ, f) be standard Lipschitz data. Then (19.1) admits a unique solution $(Y, Z, \Theta) \in S^2 \times L^2(\langle W\rangle) \times L^2(\langle\tilde{\mu}\rangle)$.*

Proof. We note that f and ξ are fixed. Using Lemma 19.1.5, we can see that for any fixed $(y, z, \theta) \in S^2 \times L^2(\langle W\rangle) \times L^2(\langle\tilde{\mu}\rangle)$, the equation

$$
\begin{cases}
dY_t = -f(\omega, t, y_t, z_t, \theta_t)dt + Z_t dW_t + \int_{\zeta\in\mathcal{Z}} \Theta_t(\zeta)\tilde{\mu}(d\zeta, dt); \\
Y_T = \xi
\end{cases}
\tag{19.5}
$$

admits a unique solution $(Y, Z, \Theta) \in S^2 \times L^2(\langle W\rangle) \times L^2(\langle\tilde{\mu}\rangle)$. We wish to show the map $\Phi : (y, z, \theta) \to (Y, Z, \Theta)$ defined in this way is a contraction under an equivalent norm.

In the setting of Theorem 19.1.6, by integrating the inequality that was proven and using Fubini's theorem, we have, for any $\beta \geq 4K + 1/2$,

$$\begin{aligned}\int_{[0,T]} e^{\beta s} E[\|\delta Y_s\|^2] ds + \frac{T}{2} \int_{[0,T]} e^{\beta t} E[\|\delta Z_t\|^2 + \|\delta \Theta_t\|^2_\nu] dt \\ \leq E\Big[T\|\xi^1 - \xi^2\|^2 + \frac{4T}{2\beta - 1} \int_{[0,T]} e^{\beta t} \|\delta_2 f_t\|^2 dt\Big].\end{aligned}$$

Therefore, if we take $(Y^1, Z^1, \Theta^1) = \Phi(y^1, z^1, \theta^1)$ and $(Y^2, Z^2, \Theta^2) = \Phi(y^2, z^2, \theta^2)$, we know that

$$\delta_2 f_t = f^1(\omega, t, y^1_t, z^1_t, \theta^1_t) - f(\omega, t, y^2_t, z^2_t, \theta^2_t),$$

and hence

$$\begin{aligned}\int_{[0,T]} e^{\beta s} E[\|\delta Y_s\|^2] ds + \frac{T}{2} \int_{[0,T]} e^{\beta t} E[\|\delta Z_t\|^2 + \|\delta \Theta_t\|^2_\nu] dt \\ \leq E\Big[\frac{4T}{2\beta - 1} \int_{[0,T]} e^{\beta t} \|\delta_2 f_t\|^2 dt\Big] \\ \leq \frac{4TK}{2\beta - 1} E\Big[\int_{[0,T]} e^{\beta t} (\|y^1_t - y^2_t\|^2 + \|z^1_t - z^2_t\|^2 + \|\theta^1_t - \theta^2_t\|^2_\nu) dt\Big].\end{aligned}$$

Fixing a value of $\beta > 2(T \vee 2)K + 1/2$, we see that we have a contraction under the equivalent norm on $L^2(P \times t) \times L^2(\langle W \rangle) \times L^2(\langle \tilde{\mu} \rangle)$ given by

$$(y, z, \theta) \mapsto \int_{[0,T]} e^{\beta s} E[\|y\|^2] ds + \frac{T}{2} \int_{[0,T]} e^{\beta t} E[\|z_t\|^2 + \|\theta_t\|^2_\nu] dt.$$

Therefore, by the contraction mapping principle (Lemma 1.5.18), there exists a unique triple (Y, Z, Θ) in the desired space which is a fixed point of Φ. From the definition of Φ, we see that Y is a semimartingale and hence the triple (Y, Z, Θ) solves the BSDE (19.1). By Lemma 19.1.4 it follows that $Y \in S^2$, as desired. □

A key property of BSDEs is that they satisfy the following 'flow' property.

Lemma 19.1.8. *Let (Y, Z, Θ) be the solution of the BSDE with standard Lipschitz data (ξ, f) at terminal time T. Let S be any stopping time $S \leq T$. Then on the interval $[0, S]$, the triple (Y, Z, Θ) is the unique solution of the BSDE with data (Y_S, f) at time S.*

Proof. We first note that (Y, Z, Θ) is a solution to the BSDE on $[0, S]$, as it satisfies the dynamics of (19.1) and has the correct terminal value. Second, we note that the equation on $[0, S]$ agrees with the BSDE on $[0, T]$ with standard Lipschitz data $(Y_S, \hat{f})$, where

$$\hat{f}(\omega, t, y, z, \theta) = I_{\{t \leq S\}} f(\omega, t, y, z, \theta).$$

This BSDE has a unique solution, and so our solution on $[0, S]$ is also unique. □

Remark 19.1.9. The requirement that f be Lipschitz in the previous lemma is simply to ensure that there exists a unique solution. If one establishes this result under weaker conditions (for example, that f is of quadratic growth in z, as in Appendix A.9), then the flow property follows.

19.2 Linear BSDEs

As usual, there are very few examples of these equations for which there is a closed form solution (for general ξ). However, for linear equations, we have the following representation. For simplicity, we consider the scalar case $m = 1$. However, a similar result is possible in general. We begin with a useful integrability result.

Lemma 19.2.1. *Let α be a predictable process taking values in $\mathbb{R}^N$ and β be a predictable process taking values in $L^2(\nu)$, such that, for some constant $K > 0$,*

$$\|\alpha_t\|^2 + \|\beta_t\|_\nu^2 < K \qquad dt \times dP\text{-almost everywhere.}$$

Then the stochastic exponential $\mathcal{E}(\alpha^\top \bullet W + \beta * \tilde{\mu})$ is a square integrable martingale.

Proof. Calculating the predictable quadratic variation, we have

$$\langle \mathcal{E}(\alpha^\top \bullet W + \beta * \tilde{\mu})\rangle_t = \int_{[0,t]} \mathcal{E}(\alpha^\top \bullet W + \beta * \tilde{\mu})_{s-}^2 (\|\alpha_s\|^2 + \|\beta_s\|_\nu^2) ds,$$

so, for some localizing sequence $T_n \uparrow \infty$,

$$\begin{aligned}
&E[I_{\{t \le T_n\}} (\mathcal{E}(\alpha^\top \bullet W + \beta * \tilde{\mu})_t)^2] \\
&\le E[(\mathcal{E}(\alpha^\top \bullet W + \beta * \tilde{\mu})_{t\wedge T_n})^2] \\
&= E[1 + \langle \mathcal{E}(\alpha^\top \bullet W + \beta * \tilde{\mu})\rangle_{t\wedge T_n}] \\
&= 1 + \int_{[0,t]} E\big[I_{\{s \le T_n\}} \mathcal{E}(\alpha^\top \bullet W + \beta * \tilde{\mu})_{s-}^2 (\|\alpha_s\|^2 + \|\beta_s\|_\nu^2)\big] ds
\end{aligned}$$

By Grönwall's inequality, it follows that

$$E[I_{\{t \le T_n\}} \mathcal{E}(\alpha^\top \bullet W + \beta * \tilde{\mu})_t^2] \le e^{Kt}.$$

Using monotone convergence, we conclude that

$$E[1 + \langle \mathcal{E}(\alpha^\top \bullet W + \beta * \tilde{\mu})\rangle_t^2] \le e^{Kt} < \infty,$$

so $\mathcal{E}(\alpha^\top \bullet W + \beta * \tilde{\mu})$ is a square integrable martingale. $\square$

In what follows, recall that $z \in \mathbb{R}^{m\times N}$, so if $m = 1$, then for $\alpha \in \mathbb{R}^N$, we have $z\alpha \in \mathbb{R}$.

Theorem 19.2.2. *Suppose $m = 1$. Consider the BSDE with linear driver*

$$f(\omega, t, y, z, \theta) = \phi_t + \rho_t y + z\alpha_t + \int_{\mathcal{Z}} \beta_t(\zeta)\theta(\zeta)\nu(d\zeta),$$

where

- ϕ *is a predictable process in* $L^2(P \times t)$,
- ρ *is a predictable process taking values in* $\mathbb{R}$,
- α *is a predictable process taking values in* $\mathbb{R}^N$,
- β *is a predictable process taking values in* $L^2(\nu)$, *with*

$$\beta_t(\zeta) \neq -1 \qquad d\nu \times dt \times dP\text{-almost everywhere,}$$

- *and there exists* K *such that*

$$|\rho_t| + \|\alpha_t\|^2 + \|\beta_t\|_\nu^2 < K \qquad dt \times dP\text{-almost everywhere.}$$

Then the process Y *in the solution of the BSDE (19.1) is given by*

$$Y_t = (\mathcal{E}(\Gamma)_t)^{-1} E\Big[\mathcal{E}(\Gamma)_T \xi + \int_{]t,T]} \mathcal{E}(\Gamma)_{s-}\phi_s ds \Big| \mathcal{F}_t\Big],$$

where $\mathcal{E}$ *denotes the stochastic exponential, and*

$$\Gamma_t = \int_{[0,t]} \rho_u du + (\alpha^\top \bullet W)_t + (\beta * \tilde{\mu})_t.$$

Proof. As $\beta_t(\zeta) \neq -1$ almost everywhere and is predictable, we know that $\Delta\Gamma_t \neq -1$ up to indistinguishability (as the jump times of μ are totally inaccessible), so the stochastic exponentials are nonzero (Lemma 15.1.5) and the proposed equation for Y is well defined. It is easy to verify that the driver f satisfies the assumptions of Theorem 19.1.7, so a unique solution (Y, Z, Θ) exists. Applying Itô's product rule to $\mathcal{E}(\Gamma)Y$, we have

$$\begin{aligned}
\frac{d(\mathcal{E}(\Gamma)_t Y_t)}{\mathcal{E}(\Gamma)_{t-}} &= dY_t + Y_t d\Gamma_t + [Y, \Gamma]_t \\
&= -\Big(\phi_t + \rho_t Y_t + Z_t\alpha_t + \int_{\mathcal{Z}} \beta_t(\zeta)\Theta_t(\zeta)\nu(d\zeta)\Big)dt \\
&\qquad + Z_t dW_t + \int_{\mathcal{Z}} \Theta_t(\zeta)\tilde{\mu}(d\zeta, dt) \\
&\qquad + Y_{t-}\Big(\rho_t dt + \alpha_t^\top dW_t + \int_{\mathcal{Z}} \beta_t(\zeta)\tilde{\mu}(d\zeta, dt)\Big) \\
&\qquad + Z_t\alpha_t dt + \int_{\mathcal{Z}} \beta_t(\zeta)\Theta_t(\zeta)\mu(d\zeta, dt) \\
&= -\phi_t dt + (Z_t + Y_t\alpha_t^\top)dW_t \\
&\qquad + \int_{\mathcal{Z}} (\beta_t(\zeta)\Theta_t(\zeta) + \Theta_t(\zeta) + Y_{t-}\beta_t(\zeta))\tilde{\mu}(d\zeta, dt).
\end{aligned}$$

From this, we see that the process

$$\mathcal{E}(\Gamma)_t Y_t + \int_{]0,t]} \mathcal{E}(\Gamma)_{s-} \phi_s ds$$

is a local martingale.

We know that

$$\mathcal{E}(\Gamma)_t = e^{\int_{[0,t]} \rho_s ds} \mathcal{E}(\alpha^\top \bullet W + \beta * \tilde{\mu})_t.$$

By Lemma 19.2.1, $\mathcal{E}(\alpha^\top \bullet W + \beta * \tilde{\mu})$ is a square integrable martingale. As $Y \in S^2$ and ρ is bounded, the same argument as in (19.4) implies that $\left\{\mathcal{E}(\Gamma)_t Y_t + \int_{]0,t]} \mathcal{E}(\Gamma)_{s-} \phi_s ds\right\}_{t \geq 0}$ is a martingale in $\mathcal{H}^1$, and so

$$\mathcal{E}(\Gamma)_t Y_t + \int_{]0,t]} \mathcal{E}(\Gamma)_{s-} \phi_s ds = E\Big[\mathcal{E}(\Gamma)_T \xi + \int_{]0,T]} \mathcal{E}(\Gamma)_{s-} \phi_s ds \Big| \mathcal{F}_t\Big].$$

The result follows by rearrangement. □

Remark 19.2.3. The key step in this proof was Lemma 19.2.1, which established that $\mathcal{E}(\alpha^\top \bullet W + \beta * \tilde{\mu})$ is a true martingale. Provided $\beta > -1$, we also know that it is positive, so defining the equivalent measure Q by

$$\frac{dQ}{dP} = \mathcal{E}(\alpha^\top \bullet W + \beta * \tilde{\mu})_T$$

and using Girsanov's theorem, we see that

$$\begin{aligned} dY_t &= -(\phi_t + \rho_t Y_t)dt + Z_t dW_t^Q + \int_{\mathcal{Z}} \Theta_t(\zeta) \tilde{\mu}^Q(d\zeta, dt) \\ &= -(\phi_t + \rho_t Y_t)dt + d(\text{some } Q\text{-martingale}), \end{aligned}$$

where

$$dW_t^Q = dW_t - \alpha_t dt$$

corresponds to an $\mathbb{R}^N$-valued Q-Brownian motion, and

$$\tilde{\mu}^Q(d\zeta, dt) = \mu(d\zeta, dt) - \nu'(d\zeta)dt = \tilde{\mu}(d\zeta, dt) - \beta_t(\zeta)\nu(d\zeta)dt$$

to the random measure μ compensated under Q. We can also write this representation in the form

$$Y_t = E^Q\Big[e^{\int_{]t,T]} \rho_s ds} \xi + \int_{]t,T]} e^{\int_{]t,s]} \rho_u du} \phi_s ds \Big| \mathcal{F}_t\Big].$$

In this way, particularly taking $\rho = 0$ and $\phi = 0$, we can see that a linear BSDE forms a natural way of encoding a change of measure.

In Appendix A.9, we consider two usefully weaker continuity assumptions on the driver f, namely where f is Lipschitz in z, but the Lipschitz constant depends on (ω, t), and where f is of quadratic growth in z. In the case where there is no direct dependence on y in the driver, and the terminal value is bounded, we show that the BSDE has a unique solution. Beyond this case, some other extensions (among many) include Darling and Pardoux [40] where the equation is up to a stopping time, Lepeltier and San Martìn where coefficients are only assumed to be continuous and of linear growth, Briand, Delyon, Hu, Pardoux and Stoica [23] (and references therein) where f is of superlinear growth in y under an asymptotic monotonicity condition, and Royer [161] where the equation is on an infinite horizon. A significant regularity result in Z, which is vital for numerical calculations, is due to Zhang [192], see also Imkeller and Dos Reis [99] in the quadratic growth case. Various numerical algorithms are available, see Bouchard and Touzi [22], Chassagneux and Richou [31], and references therein.

19.3 Comparison Theorem

We can now prove the 'comparison theorem', which is arguably the most useful result in the theory of BSDEs. As the name suggests, this allows us to compare the solutions to two BSDEs, that is, to say that if the data satisfy an inequality, then so do the solutions. This is naturally restricted to the scalar ($m = 1$) case; however, extensions are possible, with significantly more restrictive conditions (e.g. the 'viability property' of Hu and Peng [97]).

In the continuous case, the comparison theorem holds without restriction. In the setting with jumps, however, we require the following assumption.

Definition 19.3.1. *We say that f is a* balanced *driver if there exists a map $\beta : \Omega \times [0,T] \times \mathbb{R} \times \mathbb{R}^N \times L^2(\nu) \times L^2(\nu) \times \mathcal{Z} \to \mathbb{R}$ such that*

- *β is $\Sigma_p \otimes \mathcal{B}(\mathbb{R} \times \mathbb{R}^N \times L^2(\nu) \times L^2(\nu)) \otimes \mathfrak{Z}$-measurable, (i.e. predictable in (ω, t), Borel measurable in (y, z, θ, θ') and measurable in ζ)*
- *$\beta > -1$ $\nu(d\zeta)$-a.e., for all (y, z, θ, θ') and $dP \times dt$ almost all (ω, t), and*
- *for $dP \times dt$-almost all (ω, t), for all (y, z, θ, θ'),*

$$f(\omega, t, y, z, \theta) - f(\omega, t, y, z, \theta') = \int_{\mathcal{Z}} (\theta(\zeta) - \theta'(\zeta))\beta(\omega, t, y, z, \theta, \theta', \zeta)\nu(d\zeta).$$

Remark 19.3.2. In the continuous case, $\nu \equiv 0$, so all drivers are balanced.

Example 19.3.3. Let ν' be a measure on $(\mathcal{Z}, \mathfrak{Z})$ equivalent to ν, such that $(d\nu'/d\nu - 1) \in L^2(\nu)$. Then the driver

$$f(\omega, t, y, z, \theta) = \int_{\mathcal{Z}} \Big(\frac{d\nu'}{d\nu}(\zeta) - 1\Big)\theta(\zeta)\nu(d\zeta)$$

is balanced, with $\beta(\cdots, \zeta) = \frac{d\nu'}{d\nu}(\zeta) - 1$.

By Remark 19.2.3 and Girsanov's theorem, the solution to the BSDE with this driver corresponds to the conditional expectation under the measure Q where μ has compensator $\nu'(d\zeta)dt$. Therefore, if $\xi \geq \xi'$ a.s., it is easy to see that $Y_t = E^Q[\xi|\mathcal{F}_t] \geq E^Q[\xi'|\mathcal{F}_t] = Y'_t$. Establishing this fact in a more general setting is the purpose of the comparison theorem.

Theorem 19.3.4 (The Comparison Theorem). *Let (ξ, f) and (ξ', f') be standard Lipschitz data for two BSDEs, with solutions (Y, Z, Θ) and (Y', Z', Θ') respectively. Suppose*

- $\xi \geq \xi' \quad P-\text{a.s.},$
- $f(\omega, t, y, z, \theta) \geq f'(\omega, t, y, z, \theta) \quad dt \times dP - \text{a.s.}$ *for all* (y, z, θ)

and at least one of f and f' is balanced. Then

$$Y \geq Y'$$

up to indistinguishability. Furthermore, if for some $A \in \mathcal{F}_t$ we also have $I_A(Y_t - Y'_t) = 0$, then $Y = Y'$ on $A \times [t, T]$ (that is, if Y and Y' meet, they remain the same from then onwards).

Proof. Suppose f is balanced. Given our solutions (Y, Z, Θ) and (Y', Z', Θ'), with the convention that $0/0 := 0$, we define

$$\begin{aligned}
\phi_t &:= f(\omega, t, Y'_t, Z'_t, \Theta'_t) - f'(\omega, t, Y'_t, Z'_t, \Theta'_t),\\
\rho_t &:= \frac{f(\omega, t, Y_t, Z'_t, \Theta'_t) - f(\omega, t, Y'_t, Z'_t, \Theta'_t)}{Y_t - Y'_t},\\
\alpha_t &:= \frac{f(\omega, t, Y_t, Z_t, \Theta'_t) - f(\omega, t, Y_t, Z'_t, \Theta'_t)}{\|Z_t - Z'_t\|^2}(Z_t - Z'_t)^\top,
\end{aligned}$$

and $\beta_t(\zeta) = \beta(\omega, t, Y_t, Z_t, \Theta_t, \Theta'_t, \zeta)$ as in Definition 19.3.1. Writing $\delta Y = Y - Y'$, $\delta Z = Z - Z'$ and $\delta\Theta = \Theta - \Theta'$, we have the linearized equation

$$\begin{aligned}
d(\delta Y_t) = -\Big(\phi_t + \rho_t(\delta Y_t) + \delta Z_t \alpha_t + \int_{\mathcal{Z}} \beta_t(\zeta)(\delta\Theta_t(\zeta))\nu(d\zeta)\Big)dt\\
+ \delta Z_t dW_t + \int_{\mathcal{Z}} \delta\Theta_t(\zeta)\tilde{\mu}(d\zeta, dt).
\end{aligned}$$

As f is Lipschitz and balanced, we know that $\beta > -1$ and $|\rho| + \|\alpha\|^2 + \|\beta\|^2_\nu < K$ for some fixed $K > 0$. Therefore, by Theorem 19.2.2,

$$\delta Y_t = (\mathcal{E}(\Gamma)_t)^{-1} E\Big[\mathcal{E}(\Gamma)_T(\xi - \xi') + \int_{]t,T]} \mathcal{E}(\Gamma)_{s-}\phi_s ds \Big| \mathcal{F}_t\Big],$$

where $\Gamma_t = \int_{[0,t]} \rho_u du + (\alpha^\top \bullet W)_t + (\beta * \tilde{\mu})_t$. As $\beta > -1$, we know that $\mathcal{E}(\Gamma) > 0$. The inequality $Y \geq Y'$ follows from the fact that $\xi - \xi' \geq 0$ and

$\phi \geq 0$, by the assumptions of the theorem. If, in addition, $I_A(Y_t - Y'_t) = 0$, then simple rearrangement yields $I_A(\xi - \xi') = I_A \phi_s = 0$ for $s \in]t, T]$, and therefore $I_A(Y_s - Y'_s) = 0$ for $s \in]t, T]$.

If f' is balanced, instead of f, then the linearization can be taken in a different order, and the result established in the same way. □

Remark 19.3.5. If $Y = Y'$ on $A \times [t, T]$, then, from the uniqueness of the canonical semimartingale decomposition and the martingale representation theorem, we see that $Z_t = Z'_t \; dt \times dP$-a.e. and $\Theta_t(\zeta) = \Theta'_t(\zeta) \; d\nu \times dt \times dP$-a.e. on $A \times [t, T]$.

Remark 19.3.6. It is clear from the proof that the assumptions of the theorem only need to hold at the solutions themselves, that is, they can be relaxed to assuming that $\xi \geq \xi'$ a.s., f is balanced, and

$$f(\omega, t, Y'_t, Z'_t, \Theta'_t) \geq f'(\omega, t, Y'_t, Z'_t, \Theta'_t) \qquad dt \times dP - \text{a.s.}$$

where (Y', Z', Θ') is the solution of the second equation. Decomposing in a different order yields the result under the assumption that f' is balanced and

$$f(\omega, t, Y_t, Z_t, \Theta_t) \geq f'(\omega, t, Y_t, Z_t, \Theta_t) \qquad dt \times dP - \text{a.s.}$$

One can also weaken the 'balanced' assumption to specify the processes Y and Z to be one of the solutions. While these conditions are weaker, they are often difficult to verify (except in the special case where one of the solutions is deterministic).

Remark 19.3.7. Given the connection to PIDEs below, the comparison theorem stated here can be seen as the natural stochastic analogue of the maximum principle for semilinear parabolic PDEs.

One difficulty when working with the comparison theorem is that it requires that the drivers be balanced. In the continuous case, no issues arise, but we need to be careful in the presence of jumps. Providentially, this property is preserved by taking infima and suprema, as we show in the following lemma. The conditions of this result may seem overly complicated. However, they are needed to ensure sufficient measurability of the result.

Lemma 19.3.8. *For a given index set U, let $f(\cdots; u)$ be a standard balanced BSDE driver for every $u \in U$. Suppose*

(i) the maps $(y, z, \theta) \mapsto f(\omega, t, y, z, \theta; u)$ have common uniform Lipschitz constant K,

(ii) writing $\beta(\omega, t, y, z, \theta, \theta', \zeta; u)$ for the associated processes in Definition 19.3.1, (and omitting all but the last argument for clarity)

$$\operatorname{ess\,inf}_{u \in U} \beta(u) > -1 \qquad \nu(d\zeta) \times dP \times dt - \text{a.e.}, \text{ for all } (y, z, \theta, \theta')$$

the essential infimum being taken for β in the predictable $L^2(\nu)$-valued processes, for each (y, z, θ, θ'), $dt \times dP$-a.e., *(iii)* $\sup_u\{|f(\omega,t,0,0,0;u)|^2\}$ *is bounded by a predictable $dt \times dP$-integrable process,*

(iv) the maps $u \mapsto \beta(\omega,t,y,z,\theta,\theta',\zeta,u)$ are continuous, for fixed $(\omega,t,y,z,\theta,\theta',\zeta)$, and U is a countable union of compact metrizable subsets of itself.

Then there is a version of the mappings

$$\underline{f}(\omega,t,y,z,\theta) = \operatorname{ess\,inf}_{u\in U} f(\omega,t,y,z,\theta,u),$$
$$\overline{f}(\omega,t,y,z,\theta) = \operatorname{ess\,sup}_{u\in U} f(\omega,t,y,z,\theta,u),$$

which are standard balanced BSDE drivers.

Proof. We consider the definition of $\underline{f}$. The argument for $\overline{f}$ is almost identical. We face two issues. First, as the essential infimum (Theorem 1.3.40) is only defined almost everywhere, we need to be careful when defining $\underline{f}$ for all (y,z,θ), so as not to lose measurability. Second, we need to show that the result is balanced.

Fix a countable dense subset of $\mathbb{R}\times\mathbb{R}^N\times L^2(\nu)$ (where $\mathbb{R}^N$ has the ℓ_2 topology when $N=\infty$). For each fixed (y,z,θ) in this subset, define $\underline{f}(\cdot,\cdot,y,z,\theta)$ by taking the essential infimum, in the class of predictable processes, $dt\times dP$-a.e. From properties (i) and (iii), this infimum is finite almost everywhere. We shall show that $\underline{f}$ is (Lipschitz) continuous in (y,z,θ), and hence can be extended to all of $\mathbb{R}\times\mathbb{R}^N\times L^2(\nu)$.

We know that f is Lipschitz with respect to (y,z,θ), uniformly in (ω,t,u) (that is, the Lipschitz constant does not depend on (ω,t,u)). Therefore, replacing the essential infima by the limits of pointwise decreasing sequences, we see that, on our dense subset, the essential infimum equals the classical infimum over a countable subset of $\mathcal{U}$. By the standard argument that the pointwise infimum of a uniformly Lipschitz collection of functions is itself Lipschitz, we know that $\underline{f}$ is Lipschitz continuous in (y,z,θ), uniformly in (ω,t). Therefore $\underline{f}$ can be extended to all (y,z,θ) in a ($dt\times dP$-a.e.) unique (uniformly Lipschitz) continuous way. By uniqueness, this extension is a version of the essential infimum, that is,

$$\underline{f}(\omega,t,y,z,\theta) = \operatorname{ess\,inf}_u f(\omega,t,y,z,\theta,u)$$

for all (y,z,θ) and almost all (ω,t).

It remains to show that $\underline{f}$ is balanced. We know that $f(\cdots,u)$ is balanced. In particular, for any (ω,t,y,z), which we omit as arguments of f and β for clarity, and for any $\theta,\theta'\in L^2(\nu)$ (also omitted as arguments of β),

$$f(\theta,u)-f(\theta',u) = \int_{\mathcal{Z}} \big(\theta(\zeta)-\theta'(\zeta)\big)\big(\beta(\zeta,u)-1\big)\nu(d\zeta).$$

Hence, for any u,

$$\underline{f}(\theta)-f(\theta',u) \le \int_{\mathcal{Z}} \big(\theta(\zeta)-\theta'(\zeta)\big)\big(\beta(\zeta,u)-1\big)\nu(d\zeta).$$

Therefore, for any $\epsilon > 0$, there exists $u^\epsilon \in U$ (which may depend on $\omega, t, y, z, \theta, \theta'$, but can be chosen to do so measurably by the results of Appendix A.10) such that

$$\underline{f}(\theta) - \underline{f}(\theta') \leq \epsilon + \int_{\mathcal{Z}} \big(\theta(\zeta) - \theta'(\zeta)\big)\big(\beta(\zeta, u^\epsilon) - 1\big)\nu(d\zeta).$$

As β is $L^2(\nu)$ bounded, uniformly in u, taking $\epsilon \to 0$ implies[1] that there exists $\beta^* \in L^2(\nu)$ such that

$$\underline{f}(\theta) - \underline{f}(\theta') \leq \int_{\mathcal{Z}} \big(\theta(\zeta) - \theta'(\zeta)\big)\big(\beta^*(\zeta) - 1\big)\nu(d\zeta) \qquad dt \times dP\text{−a.e.}$$

Exchanging the roles of θ and θ', we obtain β_* such that

$$\underline{f}(\theta) - \underline{f}(\theta') \geq \int_{\mathcal{Z}} \big(\theta(\zeta) - \theta'(\zeta)\big)\big(\beta_*(\zeta) - 1\big)\nu(d\zeta) \qquad dt \times dP\text{−a.e.}$$

We can, therefore, define

$$\tilde{\beta}(\zeta) = \phi\beta^*(\zeta) + (1-\phi)\beta_*(\zeta) > -1,$$

where

$$\phi = \frac{\underline{f}(\theta) - \underline{f}(\theta') - \int_{\mathcal{Z}} \big(\theta(\zeta) - \theta'(\zeta)\big)\big(\beta_*(\zeta) - 1\big)\nu(d\zeta)}{\int_{\mathcal{Z}} \big(\theta(\zeta) - \theta'(\zeta)\big)\big(\beta^*(\zeta) - \beta_*(\zeta)\big)\nu(d\zeta)} \in [0,1].$$

With this $\tilde{\beta}$ we have

$$\underline{f}(\theta) - \underline{f}(\theta') = \int_{\mathcal{Z}} \big(\theta(\zeta) - \theta'(\zeta)\big)\big(\tilde{\beta}(\zeta) - 1\big)\nu(d\zeta) \qquad dt \times dP\text{−a.e.},$$

that is, $\underline{f}$ is balanced. □

Corollary 19.3.9. *Let f be a balanced driver for a BSDE. Then so are f^+ and f^- (the positive and negative parts of f).*

Proof. As $f \equiv 0$ is clearly a balanced driver, this can be obtained by observing $f^+ = f \vee 0$, and similarly for f^- (and many other examples). □

Remark 19.3.10. In Lemma 19.3.8, we only need the uniform Lipschitz constant K to ensure that the functions $\underline{f}$ and $\overline{f}$ are also Lipschitz. More generally, if we assume that the continuity of f is uniform in u, then the result would continue to hold, with $\underline{f}$ and $\overline{f}$ having the same continuity as we have assumed on f. For example, we might have linear growth in the derivative of $f(u,\cdot)$ (uniformly in u), or that $f(u,\cdot)$ has a stochastic Lipschitz constant (uniformly in u). These cases are considered in Appendix A.9.

[1] This is because any closed, convex and bounded set in $L^2(\nu)$ is weakly compact, and so any bounded sequence has a weak limit in the space (Theorem 1.7.19). Using Theorem A.10.5, this weak limit can be chosen to be measurable in its other arguments.

19.4 Markovian BSDEs

We now see how a Markovian structure can be introduced in the theory of BSDEs. Our presentation is loosely based on that in El Karoui, Peng and Quenez [65].

We suppose we have a Markov 'forward process' $X^{(t,x)}$, defined as the solution of an SDE of the form of (17.2), namely

$$\begin{cases} dX_s^{(t,x)} = b(s, X_s^{(t,x)})ds + \sigma(s, X_s^{(t,x)})dW_s + \displaystyle\int_{\mathcal{Z}} g(\zeta, s, X_{s-}^{(t,x)})\tilde{\mu}(d\zeta, ds), \\ X_s^{(t,x)} = x \in \mathbb{R}^d \qquad \text{for } s \leq t. \end{cases} \tag{19.6}$$

Here W is an $\mathbb{R}^N$-valued Brownian motion, for $N \leq \infty$, and $\tilde{\mu}$ is a compensated integer valued random measure on a Blackwell space $(\mathcal{Z}, \mathfrak{Z})$, that is, $\tilde{\mu} = \mu - \mu_p$ for some $\mu \in \tilde{\mathcal{A}}_\sigma^1$. While we usually think of x as a deterministic value, it will sometimes be convenient to let it be an independently chosen random variable, and we note that $X^{(t,x)}$ is still a càdlàg adapted process in this case. We then consider the associated 'Markovian' BSDE

$$\begin{cases} dY_s^{(t,x)} = -f(s, X_s^{(t,x)}, Y_s^{(t,x)}, Z_s^{(t,x)}, \Theta_s^{(t,x)})ds + Z_s^{(t,x)}dW_s \\ \qquad\qquad + \displaystyle\int_{\mathcal{Z}} \Theta_s^{(t,x)}(\zeta)\tilde{\mu}(d\zeta, ds), \\ Y_T^{(t,x)} = \psi(X_T^{(t,x)}). \end{cases} \tag{19.7}$$

We write the solution to this equation $(Y^{(t,x)}, Z^{(t,x)}, \Theta^{(t,x)})$. If we now allow (t, x) to vary, we see that we have a family of BSDEs parameterized by the initial condition (t, x) of the forward equation. This connection is what we will seek to understand.

Remark 19.4.1. These equations are called 'Markovian', due to the Markov property of X. However, it is *not* the case that Y is a Markov process, rather, as we shall see, Y can be written $Y_s^{(t,x)} = v(s, X_s^{(t,x)})$ for some measurable function v.

Remark 19.4.2. Here our maps are all Borel measurable, with

- $b : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : [0,T] \times \mathbb{R}^d \to \mathbb{R}^{d\times N}$ and $g : \mathbb{R}^n \times [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$,
- $f : [0,T] \times \mathbb{R}^d \times \mathbb{R}^m \times \mathbb{R}^{m\times N} \times L^2(\nu) \to \mathbb{R}^m$,
- standard Lipschitz assumptions, i.e. for some $K > 0$, for all $s \in [0,T]$, $x, x' \in \mathbb{R}^d$,

$$\begin{aligned} &\|b(s,x) - b(s,x')\|^2 + \|\sigma(s,x) - \sigma(s,x')\|^2 + \|g(\cdot,s,x) - g(\cdot,s,x')\|_\nu^2 \\ &\leq K\|x - x'\|^2, \end{aligned}$$

  and for all $y, y' \in \mathbb{R}^m$, $z, z' \in \mathbb{R}^{m\times N}$ and $\theta, \theta' \in L^2(\nu)$,

$$\|f(t,x,y,z,\theta) - f(t,x,y',z',\theta')\|^2 \leq K(\|y - y'\|^2 + \|z - z'\|^2 + \|\theta - \theta'\|_\nu^2),$$

- the growth bounds

$$\|b(t,x)\|^2 + \|\sigma(t,x)\|^2 + \|g(\cdot,t,x)\|_\nu^2 \leq K(1+\|x\|^2)$$

and, for some $p \geq 1$,

$$\|f(t,x,y,z,\theta)\|^2 + \|\psi(x)\|^2 \leq K(1+\|x\|^p),$$

where, if $p > 2$, we also require

$$E\Big[\sup_t \Big(\int_{\mathcal{Z}} \|g(\zeta,t,X_t^{(s,x)})\|\mu(d\zeta,\{t\})\Big)^p\Big] < \infty.$$

Remark 19.4.3. From Lemma 17.1.1, we know that under these assumptions, there exists a unique strong solution $X^{(t,x)}$ to (19.6), which is in $S^{2\vee p}$. From Theorem 19.1.7, for each (t,x), it follows that there exists a solution $(Y^{(t,x)}, Z^{(t,x)}, \Theta^{(t,x)})$ to the BSDE (19.7).

Theorem 19.4.4. *Suppose the assumptions of Remark 19.4.2 hold. We know the following.*

(i) There exists $C \geq 0$ such that, for $t \in [0,T]$ and $x \in \mathbb{R}^d$,

$$E\Big[\sup_{0\leq s\leq T} \|Y_s^{(t,x)}\|^2\Big] + E\Big[\int_{[0,T]} (\|Z_s^{(t,x)}\|^2 + \|\Theta_s^{(t,x)}\|_\nu^2) ds\Big] \leq C(1+\|x\|^2).$$

(ii) There exists $C \geq 0$ such that, for each $t, t' \in [0,T]$ and $x, x' \in \mathbb{R}^d$,

$$\begin{aligned} &E\Big[\sup_{0\leq s\leq T} \|Y_s^{(t,x)} - Y_s^{(t',x')}\|^2\Big] \\ &\qquad + E\Big[\int_{[0,T]} \|Z_s^{(t,x)} - Z_s^{(t',x')}\|^2 + \|\Theta_s^{(t,x)} - \Theta_s^{(t',x')}\|_\nu^2 ds\Big] \\ &\leq C(1+\|x\|^2)(\|x-x'\|^2 + |t-t'|). \end{aligned}$$

Proof. From the standard estimates on BSDEs (Theorem 19.1.6), we know that there exists C such that

$$E\Big[\sup_{t\leq s\leq T} \|Y_s^{(t,x)}\|^2\Big] + E\Big[\int_{[0,T]} (\|Z_s^{(t,x)}\|^2 + \|\Theta_s^{(t,x)}\|_\nu^2) ds\Big] \leq CE\big[\|\psi(X_T^{(t,x)})\|^2\big].$$

As ψ is Lipschitz, we have $E\big[\|\psi(X_T^{(t,x)})\|^2\big] \leq K\big(1 + E\big[\|X_T^{(t,x)}\|^2\big]\big)$. Applying Lemma 17.1.1, we see that

$$\begin{aligned} &E\big[\sup_s \|X_s^{(t,x)}\|^2\big] \\ &\leq C\Big(\|x\|^2 + \int_{[0,T]} E\big[\|b(s,X_s^{(t,x)})\|^2 + \|\sigma(s,X_s^{(t,x)})\|^2 + \|g(\cdot,s,X_s^{(t,x)})\|_\nu^2\big] ds\Big). \end{aligned}$$

and using the Lipschitz continuity of b, σ and g and Grönwall's inequality, we can show

$$E\big[\|X_T^{(t,x)}\|^2\big] \leq E\big[\sup_s \|X_s^{(t,x)}\|^2\big] \leq C(1+\|x\|^2)$$

for some C.

The second inequality follows in a similar way, using the result of Exercise 17.5.3. □

Theorem 19.4.5. *There exists a continuous deterministic function $v : [0,T] \times \mathbb{R}^d \to \mathbb{R}^m$, such that the solution to the BSDE (19.7) satisfies*

$$Y_s^{(t,x)} = v(s, X_s^{(t,x)})$$

up to indistinguishability, for any $s \geq t$.

Proof. Fix t. Let x be randomly chosen, according to some distribution on $\mathbb{R}^d$ (with support equal to $\mathbb{R}^d$), independently of W and μ. Consider the complete, right continuous filtration $\{\tilde{\mathcal{F}}_s\}_{s\geq 0}$, generated by x, $W_{s\wedge t} - W_t$ and $I_{[t,\infty]} \cdot \mu$. As W and μ are Markovian, we know that $W_{s\wedge t} - W_t$ and $I_{[t,\infty]} \cdot \mu$ are a Brownian motion and Poisson random measure in the filtration $\{\tilde{\mathcal{F}}_s\}_{s\geq 0}$.

In this filtration, there exists a solution to the forward equation (19.6), by Lemma 17.1.1. As $\{\tilde{\mathcal{F}}_s\}_{s\geq 0}$ is a subfiltration of $\{\mathcal{F}_s\}_{s\geq 0} \vee \sigma(x)$, uniqueness of solutions to SDEs implies this solution must be a version of $X^{(t,x)}$. That is, $X^{(t,x)}$ is adapted to $\{\tilde{\mathcal{F}}_s\}_{s\geq 0}$.

We can solve our BSDE (19.7) in the filtration $\{\tilde{\mathcal{F}}_s\}_{s\geq 0}$. Therefore, we have a solution such that $Y_t^{(t,x)}$ is almost surely $\sigma(x)$-measurable. By uniqueness, this must agree with the solution to the BSDE in the filtration $\{\mathcal{F}_s\}_{s\geq 0} \vee \sigma(x)$. By the Doob–Dynkin lemma (Lemma 1.3.12), this implies that, for each t, there exists a $\mathcal{B}(\mathbb{R}^d)$-measurable function $x \mapsto v(t,x)$ with $Y_t^{(t,x)} = v(t,x)$ a.s. As x is permitted to be random, and $X^{(t,x)}$ satisfies the flow property $X^{(t,x)} = X^{(t', X_{t'}^{(t,x)})}$ for all $t \geq x$, uniqueness of Y also yields

$$v(s, X_s^{(t,x)}) = Y_s^{(s, X_s^{(t,x)})} = Y_s^{(t,x)} \qquad \text{a.s.}$$

We now think of (t,x) as deterministic parameters which we vary. From Theorem 19.4.4(ii), the family of functions $\{v(t,\cdot)\}_{t\geq 0}$ we have defined is jointly continuous in x and t. By right-continuity of Y and $s \mapsto v(s, X_s^{(t,x)})$, we conclude that they are equal up to indistinguishability. □

One can also show that $Z^{(t,x)}$ and $\Theta^{(t,x)}$ can be written as functions of $(s, X_s^{(t,x)})$; however this is more usefully understood given the connection to PIDEs in the coming section.

19.5 Connections to Semilinear PIDEs

We now show the connection between these Markovian BSDEs and semilinear partial integro-differential equations. This gives a generalization of the Feynman–Kac formula, and we shall also see how the theory of viscosity solutions can be used to give this connection under fairly weak conditions.

Theorem 19.5.1 (Semilinear Feynman–Kac formula). *Let v be a $C^{1,2}_\nu$ function and suppose that:*

(i) v is a solution to the following semilinear parabolic PIDE

$$\begin{cases} \quad 0 = \dfrac{\partial v}{\partial s}(s,x) + \mathcal{L}_t v(s,x) + f(s,x,v(s,x),\partial_x v(s,x)\sigma(s,x),\tilde{v}(s,x)), \\ \psi(x) = v(T,x), \end{cases} \tag{19.8}$$

where $\tilde{v}(s,x)$ denotes the element of $L^2(\nu)$ given by the map

$$\zeta \mapsto v\big(s, x + g(\zeta,s,x)\big) - v(s,x),$$

and $\mathcal{L}_t$ is the infinitesimal generator of solutions of the forward equation (19.6), as given by Definition 17.4.1.

(ii) There exists a constant K such that, for each (s,x),

$$\|v(s,x)\|^2 + \|\partial_x v(s,x)\sigma(s,x)\|^2 + \|\tilde{v}(s,x)\|_\nu^2 \leq K(1+\|x\|^2).$$

If X is the solution to (19.6), then

$$\begin{aligned} Y_s^{(t,x)} &= v(s, X_s^{(t,x)}), \\ Z_s^{(t,x)} &= \partial_x v(s, X_s^{(t,x)})\sigma(s, X_s^{(t,x)}), \\ \Theta_s^{(t,x)}(\zeta) &= \tilde{v}(\zeta; t, X_s^{(t,x)}) \\ &= v\big(s, X_s^{(t,x)} + g(\zeta, s, X_s^{(t,x)})\big) - v(s, X_s^{(t,x)}), \end{aligned}$$

where $(Y^{(t,x)}, Z^{(t,x)}, \Theta^{(t,x)})$ is the unique solution of the BSDE (19.7), and the equalities are in S^2, $L^2(\langle W\rangle)$ and $L^2(\langle\tilde{\mu}\rangle)$ respectively. In particular,

$$Y_t^{(t,x)} = v(t,x).$$

Proof. By applying Itô's formula to $v(s, X_s^{(t,x)})$, we have

$$\begin{aligned} &dv(s, X_s^{(t,x)}) \\ &= \Big(\frac{\partial v}{\partial t}(t, X_s^{(t,x)}) + \mathcal{L}_t v(t, X_s^{(t,x)})\Big)ds + \partial_x v(s, X_s^{(t,x)})\sigma(s, X_s^{(t,x)})dW \\ &\quad + \int_{\mathcal{Z}} \tilde{v}(\zeta; t, X_s^{(t,x)})\tilde{\mu}(d\zeta, dt). \end{aligned}$$

However, since v is the solution to the stated PIDE, it follows that

$$\begin{aligned}dv(s, X_s^{(t,x)}) = &-f(t, x, v(t, x), \partial_x v(t, x)\sigma(s, x), \tilde{v}(t, x))dt \\ &+ \partial_x v(s, X_s^{(t,x)})\sigma(s, X_s^{(t,x)})dW \\ &+ \int_{\mathcal{Z}} \tilde{v}(\zeta; t, X_s^{(t,x)})\tilde{\mu}(d\zeta, dt),\end{aligned}$$

with $v(T, X_T^{(t,x)}) = \psi(X_T^{(t,x)})$. Therefore, we observe that v gives a solution to the desired BSDE. Using the bounds in (ii) and the growth bounds on X (Lemma 17.1.1), this solution lies in the space $S^2 \times L^2(\langle W\rangle) \times L^2(\langle \mu\rangle)$. By uniqueness, we know that v and $Y^{(t,x)}$ must agree, and similarly for $Z^{(t,x)}$ and $\Theta^{(t,x)}$, in the appropriate topologies. □

We now show that, conversely, in certain cases the solution of the BSDE (19.7) yields the solution of the PIDE (19.8). In particular, we shall restrict our attention to the one-dimensional case ($m = 1$), and use the comparison theorem to show that, given our assumptions on b, σ, g, f and ψ, and supposing some additional continuity assumptions on f and ψ with respect to x, the function v is a *viscosity solution* to the PIDE. We restrict our attention to the continuous case.

Before stating this result, we recall the definition of a viscosity solution. Further details of this theory, in the continuous case (when $\mathcal{L}$ contains only differential terms), can be found in Fleming and Soner [79], Elliott [66] or Crandall, Ishii and Lions [37].

Definition 19.5.2. *Suppose $v \in C([0,T] \times \mathbb{R}^d)$ satisfies $v(T, x) = \psi(x)$ for all $x \in \mathbb{R}^d$. Then v is called a* viscosity subsolution *(resp.* viscosity supersolution*) of the PDE*

$$0 = \frac{\partial v}{\partial t}(t, x) + \mathcal{L}_t v(t, x) + f(t, x, \phi(t, x), \partial_x \phi(t, x)\sigma(s, x)), \tag{19.9}$$

where

$$\mathcal{L}_t v = \sum_i b^i(t, x)\frac{\partial v}{\partial x^i}(x) + \frac{1}{2}\sum_{i,j} a^{ij}(x)\,\frac{\partial^2 v}{\partial x^i \partial x^j}(s, x),$$

if, for each $(t, x) \in [0, T] \times \mathbb{R}^d$ and $\phi \in C^{1,2}([0,T] \times \mathbb{R}^d)$ such that

- *$\phi(t, x) = v(t, x)$ and*
- *(t, x) is a local minimum (resp. local maximum) of $\phi - v$,*

we know

$$0 \le \frac{\partial \phi}{\partial t}(t, x) + \mathcal{L}_t \phi(t, x) + f(t, x, \phi(t, x), \partial_x \phi(t, x)\sigma(s, x))$$

(resp. $0 \ge \frac{\partial \phi}{\partial t}(t, x) + \mathcal{L}_t \phi(t, x) + f(t, x, \phi(t, x), \partial_x \phi(t, x)\sigma(s, x))$).

Moreover, v is called a viscosity solution *of (19.9) if it is both a viscosity subsolution and a viscosity supersolution.*

Theorem 19.5.3. *Suppose*

- *the continuity and growth assumptions of Remark 19.4.2 hold,*
- *X is a continuous process (so $g \equiv 0$),*
- *f and ψ are uniformly continuous with respect to x.*

Then the function v defined by $v(t,x) = Y_t^{(t,x)}$ is a viscosity solution of the PDE (19.9).

Furthermore, if for each $R > 0$ there exists a continuous function $m_R : \mathbb{R}^+ \to \mathbb{R}^+$, such that

- *$m_R(0) = 0$ and*
- *for all $t \in [0,T]$, $x, x' \in \mathbb{R}^d$, $y \in \mathbb{R}$, $z \in \mathbb{R}^N$ with* $\max\{\|x\|, \|x'\|, \|z\|\} < R$, we know

$$\|f(t,x,y,z) - f(t,x',y,z)\| \leq m_R\big(\|x - x'\|(1 + \|z\|)\big),$$

then v is the unique viscosity solution of (19.9).

Proof. The continuity of the function v with respect to (t,x) follows from Theorem 19.4.4(ii). We will show that v is a viscosity subsolution of (19.9), the proof that v is a viscosity supersolution is the same.

Fix $(t,x) \in [0,T] \times \mathbb{R}^d$ and let $\phi \in C^{1,2}([0,T] \times \mathbb{R}^d)$ be such that $\phi(t,x) = v(t,x)$ and $\phi \geq v$ on $[0,T] \times \mathbb{R}^d$. By localizing in x and taking a uniformly converging approximation, we can suppose without loss of generality that ϕ is C^∞ and has bounded derivatives.

We now write, for $h \geq 0$,

$$G(s,x) := \frac{\partial \phi}{\partial s}(s,x) + \mathcal{L}_s \phi(s,x) + f(s,x,\phi(s,x), \partial_x \phi(s,x))$$

for $s \in [t, t+h]$. Our aim is to show that $G(s,x) \geq 0$. As (t,x) and ϕ were arbitrary, this will prove that v is a viscosity solution of the PDE.

We have $\phi(t+h, X_{t+h}^{(t,x)}) \geq v(t+h, X_{t+h}^{(t,x)}) = Y_t^{(t,x)} + h$. Define the processes $\{(\bar{Y}_s^h, \bar{Z}_s^h)\}_{s \in [t,t+h]}$ which solve the BSDE

$$\bar{Y}_s^h = \phi(t+h, X_{t+h}^{(t,x)}) + \int_{]s,t+h]} f(r, X_r^{(t,h)}, \bar{Y}_r^h, \bar{Z}_r^h) dr - \int_{]s,t+h]} \bar{Z}_r^h dW.$$

As $\bar{Y}^h$ and $Y^{(t,x)}$ are BSDEs with the same generator, and their terminal conditions satisfy $Y_{t+h}^{(t,x)} \leq \bar{Y}_{t+h}^h$, from the comparison theorem we have

$$\bar{Y}_t^h \geq Y_t^{(t,x)} = v(t,x) = \phi(t,x).$$

Write

$$\tilde{Y}_s^h = \bar{Y}_s^h - \phi(s, X_s^{(t,x)}) - \int_{]s,t+h]} G(r,x)dr$$

and $\tilde{Z}_s^h = \bar{Z}_s^h - \partial_x\phi(s, X_s^{(t,x)})\sigma(s, X_s^{(t,x)})$. By Itô's formula, $\{(\tilde{Y}^h, \tilde{Z}^h)\}_{s\in[t,t+h]}$ is the unique solution of the BSDE

$$\begin{aligned}\tilde{Y}_s^h = \int_{]s,t+h]} f\Big(r, X_r^{(t,x)}, \phi(r, X_r^{(t,x)}) + \tilde{Y}_r^h + \int_{]r,t+h]} G(u,x)du,\\ \tilde{Z}_r^h + \partial_x\phi(r, X_r^{(t,x)})\sigma(r, X_r^{(t,x)})\Big)dr\\ + \int_{]s,t+h]} \Big(\Big(\frac{\partial\phi}{\partial r} + \mathcal{L}_r\phi\Big)(r, X_r^{(t,x)}) - G(r,x)\Big)dr - \int_{]s,t+h]} \tilde{Z}_r dW_r.\end{aligned}$$

We now show that $(\tilde{Y}^h, \tilde{Z}^h) \to (0,0)$ as $h \to 0$. By the estimate of Theorem 19.1.6, with $(Y^1, Z^1) = (\tilde{Y}^h, \tilde{Z}^h)$ and $(Y^2, Z^2) = 0$ and $f^2 = 0$, for some $K > 0$,

$$\int_{[t,t+h]} E\big[|\tilde{Y}_s^h|^2 + \|\tilde{Z}_s^h\|^2\big]ds \le K\int_{[t,t+h]} |\delta(s,h)|^2 ds,$$

where

$$\begin{aligned}&\delta(s,h)\\ &= \Big(\frac{\partial\phi}{\partial r} + \mathcal{L}_r\phi\Big)(r, X_r^{(t,x)}) - G(r,x)\\ &\quad + f\Big(r, X_r^{(t,x)}, \phi(r, X_r^{(t,x)}) + \int_{]r,t+h]} G(u,x)du, \partial_x\phi(r, X_r^{(t,x)})\sigma(r, X_r^{(t,x)})\Big).\end{aligned}$$

As $\sup_{s\in[t,t+h]} E\big[\|X_s^{(t,x)} - x\|^2\big] \to 0$ as $h \to 0$ (from Exercise 17.5.3) and since all the coefficients and ϕ and its derivatives are uniformly continuous with respect to x, it follows that

$$\lim_{h\downarrow 0} \sup_{s\in[t,t+h]} E\big[|\delta(s,h)|^2\big] = 0.$$

Therefore, we see

$$\int_{[t,t+h]} E\big[|\tilde{Y}_s^h|^2 + \|\tilde{Z}_s^h\|^2\big]ds \le K\int_{[t,t+h]} |\delta(s,h)|^2 ds \le h\epsilon(h)$$

for some function ϵ with $\epsilon(h) \to 0$ as $h \to 0$. Taking an expectation, we also see

$$\bar{Y}_t^h - \phi(t,x) - \int_{]s,t+h]} G(r,x)dr = \tilde{Y}_t^h = E\Big[\int_{]t,t+h]} \delta'(r,h)dr\Big],$$

where

$$\begin{aligned}\delta'(s,h) = \Big(\frac{\partial\phi}{\partial r} + \mathcal{L}_r\phi\Big)(r, X_r^{(t,x)}) - G(r,x) \\ + f\Big(r, X_r^{(t,x)}, \phi(r, X_r^{(t,x)}) + \tilde{Y}_r^h + \int_{]r,t+h]} G(u,x)du, \\ \tilde{Z}_r^h + \partial_x\phi(r, X_r^{(t,x)})\sigma(r, X_r^{(t,x)})\Big).\end{aligned}$$

Since f is Lipschitz, we also know that $|\delta(r,h) - \delta'(r,h)| \leq K(|\tilde{Y}_r| + \|\tilde{Z}_r\|)$. It follows that $\tilde{Y}_t^h = h\epsilon(h)$ for some ϵ with $\epsilon(h) \to 0$ as $h \to 0$, and, as $\bar{Y}_t^h \geq \phi(t,x)$, we have

$$\int_{[t,t+h]} G(r,x)dr \geq -h\epsilon(h).$$

Dividing by h and letting $h \to 0$, we obtain

$$G(t,x) = \frac{\partial\phi}{\partial s}(s,x) + \mathcal{L}_s\phi(s,x) + f(s,x,\phi(s,x),\partial_x\phi(s,x)) \geq 0,$$

so v is a viscosity subsolution of (19.9).

The uniqueness statement is then simply the statement that the PDE (19.9) has at most one viscosity solution, given the stated growth bounds. This result can be found in Ishii and Lions [100]. □

Remark 19.5.4. Philosophically, the relationship between viscosity solutions and BSDEs is also well founded in the non-local case (where X may jump), provided f is a balanced driver. The difficulty is that, without stronger assumptions, we cannot guarantee that ϕ lies in the domain of $\mathcal{L}_t$ (i.e. $\phi \in C^{1,2}_\nu$), and the integral term in $\mathcal{L}_t\phi$ may not be well defined. It is possible to overcome this difficulty, though the theory becomes a little more involved. See Barles, Buckdahn and Pardoux [3] for details in a slightly restricted setting.

Remark 19.5.5. In the case where all the coefficients are C^3, and the filtration is generated by a finite-dimensional Brownian motion, Pardoux and Peng [146] show that the BSDE solution $v(t,x) = Y_t^{(t,x)}$ belongs to $C^{1,2}([0,T]\times\mathbb{R}^d, \mathbb{R}^m)$, and is therefore a classical solution to the corresponding PDE.

Remark 19.5.6. We have here supposed throughout that the forward process X is finite dimensional. This can be relaxed, but the notion of solution to the PDEs considered becomes more delicate (one cannot readily use viscosity solutions, given the infinite dimensional nature of the problem). See Confortola [36] for one approach in this setting. The Markov assumption can also be relaxed, and this leads to the theory of *path dependent* PDEs, a special case of those considered by Ekren, Keller, Touzi and Zhang [64].

19.6 Exercises

Exercise 19.6.1. Suppose that f is standard, balanced and convex in (y, z, θ) and that $(Y^\xi, Z^\xi, \Theta^\xi)$ solves the BSDE with data (ξ, f). Show that the map $\xi \mapsto Y_t^\xi$ is convex for all t.

Exercise 19.6.2. Let f be a standard and balanced driver for a BSDE. Let Y^ξ be as in Exercise 19.6.1.

(i) Show that if $f(\omega, t, y, 0, 0) = 0$, then for any $\mathcal{F}_t$-measurable ξ, we have $Y_t^\xi = \xi$.
(ii) Show that if f does not depend on y and satisfies (i), then for any $\mathcal{F}_t$-measurable ξ and any $\mathcal{F}_T$-measurable ξ', we have $Y_t^{\xi+\xi'} = Y_t^{\xi'} + \xi$.
(iii) Show that if f is positively homogenous, that is for any $\lambda > 0$ we have $f(\omega, t, \lambda y, \lambda z, \lambda\theta) = \lambda f(\omega, t, y, z, \theta)$ then $Y^{\lambda\xi} = \lambda Y^\xi$

Exercise 19.6.3. Let f and f' be balanced convex drivers for a BSDE. Suppose for simplicity that f and f' do not depend on y. Define the inf-convolution

$$\tilde{f}(z, \theta) = \inf\left\{f(z - z', \theta - \theta') + f'(z', \theta') : z' \in \mathbb{R}^N, \theta' \in L^2(\nu)\right\}.$$

(i) Show that $\tilde{f}$ is convex and balanced.
(ii) For a given terminal value $\tilde{\xi} \in L^2$, let $\tilde{Y}$ be the solution to the BSDE with driver $\tilde{f}$. Show that

$$\tilde{Y} \leq \inf\{Y^{\xi-\xi'} + Y^{\xi'} : \xi' \in L^2\}.$$

(iii) Assuming there exists $(\hat{Z}, \hat{\Theta}) \in L^2(\langle W\rangle) \times L^2(\langle\tilde{\mu}\rangle)$ such that $\tilde{f}(Z_t, \Theta_t) = f(Z - \hat{Z}, \Theta - \hat{\Theta}) + f'(\hat{Z}, \hat{\Theta})$, show that $\tilde{Y} = \inf\{Y^{\xi-\xi'} + Y^{\xi'} : \xi' \in L^2\}$.

This question is based on ideas from Barrieu and El Karoui [5]

Exercise 19.6.4. Suppose we are in the pure jump setting (where there is no diffusion term), and $\nu(\mathcal{Z}) < \infty$. Let f be the linear BSDE driver

$$f(\omega, t, y, \theta) = \alpha \int_{\mathcal{Z}} \theta(\zeta)\nu(d\zeta),$$

for $\alpha \in \mathbb{R}$.

(i) Show that, for any $\xi \in L^2(\mathcal{F}_T)$, the corresponding BSDE has a solution.
(ii) Let $\xi = \mu([0, T] \times \mathcal{Z})$, the random variable which counts the number of jumps (of any size) before time T. Give an explicit solution to the BSDE with this terminal value.
(iii) Show that, if $\alpha = -1$, then the strict comparison theorem fails, and if $\alpha < -1$, the comparison theorem fails.

Exercise 19.6.5. Consider the scalar BSDE with driver

$$f(\omega, t, y, z, \theta) = -\alpha y + g(\omega, t, z, \theta),$$

where g is a Lipschitz balanced function and $|g(\omega, t, 0, 0)| < C$ for some $C \in \mathbb{R}$. Let $Y^{(T)}$ be the solution to the BSDE with terminal value $Y_T^{(T)} = 0$.

(i) Show that $|Y^{(T)}| \leq C/\alpha$ for all T, by considering $e^{-\alpha t}|Y_t^{(T)}|$.
(ii) Show that $Y^{(T)}$ converges ucp as $T \to \infty$, by considering the process $e^{-\alpha t}|Y_t^{(T)} - Y_t^{(T')}|$, for $T' > T$.
(iii) Show that there is exactly one solution (Y, Z, Θ) with Y bounded to the equation

$$dY_t = -f(\omega, t, Y_{t-}, Z_s, \Theta_s)dt + Z_t dW_t + \int_{\mathcal{Z}} \Theta_t(\zeta)\tilde{\mu}(d\zeta, dt)$$

by considering $e^{-\alpha t}|Y_t - Y_t'|$ for two solutions Y, Y'.

Hint: In each of these arguments, you first should change measure to remove a drift term, then apply Tanaka's formula. This argument is based on that in Royer [161].

Part V

Applications

20

Control of a Single Jump

In this and the coming chapter, we use the mathematical machinery we have developed to consider problems related to the optimal control of a random process. To begin with, we consider the simple case of a single jump process, as in Chapter 13, where a controller can determine the rate at which the jump occurs, but faces some cost for doing so. This example will allow us to demonstrate the main methods used in optimal control, before moving on to more technically demanding problems.

It appears necessary in these situations to require all measures corresponding to admissible controls to be mutually absolutely continuous. Otherwise, various conditional expectations are defined only up to control-dependent sets. We first discuss the form of the absolutely continuous measures describing the process, and prove that the absolute continuity of the measures implies the absolute continuity of the Lévy systems and, more significantly, that the converse holds.

The results of this section were first presented in the paper [49] of M.H.A. Davis and one of the authors, extending earlier work of Boel and Varaiya [20]. Pliska [150] has obtained related results in terms of infinitesimal generators for the optimal control of Markov jump processes. The form of the infinitesimal generators is closely related to that of the Hamiltonian function $H(t, u)$ of Corollary 20.3.8 below, a connection which will become more pronounced in Chapter 21.

20.1 Describing Measure Changes

We begin by describing the single jump process, and formulating a method to describe changes of measure in this setting. This section is necessarily formal; the reader familiar with Lévy systems for random measures and willing to accept that these can be modified via measure changes can skip directly to Section 20.2.

S.N. Cohen, R.J. Elliott, *Stochastic Calculus and Applications*,
Probability and Its Applications, DOI 10.1007/978-1-4939-2867-5_20

Consider a single jump process, as described in Chapter 13. That is, we consider a process $\{z_t\}_{t\geq 0}$, with values in a Blackwell space $(\mathcal{Z}, \mathfrak{Z})$ which remains at its initial point $z_0 \in \mathcal{Z}$ until a random time T, when it jumps to a new random position z. The underlying probability space is again taken to be

$$\Omega = [0, \infty] \times \mathcal{Z},$$

with the σ-algebra $\mathcal{B} \otimes \mathfrak{Z}$. A sample path of the process is

$$z_t(\omega) = \begin{cases} z_0 & \text{if } t < T(\omega), \\ z(\omega) & \text{if } t \geq T(\omega). \end{cases}$$

A probability measure P is given on $(\Omega, \mathcal{B} \otimes \mathfrak{Z})$, and we write $\hat{P}$ when we think of P as a measure on the concrete space $[0, \infty] \times \mathcal{Z}$. We suppose

$$\begin{aligned} 0 &= P(z(\omega) = z_0) = \hat{P}([0, \infty] \times \{z_0\}), \\ 0 &= P(T = 0) = \hat{P}(\{0\} \times \mathcal{Z}). \end{aligned}$$

The definitions and notation from the first section of Chapter 13 for the single jump process will now be used without further explanation.

In the sequel we shall discuss the same concepts associated with a second measure $\overline{P}$ on Ω. Such functions, etc., will be denoted by $\overline{F}_t$, $\overline{\lambda}$ and so on.

Suppose $\overline{P}$ is absolutely continuous with respect to P. Then there is a Radon–Nikodym derivative $L = d\overline{P}/dP$. Write $L_t = E[L|\mathcal{F}_t]$. Exploiting the fact that $\Omega = [0, \infty] \times \mathcal{Z}$, we can write the random variable L as $L(\omega) = \hat{L}(T, z)$, so, from Lemma 13.1.13,

$$\begin{aligned} L_t &= I_{\{t \geq T\}} \hat{L}(T, z) + I_{\{t < T\}} F_t^{-1} \int_{]t, \infty] \times \mathcal{Z}} \hat{L}(s, \zeta) \mu(ds, d\zeta) \\ &= \hat{L}(T, z) I_{\{t \geq T\}} + I_{\{t < T\}} (\overline{F}_t / F_t). \end{aligned}$$

From Theorem 13.1.15, the martingale L has a representation as a stochastic integral

$$L_t = 1 + \int_{\mathcal{Z} \times [0, t]} g(s, \zeta) \tilde{\mu}(d\zeta, ds), \tag{20.1}$$

where

$$\begin{aligned} g(s, \zeta) &= \hat{L}(s, \zeta) - (\overline{F}_s / F_s) I_{\{s < c\}}, \\ \tilde{\mu}(A, t) &= \mu(A, t) - \mu_p(A, t) = I_{\{t \geq T\}} I_{\{z \in A\}} - \int_{[0, t \wedge T]} \lambda(A, s) d\Lambda(s). \end{aligned}$$

With $\overline{c} = \inf\{t : \overline{F}_t = 0\}$ we see $\overline{c} \leq \infty$ and set $g(s, \zeta) = 0$ for $s > \overline{c}$.

Theorem 20.1.1. *Write*

$$\phi(s,\zeta) = \begin{cases} (F_{s-}/\overline{F}_{s-})\big(\hat{L}(s,\zeta) - (\overline{F}_s/F_s)\big) & \textit{for } s < \overline{c}, \\ 0 & \textit{for } s > \overline{c}, \\ 0 & \textit{for } s = \overline{c} \textit{ if } \overline{F}_{\overline{c}-} = 0 \textit{ or } \overline{c} = \infty, \\ (F_{\overline{c}-}/\overline{F}_{\overline{c}-})\hat{L}(\overline{c},\zeta) & \textit{for } s = \overline{c} \textit{ if } \overline{c} < \infty \textit{ and } \overline{F}_{\overline{c}-} \neq 0. \end{cases}$$

Then

$$\begin{aligned} L_t &= \mathcal{E}(\phi * \tilde{\mu})_t \\ &= \exp\Big(- \int_{[0,t\wedge T]\times \mathcal{Z}} \phi(s,\zeta)\lambda(d\zeta,s)d\Lambda_s \Big) \\ &\quad \times \Big(1 + \phi(T,z)I_{\{t\geq T\}} + I_{\{t\geq T\}} \int_{\mathcal{Z}} \phi(T,\zeta)\lambda(d\zeta,T)\, \frac{\Delta F_T}{F_{T-}} \Big) \\ &\quad \times \prod_{\substack{u\leq t\wedge T \\ u\neq T}} \Big(1 + \int_{\mathcal{Z}} \phi(u,\zeta)\lambda(d\zeta,u)\, \frac{\Delta F_u}{F_{u-}} \Big). \end{aligned}$$

where $\mathcal{E}$ is the Doléans-Dade exponential.

Proof. If $s \leq T$ then $L_{s-} = \overline{F}_{s-}/F_{s-}$. Therefore, for $s \leq T$ and $s < \overline{c}$ (and, if $\overline{F}_{\overline{c}-} \neq 0$, for $s \leq T \wedge \overline{c}$), we know $L_{s-} > 0$ so

$$\phi(s,\zeta) = g(s,\zeta)L_{s-}^{-1} = (F_{s-}/\overline{F}_{s-})\big(\hat{L}(s,\zeta) - (\overline{F}_s/F_s)\big).$$

Writing

$$\begin{aligned} M_t := (\phi * \tilde{\mu})_t &= \int_{[0,t]\times\mathcal{Z}} \phi(s,\zeta)\tilde{\mu}(ds,d\zeta) \\ &= \phi(T,z)I_{t\geq T} + \int_{[0,t\wedge T]\times\mathcal{Z}} \phi(s,\zeta)\lambda(d\zeta,s)d\Lambda_s, \end{aligned}$$

we notice that (20.1) can be written

$$L_t = 1 + \int_0^t L_{s-}dM_s\,.$$

Consequently, by the exponential formula of Theorem 15.1.8, as M is a pure jump martingale,

$$L_t = \mathcal{E}(M)_t = e^{M_t}\prod_{u\leq t}(1+\Delta M_u)e^{-\Delta M_u}.$$

At the discontinuities of F, we know

$$\Delta M_u = \int_{\mathcal{Z}} \phi(u,\zeta)\lambda(d\zeta,u)\, \frac{\Delta F_u}{F_{u-}},$$

provided $u \neq T$, and at the jump time T,

$$\Delta M_T = \phi(T,\zeta) + \int_{\mathcal{Z}} \phi(T,\zeta)\lambda(d\zeta,T)\,\frac{\Delta F_T}{F_{T-}}\,.$$

Substituting, we see that L has the stated form. □

We now see how a change of measure from P to $\overline{P}$ impacts the Lévy system of the random measure.

Theorem 20.1.2. *Suppose $(\overline{\lambda},\overline{\Lambda})$ is the Lévy system of μ under $\overline{P}$. Then, writing*

$$\gamma(A,s) = \int_A \Big(1+\phi+\,\frac{\Delta F_s}{F_{s-}}\,\int_{\mathcal{Z}}\phi d\lambda\Big)d\lambda$$

for $A \in \mathfrak{Z}$, we know that, $d\overline{F}$-a.e.,

$$\overline{\lambda}(A,s) = \gamma(A,s)/\gamma(\mathcal{Z},s) \qquad \text{and} \qquad \overline{\Lambda}_t = \int_{]0,t]}\gamma(\mathcal{Z},s)d\Lambda_s\,.$$

Proof. For $t>0$ and $A\in\mathfrak{Z}$,

$$\begin{aligned}\overline{F}{}^A_t = \overline{P}(T>t, z\in A) = \hat{\overline{P}}(\,]t,\infty]\times A) &= \int_{]t,\infty]\times A}\hat{L}(s,\zeta)\,d\hat{P}\\ &= -\int_{]t,\infty]}\int_A L(s,x)\lambda(dx,s)dF_s\,.\end{aligned}$$

However,

$$\overline{F}{}^A_t = -\int_{]t,\infty]}\overline{\lambda}(A,s)d\overline{F}_s = -\int_{]t,\infty]}\overline{\lambda}(A,s)\,\frac{d\overline{F}_s}{dF_s}\,dF_s\,.$$

So, dF-a.e.,

$$\overline{\lambda}(A,s)\,\frac{d\overline{F}_s}{dF_s} = \int_A L(s,\zeta)\lambda(dx,s) = \int_A\Big(\frac{\overline{F}_{s-}}{F_{s-}}\,\phi(s,\zeta)+\frac{\overline{F}_s}{F_s}\Big)\lambda(d\zeta,s).$$

Therefore, for $s<\overline{c}$ (and, if $\overline{F}_{\overline{c}-}\neq 0$, for $s\le\overline{c}$), $d\overline{F}$-a.e.,

$$\begin{aligned}\overline{\lambda}(A,s)\,\frac{F_s}{\overline{F}_{s-}}\,\frac{d\overline{F}_s}{dF_s} &= \int_A\Big(\frac{F_s}{F_{s-}}\,\phi+\frac{\overline{F}_s}{\overline{F}_{s-}}\Big)\lambda(d\zeta,s)\\ &= \int_A\Big(\Big(1+\frac{\Delta F_s}{F_{s-}}\Big)\phi+\Big(1+\frac{\Delta\overline{F}_s}{\overline{F}_{s-}}\,\Big)\Big)\lambda(d\zeta,s).\end{aligned} \tag{20.2}$$

If s is a point of continuity of F, then it is also a point of continuity of $\overline{F}$ and $\Delta F_s = \Delta\overline{F}_s = 0$. If $\Delta F_s \neq 0$ then the Radon–Nikodym derivative gives $d\overline{F}/dF(s) = \Delta\overline{F}_s/\Delta F_s$ and the left-hand side of (20.2) is

$$\overline{\lambda}(A,s)\ \frac{(F_{s-}+\Delta F_s)}{\overline{F}_s}\ \frac{\Delta \overline{F}_s}{\Delta F_s} = \overline{\lambda}(A,s)\ \frac{\Delta \overline{F}_s}{\overline{F}_{s-}}\ \Big(1+\frac{F_{s-}}{\Delta F_s}\Big).$$

Evaluating (20.2) with $A=\mathcal{Z}$, so $\overline{\lambda}(\mathcal{Z},s)=1=\lambda(\mathcal{Z},s)$, we find

$$\frac{\Delta \overline{F}_s}{\overline{F}_{s-}} = \frac{\Delta F_s}{F_{s-}} \int_{\mathcal{Z}} \Big(1+\phi+\frac{\Delta F_s}{F_{s-}}\,\phi\Big)\lambda(d\zeta,s) \qquad \text{if } \Delta F_s \neq 0 \tag{20.3}$$

and

$$\frac{F_s}{\overline{F}_{s-}}\ \frac{d\overline{F}}{dF}(s) = \int_{\mathcal{Z}} (1+\phi)\lambda(d\zeta,s) \qquad \text{if } \Delta F_s = 0.$$

Substituting in (20.2) and rearranging, we have, if $\big(1+(\Delta F_s/F_{s-})\big)\neq 0$,

$$\begin{aligned}&\overline{\lambda}(A,s)\\ &= \int_A \Big(1+\phi+\frac{\Delta F_s}{F_{s-}}\int_{\mathcal{Z}}\phi\,d\lambda\Big)\lambda(d\zeta,s)\Big/\int_{\mathcal{Z}}\Big(1+\phi+\frac{\Delta F_s}{F_{s-}}\int_{\mathcal{Z}}\phi d\lambda\Big)\lambda(d\zeta,s)\\ &= \gamma(A,s)/\gamma(\mathcal{Z},s)\end{aligned}$$

$d\overline{F}_s$-a.e. for $s<\overline{c}$, and for $s\le \overline{c}$ if $\overline{F}_{\overline{c}-}\neq 0$.

Now $\big(1+(\Delta F_s/F_{s-})\big)=0$ only if $s=c$, $c<\infty$ and $F_{c-}\neq 0$. This situation is only of interest here if also $\overline{c}=c$ and $\overline{F}_{c-}\neq 0$. However, in this case it is easily seen that substituting $\phi(c,x)=(F_{c-}/\overline{F}_{c-})L(c,x)$ in (20.2) gives the correct expression for $\overline{\lambda}(A,c)=\lambda(A,c)$, because

$$L(c,x) = (\Delta \overline{F}_c/\Delta F_c)\big(d\overline{\lambda}/d\lambda(c)\big).$$

Therefore, we have the desired representation of $\overline{\lambda}$.

We now find $\overline{\Lambda}$. We have

$$\overline{\Lambda}_t = -\int_{]0,t]} \frac{d\overline{F}_s}{\overline{F}_{s-}} = \int_{]0,t]} \frac{F_{s-}}{\overline{F}_{s-}}\ \frac{d\overline{F}}{dF}(s)\ d\Lambda_s\,.$$

If F is continuous at s, again $\Delta\overline{F}_s=\Delta F_s=0$ and, evaluating (20.2) for $A=\mathcal{Z}$,

$$\frac{d\overline{\Lambda}}{d\Lambda}(s) = \frac{F_s}{\overline{F}_s}\frac{d\overline{F}}{dF}(s)\ = \int_{\mathcal{Z}}(1+\phi)\lambda(d\zeta,s).$$

If F is not continuous at s, then $d\overline{F}_s/dF_s=\Delta\overline{F}_s/\Delta F_s$ and, from (20.3),

$$\frac{d\overline{\Lambda}_s}{d\Lambda_s} = \frac{\Delta\overline{F}_s}{\Delta F_s}\ \frac{F_{s-}}{\overline{F}_{s-}} = \int_{\mathcal{Z}}\Big(1+\phi+\frac{\Delta F_s}{F_{s-}}\,\phi\Big)\lambda(d\zeta,s).$$

As $\int_{\mathcal{Z}}\phi d\lambda = \int_{\mathcal{Z}}(\int_{\mathcal{Z}}\phi d\lambda)d\lambda$, we have

$$\overline{\Lambda}_t = \int_{[0,t]}\int_{\mathcal{Z}}\Big(1+\phi+\frac{\Delta F_s}{F_{s-}}\int_{\mathcal{Z}}\phi d\lambda\Big)\lambda(d\zeta,s)d\Lambda_s = \int_{]0,t]}\gamma(\mathcal{Z},s)d\Lambda_s\,.$$

$\square$

Notation 20.1.3. We denote by $\mathcal{K}$ the set of all deterministic integrated rate functions, that is, the right continuous, nondecreasing (deterministic) functions $\Lambda : [0, \infty] \to [0, \infty]$ such that

(i) $\Lambda_0 = 0$,
(ii) $\Delta\Lambda_u = \Lambda_u - \Lambda_{u-} \leq 1$ for all points of discontinuity u,
(iii) if $\Delta\Lambda_u = 1$ then $\Lambda_t = \Lambda_u$ for $t \geq u$.

If $\Lambda \in \mathcal{K}$ then Λ has a unique decomposition $\Lambda = \Lambda^c + \Lambda^d$, where $\{\Lambda^d_t = \sum_{s \leq t} \Delta\Lambda_s\}_{t \geq 0} \in \mathcal{K}$ and $\Lambda^c \in \mathcal{K}$ is continuous. Note that Λ_t may equal $+\infty$ for finite t.

Lemma 20.1.4. *The formulae*

$$\begin{cases} F_t = \exp(-\Lambda^c_t) \prod_{u \leq t} (1 - \Delta\Lambda_u), & \text{(20.4a)} \\ \Lambda_t = -\int_{[0,t]} F^{-1}_{s-} dF_s\,, & \text{(20.4b)} \end{cases}$$

define a bijection (and its inverse) between the set $\mathcal{K}$ *and the set of all probability distributions on* $[0, \infty]$, *parameterized by their cumulative distribution functions* $G_t = 1 - F_t$.

Proof. Clearly, if $\Lambda \in \mathcal{K}$, then F_t defined by (20.4a) is monotonic decreasing and right continuous, $F_0 = 1$ and $0 \leq F_t \leq 1$. Therefore, $G_t = 1 - F_t$ is a cumulative probability distribution on $[0, \infty]$.

Conversely, if G is a cumulative probability distribution function, then for $F_t = 1 - G_t$ and Λ_t given by (20.4b), it follows that $\Lambda \in \mathcal{K}$.

From Theorem 15.1.8, if $\Lambda \in \mathcal{K}$, then F defined by (20.4a) is the unique solution of the equation

$$\begin{cases} dF_t = -F_{t-} d\Lambda_t\,, \\ F_0 = 1. \end{cases}$$

This shows the stated formulae define a bijection. □

Remark 20.1.5. If $\Lambda^d \equiv 0$ and Λ^c is absolutely continuous with respect to Lebesgue measure, then there is a measurable function α such that

$$\Lambda^c_t = \int_{[0,t]} \alpha_s ds.$$

The function α_s is often called the "rate" of the jump process. However, there are continuous increasing functions which are singular with respect to Lebesgue measure, so to discuss the optimal control of the single jump process we suppose a general "(integrated) base rate" $\Lambda = \Lambda^c + \Lambda^d$ is given.

Lemma 20.1.6. *Suppose $\overline{\Lambda} \in \mathcal{K}$ is a second process whose associated Stieltjes measure is absolutely continuous with respect to (the Stieltjes measure associated with) Λ. Then the associated $\overline{F}_t$ has the form*

$$\overline{F}_t = F_t \exp\Big(-\int_{[0,t]} (\alpha_s - 1) d\Lambda_s^c\Big)\Big(\prod_{u\le t} \frac{\big(1-\alpha_u \Delta\Lambda_u^d\big)}{(1-\Delta\Lambda_u^d)}\Big).$$

Here $\alpha = d\overline{\Lambda}/d\Lambda$ is the Radon–Nikodym derivative and F is defined by (20.4a).

Furthermore, $\alpha_u \Delta\Lambda_u^d \le 1$ and, if $\alpha_u\Delta\Lambda_u^d = 1$, then $\alpha_t = 0$ for all $t \ge u$.

Proof. By hypothesis,

$$\overline{\Lambda}_t = \int_{[0,t]} \alpha_s d\Lambda_s^c + \sum_{u\le t} \alpha_u \Delta\Lambda_u^d\,,$$

so, from (20.4a),

$$\begin{aligned}\overline{F}_t &= \exp\Big(-\int_{[0,t]} \alpha_s d\Lambda_s^c\Big)\prod_{u\le t}\big(1-\alpha_u\Delta\Lambda_u^d\,\big)\\ &= F_t \exp\Big(-\int_{[0,t]} (\alpha_s - 1) d\Lambda_s^c\Big)\prod_{u\le t} \frac{\big(1-\alpha_u \Delta\Lambda_u^d\big)}{(1-\Delta\Lambda_u^d)}\,.\end{aligned}$$

The conditions on α follow from Lemma 20.1.4 and the definition of $\mathcal{K}$. $\square$

Remark 20.1.7. Because

$$-d\overline{F}_t = \overline{F}_{t-}d\overline{\Lambda}_t = (\overline{F}_t/F_{t-})\alpha_t F_{t-}d\Lambda_t = -(\overline{F}_{t-}/F_{t-})\alpha_t dF_t,$$

the probability distribution associated with $\overline{F}$ above is certainly absolutely continuous with respect to that associated with F. To ensure the converse, it would be sufficient to require that, for some positive integer n,

$$1/n \le \alpha_s \le \min\big(n, (n^{-1}-1)F_{s-}/\Delta F_s\big)$$

for all s.

The above discussion only concerns the rate Λ, describing *when* the jump occurs. Consider now the other component λ of the Lévy system, which describes *where* the jump goes. Because $(\mathcal{Z}, \mathfrak{Z})$ is a Blackwell space, the measures $\lambda(\cdot, s)$ can be chosen to be a regular family of conditional probability distributions, and so will satisfy

(i) $\lambda(A, s) \ge 0$ for $A \in \mathfrak{Z}$, $\quad s > 0$,
(ii) for each $A \in \mathfrak{Z}$ we know $\lambda(A,\,\cdot)$ is Borel measurable,
(iii) for all $s \in \,]0, c[$ (except perhaps on a set of $d\Lambda$-measure 0), we know $\lambda(\,\cdot, s)$ is a probability measure on $(\mathcal{Z}, \mathfrak{Z})$ and, if $c < \infty$ and $\Lambda_{c-} < \infty$, then $\lambda(\,\cdot, c)$ is a probability measure.

Lemma 20.1.8. *There is a bijection between probability measures μ on $(\Omega, \mathcal{B} \times \mathfrak{Z})$ and Lévy systems (λ, Λ), where λ satisfies (i – iii) above and $\Lambda \in \mathcal{K}$.*

Proof. Definition 13.1.4 indicates how a Lévy system is determined by a measure μ. Conversely, given a pair (λ, Λ), because $\Lambda \in \mathcal{K}$ we can determine a function F by (20.4b). Then for $A \in \mathfrak{Z}$, we define

$$P(T \le t, z \in A) = \hat{P}(\,]0,t] \times A) = -\int_{]0,t]} \lambda(A,s)dF_s\,.$$

$\square$

We now establish the converse of Theorem 20.1.2, that is, if the Lévy systems of two measures $\overline{P}, P$, on $(\Omega, \mathcal{B} \times \mathfrak{Z})$ are absolutely continuous, then the measures are absolutely continuous.

Theorem 20.1.9. *Suppose $P, \overline{P}$ have Lévy systems (λ, Λ) and $(\overline{\lambda}, \overline{\Lambda})$. Write $\overline{c} = \inf\{t : \overline{F}_t = 0\}$, and suppose $\overline{c} \le c$, $d\overline{\Lambda} \ll d\Lambda$ on $[0, \overline{c}]$ and $\overline{\lambda}(\cdot,t) \ll \lambda(\cdot,t)$, at least $d\Lambda$-a.e. Then $\overline{P} \ll P$, with Radon–Nikodym derivative*

$$\frac{d\overline{P}}{dP}(t,\zeta) = \hat{L}(t,\zeta) = \alpha_t \beta(t,\zeta) \exp\Big(-\int_{[0,t]} (\alpha_s - 1) d\Lambda_s^c\Big) \Pi_{t-} I_{\{t \le \overline{c}\}}\,,$$

where

$$\Pi_t = \prod_{u \le t} \frac{\big(1 + (\Delta F_u / F_{u-})\alpha_u\big)}{\big(1 + (\Delta F_u / F_{u-})\big)}\,.$$

Here

$$\alpha_t = \frac{d\overline{\Lambda}}{d\Lambda}(t) \quad \textit{and} \quad \beta(t,\zeta) = \frac{d\overline{\lambda}}{d\lambda}(t,\zeta).$$

Proof. Define $\hat{L}(t,\zeta)$ by the above expression and write

$$\eta(t) = \exp\Big(-\int_{[0,t]} (\alpha_s - 1) d\Lambda_s^c\Big).$$

Then, because $\int_{\mathcal{Z}} \beta(t,\zeta) d\lambda = 1$ a.s.,

$$E[\hat{L}(T,z)] = -\int_{[0,\overline{c}]} \alpha_t \eta(t) \Pi_{t-} dF_t\,.$$

From Lemma 20.1.6 and equations (20.4a) and (20.4b), $\eta(t)\Pi_{t-} = \overline{F}_{t-}/F_{t-}$. As measures on $[0,\infty]$,

$$d\overline{\Lambda}_t = -\frac{d\overline{F}_t}{\overline{F}_{t-}} = -\alpha_t \frac{dF_t}{F_{t-}} = \alpha_t d\Lambda_t,$$

so

$$E[L(T,z)] = -\int_{[0,\overline{c}]} \alpha_t \overline{F}_{t-} \frac{dF_t}{F_{t-}}$$
$$= -\int_{[0,\overline{c}]} \overline{F}_{t-} \frac{d\overline{F}_{t-}}{\overline{F}_{t-}} = \overline{F}_{0-} - \overline{F}_{\overline{c}} = 1.$$

A probability measure $P^* \ll P$ can, therefore, be defined on $(\Omega, \mathcal{B} \times \mathfrak{Z})$ by $dP^*/dP = \hat{L}(T,z)$. For $t < \overline{c}$ we have

$$L_t = E[\hat{L}(T,z)|\mathcal{F}_t] = \hat{L}(T,z)I_{\{t \geq T\}} + I_{\{t<T\}} F_t^{-1} \int_{]t,\overline{c}] \times \mathcal{Z}} \hat{L}(s,\zeta)\mu(ds, d\zeta).$$

By similar calculations to those above, the last term can be written

$$F_t^{-1} \int_{]t,\overline{c}]} \alpha_s \eta(s) \Pi_{s-} dF_s = \overline{F}_t / F_t = \eta(t)\Pi_t \,,$$

so

$$L_t = \alpha_T \beta(T,z)\eta(T)\Pi_{T-} I_{\{t \geq T\}} + I_{\{t<T\}}\, \eta(t)\Pi_t \,.$$

In the notation of Theorem 20.1.1,

$$\phi(t,\zeta) = \begin{cases} \frac{F_{t-}}{\overline{F}_{t-}}\Big(\hat{L}(t,\zeta) - \frac{\overline{F}_t}{F_t}\Big) & \text{if } t < \overline{c}, \\ 0 & \text{if } t > \overline{c}, \\ 0 & \text{if } t = \overline{c} \text{ and } \overline{c} = \infty \text{ or } \overline{F}_{\overline{c}-} = 0, \\ \hat{L}(\overline{c},\zeta) = \alpha(\overline{c})\beta(\overline{c},\zeta) & \text{if } \overline{c} < \infty \text{ and } \overline{F}_{\overline{c}-} \neq 0, \end{cases}$$

and, in the case $t < \overline{c}$, this simplifies to

$$\phi(t,\zeta) = \frac{F_{t-}}{\overline{F}_{t-}}\Big(\hat{L}(t,\zeta) - \frac{\overline{F}_t}{F_t}\Big)$$
$$= \big(\alpha(t)\beta(t,\zeta)\Pi_{t-}\eta(t)I_{\{t \leq \overline{c}\}} - \eta(t)\Pi_t\big)\eta(t)^{-1}\Pi_{t-}^{-1}$$
$$= \alpha(t)\beta(t,\zeta)I_{\{t \leq \overline{c}\}} - \frac{\big(1 + \Delta F_t/F_{t-}\alpha(t)\big)}{(1 + \Delta F_t/F_{t-})}.$$

The Lévy system (λ^*, Λ^*) associated with P^* is given by

$$\lambda^*(A,s) = \gamma(A,s)/\gamma(\mathcal{Z},s), \qquad \Lambda_t^* = \int_{]0,t]} \gamma(\mathcal{Z},s)d\Lambda_s,$$

where

$$\gamma(A,s) = \int_A \Big(1 + \phi + \frac{\Delta F_s}{F_{s-}} \int_{\mathcal{Z}} \phi d\lambda\Big) d\lambda.$$

Substituting the above expressions for ϕ, we have

$$\int_{\mathcal{Z}} \phi \, d\lambda = \alpha_t I_{\{t \leq \overline{c}\}} - \frac{\big(1 + (\Delta F_t/F_{t-})\alpha(t)\big)}{(1 + \Delta F_t/F_{t-})}$$

and

$$\Big(1 + \frac{\Delta F_t}{F_{t-}}\Big) \int_{\mathcal{Z}} \phi \, d\lambda = \alpha_t - 1.$$

It follows that

$$\frac{d\Lambda_t^*}{d\Lambda_t} = \alpha(t) \quad \text{and} \quad \frac{d\lambda^*}{d\lambda} = \beta(t, \zeta),$$

so $\Lambda^* = \overline{\Lambda}$ and $\lambda^* = \overline{\lambda}$. By Lemma 20.1.8, $\overline{P} = P^* \ll P$ and the result is proven. □

20.2 The Control Problem

We now consider a jump process control problem by supposing the process $\{z_t\}_{t\geq 0}$ is governed by an indexed family of probability measures $\{P^u : u \in \mathcal{U}\}$ on $(\Omega, \mathcal{B}\times\mathfrak{Z})$. A controller chooses a control u so as to minimize some expected cost function, where the expectation is taken under P^u.

We assume all the measures P^u are absolutely continuous with respect to some "base measure" P. If (λ, Λ) and (λ^u, Λ^u) are, respectively, the Lévy systems associated with P and P^u then $\lambda^u \ll \lambda$ and $d\Lambda^u \ll d\Lambda$. Conversely, if λ^u and Λ^u are given and $\lambda^u \ll \lambda$, $d\Lambda^u \ll d\Lambda$, then Theorem 20.1.9 gives the formula for dP^u/dP, showing that $P^u \ll P$. The measure P^u, is therefore, determined by the Radon–Nikodym derivatives

$$\alpha_s^u = \alpha^u(s) = \frac{d\Lambda^u}{d\Lambda}(s) \quad \text{and} \quad \beta^u(s,\zeta) = \frac{d\lambda^u}{d\lambda}(s,\zeta).$$

Definition 20.2.1. *Suppose U is a compact set in a metric space (with its Borel σ-algebra). The set of* admissible controls $\mathcal{U}$ *is the set of measurable functions $u : \mathbb{R}^+ \to U$.*

To see the effect of a control, we suppose we are given functions

$$\alpha : \mathbb{R}^+ \times U \to \mathbb{R}^+,$$
$$\beta : \mathbb{R}^+ \times \mathcal{Z} \times U \to \mathbb{R}^+$$

satisfying the following conditions, for all $(s, \zeta, u) \in \mathbb{R}^+ \times \mathcal{Z} \times U$,

(i) $\alpha(s, \cdot)$ and $\beta(s, \zeta, \cdot)$ are continuous for all (s, ζ),
(ii) $\alpha(\cdot, u)$ and $\beta(\cdot, \cdot, u)$ are measurable for all $u \in U$,
(iii) $0 < c_1 \leq \alpha(s, u) \leq c(s)$,
(iv) $0 < c_2 \leq \beta(s, \zeta, u) \leq c_3$,
(v) $\int_{\mathcal{Z}} \beta(s, \zeta, u)\lambda(s, dx) = 1$.

Here c_1, c_2, c_3 are constants and $c(\cdot)$ is a measurable function such that $c(s) < \min\big(c_4,\, 1/\Delta\Lambda(s)\big)$ for each s and some finite constant $c_4 > 0$.

For $u \in \mathcal{U}$ write $\alpha^u = \alpha\big(s, u(s)\big)$ and $\beta^u = \beta\big(s, \zeta, u(s)\big)$. Note that u then controls the probability $d\Lambda^u$, of when the jump occurs, and $\lambda^u(d\zeta, s)$, of where it goes. If $L^u(t, z)$ is defined by Theorem 20.1.9, a measure corresponding to $u \in \mathcal{U}$ is given by $dP^u = L^u(T, z)dP$.

Remark 20.2.2. The assumption that U is a compact set can be weakened, see Remark 20.3.14.

Remark 20.2.3. Under the above conditions, P^u and P are mutually absolutely continuous, that is, they have the same null sets, so statements made 'almost surely' are unambiguous.

Remark 20.2.4. The predictable σ-algebra on $\mathbb{R}^+ \times \Omega$ is the σ-algebra generated by the (real) left-continuous functions (Theorem 7.2.4). In the present situation, the fundamental process z stops after the single jump time $T(\omega)$, so, because our controls are just deterministic functions used up to time $T(\omega)$, the space $\mathcal{U}$ could equally be defined as including predictable functions. In the single jump control problem there is no element of "feedback".

Suppose a cost is associated with the jump process and has the following form:

$$G(T,z) + \int_{[0,T]\times\mathcal{Z}} c\big(s,\zeta,u(s)\big)d\mu_p^u(s),$$

where $d\mu_p^u = \lambda^u(d\zeta,s)d\Lambda_s^u$ is the compensator of the jump process under the controlled measure. (By including a Radon–Nikodym derivative in the cost, the integral could equally be taken with the compensator under the base measure.) Here G and c are real valued, measurable and bounded. If the control u is used, the expected total cost is

$$J(u) = E_u\Big[G(T,z) + \int_{[0,T]\times\mathcal{Z}} c\big(s,\zeta,u(s)\big)\alpha^u(s)\beta^u(s,\zeta)\lambda(d\zeta,s)d\Lambda_s\Big],$$

where E_u denotes the expectation with respect to the measure P^u determined as above by the Radon-Nikodym derivatives $\big(\alpha^u(s),\beta^u(s,\zeta)\big)$. However, as μ_p^u is the compensator of μ and c is bounded, this apparently more general cost can be written more simply as a terminal cost

$$J(u) = E_u\Big[G(T,z) + \int_{[0,T]\times\mathcal{Z}} c\big(s,\zeta,u(s)\big)d\mu\Big] = E_u\big[f\big(T,z,u(T)\big)\big], \quad (20.5)$$

where

$$f(s,\zeta,u) = G(s,\zeta) + c(s,\zeta,u).$$

We suppose, therefore, that the cost is of the form (20.5), where f is real, measurable, and bounded. $J(u)$ is consequently finite for all $u \in \mathcal{U}$. We also assume that c, and hence f, is continuous in u.

The optimal control problem is to determine how $u \in \mathcal{U}$ should be chosen so that $J(u)$ is minimized.

Lemma 20.2.5. *Suppose control $u \in \mathcal{U}$ is used up to time t and control $v \in \mathcal{U}$ is used from time t onwards. That is, consider a control*

$$w(s) = I_{\{s \le t\}} u + I_{\{s > t\}} v \in \mathcal{U}.$$

The resulting expected final cost, given information $\mathcal{F}_t$, we denote

$$\psi(u, v, t) = E_w \big[f\big(T, z, w(T)\big) | \mathcal{F}_t \big].$$

If

$$\psi(v, t) := I_{\{t < T\}} E_w \big[f\big(T, z, v(T)\big) | T > t \big],$$

then $\psi(v, t)$ is independent of u and

$$\psi(u, v, t) = f\big(T, z, u(T)\big) I_{\{t \ge T\}} + \psi(v, t).$$

Proof. By the definition of conditional expectation, we can write

$$\begin{aligned} \psi(u, v, t) = E_w \big[f\big(T, z, w(T)\big) \big| T \le T \big] I_{\{T \le t\}} \\ + E_w \big[f\big(T, z, w(T)\big) \big| T > t \big] I_{\{T > t\}}, \end{aligned}$$

so the decomposition is immediate from the form of $w(s)$.

To show $\psi(v, t)$ is independent of u, first write

$$\frac{dP^w}{dP} = L^w(T, z) = L$$

and

$$L_t = E[L | \mathcal{F}_t].$$

We then define $L^t = L / L_t$ (with the convention that $0/0 := 1$), so

$$\begin{aligned} L^t = I_{\{t \ge T\}} + I_{\{t < T\}} \bigg(\alpha^v(T) \beta^v(T, z) \exp \Big(- \int_{]t, T]} \big(\alpha^v(s) - 1 \big) d\Lambda^c_s \Big) \\ \times \prod_{t < s < T} \frac{\big(1 - \alpha^v(s) \Delta\Lambda(s)\big)}{\big(1 - \Delta\Lambda(s)\big)} \bigg). \end{aligned}$$

Clearly, L^t does not depend on $u \in \mathcal{U}$ and

$$E[L^t | \mathcal{F}_t] = E[L | \mathcal{F}_t] / L_t = 1 \quad \text{a.s.}$$

By Bayes' rule (Exercise 5.7.1),

$$\begin{aligned} \psi(v, t) &= \frac{E\big[I_{\{t < T\}} L_t L^t f\big(T, z, v(T)\big) | \mathcal{F}_t \big]}{E[L_t L^t | \mathcal{F}_t]} \\ &= E\big[I_{\{t < T\}} L^t(T, z) f\big(T, z, v(T)\big) | \mathcal{F}_t \big]. \end{aligned}$$

The expectations here are with respect to the base measure P, so the last term is independent of u. $\square$

20.3 Three Optimality Principles

We have seen how to determine the expected cost for a *given* control u. We now investigate *optimal* controls, which are those which minimize the expected cost.

Definition 20.3.1. *The* value function *for the control problem is defined by*

$$V_t = I_{\{T>t\}} \inf_{u\in\mathcal{U}} \Big\{ E_u\big[f\big(T, z, u(T)\big)\big|T > t\big]\Big\}.$$

Because $T > 0$ *a.s., we have*

$$V_0 = J^* = \inf_{u\in\mathcal{U}} J(u),$$

and certainly $V_T = 0$*. We say a policy* u^* *is optimal if it attains the minimum* $J(u^*) = J^*$*.*

Remark 20.3.2. For each t, the infimum in the definition of V is just the infimum of a set of real numbers, and so is well defined. However, if $\mathcal{U}$ is uncountable, this definition does not guarantee the measurability of V with respect to time. The following lemma removes this concern.

Lemma 20.3.3. *The function V is càdlàg.*

Proof. We know that, for any $u \in \mathcal{U}$,

$$E_u\big[f\big(T, z, u(T)\big)\big|\mathcal{F}_t\big] = I_{\{t\geq T\}} f\big(T, z, u(T)\big) + I_{\{t<T\}} E_u\big[f\big(T, z, u(T)\big)\big|T > t\big].$$

By the martingale representation theorem (Theorem 13.1.15), if $\tilde{\mu}^u$ is the P^u-martingale random measure obtained by compensating μ, then there exists a function $g^u : [0,\infty] \times \mathcal{Z} \to \mathbb{R}$ such that

$$\begin{aligned}
&E_u\big[f\big(T, z, u(T)\big)\big|\mathcal{F}_t\big] \\
&= E_u\big[f\big(T, z, u(T)\big)\big] + \int_{]0,t]\times\mathcal{Z}} g^u(s,\zeta)d\tilde{\mu}^u \\
&= E_u\big[f\big(T, z, u(T)\big)\big] + \int_{]0,t]\times\mathcal{Z}} g^u(s,\zeta)d\mu \\
&\qquad - \int_{]0,t]\times\mathcal{Z}} g^u(s,\zeta)\alpha^u(s)\beta^u(s,\zeta)\lambda(d\zeta,s)d\Lambda_s.
\end{aligned}$$

As f is bounded and α^u and β^u satisfy the bounds following Definition 20.2.1, we know g^u is bounded, uniformly in u. Hence, we have the equation, for $r < t$,

$$\begin{aligned}
&I_{\{t<T\}} E_u\big[f\big(T, z, u(T)\big)\big|T > t\big] - I_{\{r<T\}} E_u\big[f\big(T, z, u(T)\big)\big|T > r\big] \\
&= I_{\{t\geq T>r\}} f\big(T, z, u(T)\big) + \int_{]r,t]\times\mathcal{Z}} g^u(s,\zeta)d\mu \\
&\quad - \int_{]r,t]\times\mathcal{Z}} g^u(s,\zeta)\alpha^u(s)\beta^u(s,\zeta)\lambda(d\zeta,s)d\Lambda_s.
\end{aligned}$$

From the bounds following Definition 20.2.1, this shows that the process

$$t \mapsto I_{\{t<T\}} E_u\big[f\big(T, z, u(T)\big)\big|T > t\big]$$

is a.s. absolutely continuous with respect to the (càdlàg) Stieltjes measure $d\mu + d\Lambda$ on $[0, \infty]$, uniformly in u (that is, the Radon–Nikodym derivatives are uniformly integrable with respect to u).

From Theorem 1.3.40, we know that there exists a sequence $u^1, u^2, \ldots$, such that

$$I_{\{T>t\}} E_{u^n}\big[f\big(T, z, u^n(T)\big)\big|T > t\big] \quad \downarrow \quad V_t, \quad (d\mu + d\Lambda + dt)\text{-a.e.}$$

Using a uniform absolute continuity estimate, we conclude that V is càdlàg. It also follows that the convergence of this sequence is for all t. □

The function V describes the 'remaining cost', conditional on the information available at time t. It is also convenient to consider a process M^u describing our knowledge of the total cost, given the previously used control. Using this process, we obtain the first of our optimality principles.

Theorem 20.3.4 (Martingale Optimality Principle). *Let*

$$M_t^u = f\big(T, z, u(T)\big) I_{\{t\geq T\}} + V_t.$$

Then the following hold:

(i) $\{M_t^u\}_{t\geq 0}$ *is a* $(\Omega, \{\mathcal{F}_t\}_{t\geq 0}, P^u)$*-submartingale for any* $u \in \mathcal{U}$.
(ii) $u^* \in \mathcal{U}$ *is optimal if and only if* M^{u^*} *is a* $(\Omega, \{\mathcal{F}_t\}_{t\geq 0}, P^{u^*})$*-martingale.*

In particular, $M_0^{u^*} = V_0 = J^*$, *and*

$$M_t^{u^*} = \sup_{u\in\mathcal{U}} E_u[M_\tau^{u^*}|\mathcal{F}_t]$$

for τ *any* $\{\mathcal{F}_t\}_{t\geq 0}$*-stopping time with* $\tau \geq t$.

Proof. Using the notation of Lemma 20.2.5, for all $u \in \mathcal{U}$, for any $s \geq t$, we have

$$\begin{aligned}
V_t &= I_{\{T>t\}} \inf_w E_w\big[f\big(T, z, w(T)\big)\big|T > t\big] \\
&= I_{\{T>t\}} \inf_{u,v} E_w\big[\psi(u, v, s)\big|T > t\big] \\
&\leq I_{\{T>t\}} E_u\big[I_{\{T\leq s\}} f\big(T, z, u(T)\big)\big|T > t\big] + I_{\{T>t\}} \inf_v E_w\big[I_{\{T>s\}}\psi(v, s)\big|T > t\big] \\
&\leq I_{\{T>t\}} E_u\big[I_{\{T\leq s\}} f\big(T, z, u(T)\big)\big|T > t\big] \\
&\quad + I_{\{T>t\}} E_u\big[I_{\{T>s\}} \inf_v E_v[\psi(v, s)|T > s]\big|T > t\big] \\
&= I_{\{T>t\}} E_u\big[I_{\{T\leq s\}} f\big(T, z, u(T)\big)\big|T > t\big] + I_{\{T>t\}} E_u[V_s|T > t].
\end{aligned}$$

Therefore,

$$\begin{aligned}
M_t^u &= f\big(T, z, u(T)\big)I_{t\geq T} + V_t \\
&\leq f\big(T, z, u(T)\big)I_{t\geq T} + I_{\{T>t\}}E_u\big[f\big(T, z, u(T)\big)I_{\{T\leq s\}}\big|T>t\big] \\
&\qquad + I_{\{T>t\}}E_u[V_s|T>t]. \\
&\leq E_u\big[f\big(T, z, u(T)\big)I_{\{T\leq s\}} + V_s\big|\mathcal{F}_t\big] \\
&= E_u[M_s^u|\mathcal{F}_t].
\end{aligned}$$

As f is bounded, V and M^u are bounded, so M^u is a P^u-submartingale.

From the optional stopping theorem (Theorem 5.3.1) and the fact M is stopped at T, for any stopping time $\tau \geq 0$, we obtain the inequality

$$J^* = V_0 = M_0^u \leq E_u[M_\tau^u] \leq E_u[M_T^u] = E_u\big[f\big(T, z, u(T)\big)\big].$$

However, if (and only if) u^* is optimal, we have $E_{u^*}\big[f\big(T, z, u^*(T)\big)\big] = J^*$, so we see that M^{u^*} is a P^{u^*}-martingale (Theorem 5.4.6). □

Given the simplicity of our setting, it is easy to obtain the following version of Bellman's dynamic programming principle.

Theorem 20.3.5 (Dynamic Programming Principle). *We say a control $u^* \in \mathcal{U}$ is optimal at time t if*

$$E_{u^*}\big[f\big(T, z, u^*(T)\big)\big|T>t\big] = \inf_{u\in\mathcal{U}}\big\{E_u\big[f\big(T, z, u(T)\big)\big|T>t\big]\big\}.$$

The following statements hold.

(i) For $s<t$, if u^* is an optimal control at s, then u^* is also an optimal control at t.

(ii) For $s<t$, if u^* is an optimal control at s and v^* is an optimal control at t, then $w^* = I_{\{r<t\}}u^* + I_{\{r\geq t\}}v^*$ is an optimal control at s.

Proof. Statement (i) follows from the fact $w^* \in \mathcal{U}$, so M^{w^*} is a P^{w^*}-supermartingale and, by optimality of v^*, we know $M_t^{u^*} \leq M_t^{v^*}$. Monotonicity of the expectation then implies that

$$M_s^{u^*} = E_{u^*}[M_t^{u^*}|\mathcal{F}_s] \leq E_{u^*}[M_t^{v^*}|\mathcal{F}_s] = E_{w^*}[M_t^{w^*}|\mathcal{F}_s] \leq M_s^{w^*} \leq M_s^{u^*},$$

and we see that these are all equalities, so w^* is optimal at s. We also observe that $E_{u^*}[M_t^{u^*}|\mathcal{F}_s] = E_{u^*}[M_t^{v^*}|\mathcal{F}_s]$, so $M_t^{u^*} = M_t^{v^*}$, and hence u^* is also optimal at t, as stated in (ii). □

Remark 20.3.6. In this context, we also know that the value function satisfies the following 'Markovian' properties:

(i) The value function V is equal to a deterministic function of time and of the state variable $I_{\{t>T\}}$.

(ii) If an optimal control u^* exists, then it is also equal to a deterministic function of $(t, I_{\{t>T\}})$.

We shall see shortly how to derive a differential equation for V.

Using the uniqueness of the Doob–Meyer decomposition and the above principle of optimality, we now characterize an optimal control $u^* \in \mathcal{U}$ in terms of the minimization of a certain function. Functions, processes and measures associated with u^* will be denoted by f^*, μ_p^*, P^* etc. Recall that $\tilde{\mu}^* = \mu - \mu_p^*$ denotes the compensated jump measure under the measure P^*.

Theorem 20.3.7 (Minimum Principle). *A control $u^* \in \mathcal{U}$ is an optimal control if and only if there is a measurable function $g : \Omega \to \mathbb{R}$ such that*

$$M_t^* = J^* + \int_{]0,t]\times\mathcal{Z}} g(s,\zeta)d\tilde{\mu}^*, \tag{20.6}$$

where the integral is a martingale under P^* and, at almost every point (t,ζ), the control $u^*(\omega)$ minimizes the Hamiltonian

$$H(t,u) := \alpha^u(t)\Big(\int_{\mathcal{Z}} \big((g + f^u - f^*)\beta^u\big)(t,\zeta)\lambda(d\zeta,t)\Big). \tag{20.7}$$

Proof. Suppose $u^* \in \mathcal{U}$ is optimal. Then, from the martingale representation result (Theorem 13.1.15), we know that (20.6) is satisfied by the function g, where

$$\begin{aligned} g(s,\zeta) &= f^*(s,\zeta) - E^*[f^*(T,z)] \\ &\quad + I_{\{s<c\}}\frac{1}{F_s^*}\int_{]0,s]\times\mathcal{Z}} (f^*(s',\zeta') - E^*[f^*(T,Z)])dP^*(s',\zeta') \\ &= f^*(s,\zeta) - (F_s^*)^{-1}\int_{]s,\infty]\times\mathcal{Z}} f^*(s',\zeta')dP^*(s',\zeta'). \end{aligned} \tag{20.8}$$

For any other control $u \in \mathcal{U}$,

$$\begin{aligned} M_t^u &= M_t^* + (f^u - f^*)I_{\{t\geq T\}} \\ &= J^* + \int_{]0,t]\times\mathcal{Z}} g\,(d\mu - d\mu_p^u + d\mu_p^u - d\mu_p^*) \\ &\quad + \int_{]0,t]\times\mathcal{Z}} (f^u - f^*)d\tilde{\mu}^u + \int_{]0,t]\times\mathcal{Z}} (f^u - f^*)d\mu_p^u \\ &= J^* + \int_{]0,t]\times\mathcal{Z}} (g + f^u - f^*)d\tilde{\mu}^u + \int_{]0,t]\times\mathcal{Z}} (g + f^u - f^*)d\mu_p^u \\ &\quad - \int_{]0,t]\times\mathcal{Z}} (g + f^* - f^*)d\mu_p^*. \end{aligned} \tag{20.9}$$

This expresses M^u as the sum of a P^u-local martingale and a predictable process of integrable variation. Consequently M^u is a P^u-special semimartingale and, from Theorem 11.6.10, this decomposition is unique.

However, from Theorem 20.3.4, M^u is a P^u-submartingale, so

$$\int_{]0,t]\times\mathcal{Z}}(g+f^u-f^*)d\mu_p^u-\int_{]0,t]\times\mathcal{Z}}(g+f^*-f^*)d\mu_p^*$$

must be nondecreasing in t. Because

$$\mu_p^u(t,A)=\int_{]0,t\wedge T]\times A}\alpha^u(s)\beta^u(s,x)\lambda(dx,s)d\Lambda_s$$

and

$$\mu_p^*(t,A)=\int_{]0,t\wedge T]\times A}\alpha^*(s)\beta^*(s,x)\lambda(dx,s)d\Lambda_s,$$

we conclude $H(t,u)\geq H(t,u^*)$, and so obtain the minimum principle (20.7).

Conversely, suppose $g:\Omega\to\mathbb{R}$ is a measurable process such that (20.6) is satisfied and that $u^*\in\mathcal{U}$ minimizes the Hamiltonian (20.7). We know that the integral in (20.6) is a martingale under P^* and

$$J^*=E_{u^*}[f^*(T,z)]=J(u^*).$$

For any other $u\in\mathcal{U}$, write

$$A_t^u=\int_{]0,t]\times\mathcal{Z}}(g+f^u-f^*)d\mu_p^u-\int_{]0,t]\times\mathcal{Z}}(g+f^*-f^*)d\mu_p^*.$$

Because u^* minimizes the Hamiltonian, A_t^u is a nondecreasing process, so evaluating $E_u[M_T^u]$ we have

$$J(u)=E_u[M_T^u]=J^*+E_u[A_T^u].$$

Therefore, $J^*\leq J(u)$ and $u^*\in\mathcal{U}$ is optimal. □

Corollary 20.3.8. *The Hamiltonian* (20.7) *can be written*

$$H(t,u)=\alpha^u(t)\Big(\int_{\mathcal{Z}}f^u\beta^u d\lambda-\eta(t)\Big),$$

where

$$\eta(t)=E^*[f^*(T,z)|T>t].$$

Proof. We have seen in (20.8) that

$$\begin{aligned}g(s,\zeta)&=f^*(s,\zeta)-(F_t^*)^{-1}\int_{]t,\infty]\times\mathcal{Z}}f^*(s',\zeta')dP^*(s',\zeta')\\&=f^*(s,\zeta)-E^*[f^*(T,z)|T>t].\end{aligned}$$

Substituting in (20.7) the result follows. □

Remark 20.3.9. Using this form of the Hamiltonian, the above minimum principle appears similar to those of Pliska [150] and Rishel [156]. In fact in [156] Rishel gives a system of "adjoint equations" which are satisfied by his analog of the function $\eta(t)$. The following theorem gives the equivalent result in the present context.

Remark 20.3.10. From this form of the Hamiltonian, and the assumptions on α, β and f, we see that $H(t,\cdot)$ is continuous for all t, and $H(\cdot, u)$ is measurable for all u.

Theorem 20.3.11. *For $t < c^*$, the function $\eta(t)$ satisfies*

$$\begin{aligned}\eta(t) - \eta(0) = \int_{]0,t]} \Big\{ \frac{\eta(s)}{1-\alpha^*(s)\Delta\Lambda(s)} - \gamma(s) \Big\} \alpha^*(s) d\Lambda(s) \\ - \sum_{s\le t} \frac{\big(\alpha^*(s)\Delta\Lambda(s)\big)^2}{1-\alpha^*(s)\Delta\Lambda(s)}\,\gamma(s),\end{aligned}$$

where

$$\gamma(s) = \int_{\mathcal{Z}} f^*(\zeta, s)\lambda^*(d\zeta, s).$$

Proof. For $t < c^*$ we have

$$\begin{aligned}\eta(t) &= \frac{1}{F_t^*}\int_{]t,\infty]\times\mathcal{Z}} f^* dP^* = -\frac{1}{F_t^*}\int_{]t,\infty]}\int_{\mathcal{Z}} f^* d\lambda^* dF_s^* \\ &= -\frac{1}{F_t^*}\int_{]t,\infty]} \gamma(s) dF_s^*.\end{aligned}$$

Applying the product formula for Stieltjes integrals (Theorem 1.3.43), this gives

$$\begin{aligned}&\eta(t) - \eta(0) \\ &= \int_{]0,t]} \frac{1}{F_{s-}^*}\gamma(s) dF_s^* - \int_{]0,t]} F_{s-}^*\,\eta(s-) d\Big(\frac{1}{F_s^*}\Big) + \sum_{s\le t} \Delta\Big(\frac{1}{F_s^*}\Big)\gamma(s)\Delta F_s^*.\end{aligned} \tag{20.10}$$

Now

$$\begin{aligned}d\Lambda^*(s) = \alpha^*(s)d\Lambda(s) = -\frac{dF_s^*}{F_{s-}^*}, \\ \Delta F_s^* = -F_{s-}^*\,\alpha^*(s)\Delta\Lambda(s),\end{aligned}$$

and

$$F_s^* = F_{s-}^* + \Delta F_s^* = \big(1 - \alpha^*(s)\Delta\Lambda(s)\big)F_{s-}^*.$$

Similarly,

$$\Delta\Big(\frac{1}{F_s^*}\Big) = \frac{1}{F_s^*}\,\alpha^*(s)\Delta\Lambda(s).$$

Substituting this in (20.10), we obtain the desired equation. □

Corollary 20.3.12. *Suppose $\Lambda(t)$ is continuous. Then $\eta(t)$ satisfies the differential equation*

$$\frac{d\eta}{d\Lambda}(s) = \alpha^*(s)\big(\eta(s) - \gamma(s)\big) = -H\big(s, u^*(s)\big).$$

Proof. When Λ is continuous, $\Delta\Lambda \equiv 0$ and the sum on the right of (20.10) disappears. Therefore, $\eta(s-) = \eta(s)$ and, from Corollary 20.3.8, the integrand of the remaining term is just $-H\big(s, u^*(s)\big)$. □

Remark 20.3.13. For continuous F (and so continuous Λ), the minimum principle can be written in the compact form

$$\begin{aligned} H(t, u, \eta) &= \alpha^u(t)\Big[\int_{\mathcal{Z}} f^u \beta^u d\lambda - \eta\Big], \\ \frac{d\eta}{d\Lambda}(t) &= -\inf_{u\in U} H\big(t, u, \eta(t)\big), \end{aligned}$$

with initial condition $\eta(0) = J^*$. An optimal control u^* is then obtained as a function of $(t, \eta(t))$, given implicitly as the minimizer of $H\big(t, u(t), \eta(t)\big)$. More usefully in practice, if we are in the case where $c < \infty$ and $F_{c-} > 0$, then we have the convenient representation

$$\frac{d\eta}{d\Lambda}(t) = -\inf_{u\in U} H\big(t, u, \eta(t)\big) \quad \text{and} \quad \eta(c) = \inf_{u\in U} \int_{\mathcal{Z}} f(c, z, u)\beta^u d\lambda.$$

which is an ordinary differential equation, and can be solved backwards from time c, without knowing the value of J^* *a priori*. From the definition of η, our value function can be written

$$V_t = I_{\{t<T\}}\, \eta_t,$$

so this provides a convenient means of calculating V_t. This is essentially the appropriate form of the Hamilton–Jacobi–Bellman equation (cf. Definition 21.4.9) in this context.

Remark 20.3.14. The existence of a *pointwise* minimizer of $H\big(t, u, \eta(t)\big)$ is guaranteed by the continuity of H with respect to u, and the fact that U is compact. This raises the question of whether the minimizer of $H\big(t, u, \eta(t)\big)$ can be chosen so as to be measurable with respect to time, as otherwise it would not be in $\mathcal{U}$. It turns out that there *does* exist a dt-a.e. optimal measurable control, as we have assumed H is continuous in u, measurable in t, and U is compact. The proof of (a more general version of) this result is the purpose of Appendix A.10.

The results of Appendix A.10 will also allow us to relax the compactness of U, with the effect that we can only guarantee the existence of controls which are arbitrarily close to optimal. These details will be considered more closely in the setting of the coming chapter.

20.4 Exercises

Exercise 20.4.1. Consider a duel between two players, whose accuracy improves over time. Each player has a single shot, which they decide to fire at a randomly chosen time. Suppose the probability of a hit from a shot fired by Player i at t is given by $\pi_i(t)$ for some function $\pi : [0,1] \to [0,1]$. If one player shoots before time 1 and misses, then the other player wins. On the other hand, there may be an advantage to shooting before the other player.

Let T be the time of the first shot fired, and $\mathcal{Z} = \{1, 2\}$, so that z indicates which player loses the game. Suppose Player 2 shoots in the interval $]t, t+dt]$ with rate $\frac{r_t}{1-t}dt$, where $r_t \in]0, \infty[$. Player 1 chooses to shoot with rate $\frac{u_t}{1-t}dt$, where u_t is the control to be chosen. We define the 'base measure' P to be that given by $u_t = r_t = 1/2$, so T has a uniform distribution under P, and

$$P(z = 1|T = t) = \frac{1 - \pi_1(t) + \pi_2(t)}{2} =: h(t).$$

We then have

$$\alpha^u(t) = r_t + u_t \qquad \beta(u, t, 1) = 1 - \beta(u, t, 2) = \frac{(1 - \pi_1(t))u_t + \pi_2(t)r_t}{u_t + r_t} \Big/ h(t).$$

Suppose each player associates a cost 1 with losing and no cost with winning.

(i) For a given r, find a differential equation implicitly describing the optimal policy u for Player 1, when $U = [1/k, k]$ for some $k > 0$.

(ii) By symmetry, in the case $\pi_1(t) = \pi_2(t) = t^c$ for some $c > 0$, find a Nash equilibrium for the game (that is, policies u and r such that u is optimal for Player 1 given r, and r is optimal for Player 2 given u), when both players must select policies from the set $U = [1/k, k]$. Give an intuitive explanation why this strategy is reasonable.

21

Optimal Control of Drifts and Jump Rates

We now discuss the optimal control of the solution to a stochastic differential equation, of a type similar to those considered in Chapter 17.

We explore two formulations of our control problem. In the first, we consider a control which affects the probability measure directly. In the second, we suppose that we have a reference process X, and we can add a drift to the dynamics of X, in a weak sense. Using Girsanov's theorem, this corresponds to a change of measure, and so fits within the first setting. We shall see that, provided our control acts only on the drift of our process and the compensator of the jump measure, then it is possible to express the value function for our control problem through a BSDE.

The value function for the optimal control can then be found by minimizing the driver of the BSDE, which plays the role of the Hamiltonian for our control problem. Martingale and minimum principles appear naturally from this formulation. In the Markovian case, this also allows us to formulate a PDE for the optimal value function, which is the famous Hamilton–Jacobi–Bellman equation.

Earlier work is reviewed in Fleming [80]. Other techniques in this area (including those in the first edition of this book) depended more on ad-hoc arguments, rather than applications of the theory of BSDEs. The broad approach taken here is due to Quenez [154], however it is worth noting that the connection between BSDEs and optimal control has been fundamental since the early work of Bismut [17].

This framework does not cover all problems considered in optimal stochastic control, in particular, the controls considered here cannot affect the diffusion coefficient σ. However, a full consideration of the general theory requires careful analysis, and can be found in the books of Krylov [120], Yong and Zhou [188], Fleming and Soner [79], Touzi [177] and Pham [149]. For problems with jumps, see also Øksendal and Sulèm [143].

S.N. Cohen, R.J. Elliott, *Stochastic Calculus and Applications*,
Probability and Its Applications, DOI 10.1007/978-1-4939-2867-5_21

21.1 Continuous Time Control

Basic setting: For a given filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$, suppose that we have an N-dimensional Brownian motion $W = \{W^1, W^2, \ldots, W^N\}$ (for $N \leq \infty$) and a compensated random measure $\tilde{\mu} = \mu - \mu_p$, with compensator $\mu_p(d\zeta, dt) = \nu(d\zeta)dt$. As in Chapter 19, we assume that $(W, \tilde{\mu})$ together have the predictable representation property in the filtration $\{\mathcal{F}_t\}_{t\geq 0}$. (However, we do not require $\{\mathcal{F}_t\}_{t\geq 0}$, to be the filtration generated by $(W, \tilde{\mu})$.)

Controls: Let U be a space of controls, which we assume is equal to a countable union of compact metrizable subsets of itself, for example $\mathbb{R}$ or $\mathbb{N}$. (This technical assumption is only required to enable us to prove the measurability of optimal controls). We write $\mathcal{U}$ for the space of $\{\mathcal{F}_t\}_{t\geq 0}$-predictable processes taking values in U.

To model the effect of a control, we suppose we have functions

$$\begin{aligned}
&\alpha : \Omega \times [0, T] \times U \to \mathbb{R}^{1\times N},\\
&\beta : \mathcal{Z} \times \Omega \times [0, T] \times U \to]0, \infty[
\end{aligned}$$

A controller will act to modify the measure P under which our system evolves, replacing it with the measure P^u defined by

$$\frac{dP^u}{dP} = \mathcal{E}\Big(\int_{[0,\cdot]} \alpha(\omega, t, u_t) dW_t + \int_{\mathcal{Z}\times[0,\cdot]} \big(\beta(\zeta, \omega, t, u_t) - 1\big)\tilde{\mu}(d\zeta, dt)\Big)_T,$$

where $\mathcal{E}$ is the Doléans-Dade exponential (Definition 15.1.1) and T is a deterministic terminal time. We shall write E for the expectation under P and E_u for the expectation under P^u. For ease of notation, we define

$$\Lambda_t^u := \frac{dP^u}{dP}\Big|_{\mathcal{F}_t} = \mathcal{E}\Big(\int_{[0,\cdot]} \alpha(\omega, t, u_t) dW_t + \int_{\mathcal{Z}\times[0,\cdot]} \big(\beta(\zeta, \omega, t, u_t) - 1\big)\tilde{\mu}(d\zeta, dt)\Big)_t.$$

For simplicity, we shall begin by assuming that the functions α and β are uniformly bounded, which guarantees this defines a true probability measure P^u equivalent to P, that is, Λ^u is a strictly positive martingale. In fact, under this assumption, Λ^u is a positive square integrable martingale, as can be seen using Lemma 19.2.1. (See Remark 21.3.11 for generalizations of this assumption.)

By applying Girsanov's theorem (Theorem 15.2.6, see also Corollaries 15.3.4 and 15.3.7), we see that

$$W^u := W - \int_{[0,\cdot]} \alpha(\omega, t, u_t) dt$$

is a P^u Brownian motion, while the compensator of μ under P^u is given by

$$\mu_p^u(d\zeta, dt) := \beta(\zeta, \omega, t, u_t)\nu(d\zeta)dt.$$

Therefore, we see that our controller is effectively modifying the drift of the Brownian motion, and the rates of jumps of different sizes.

Remark 21.1.1. It is important to note that the Brownian motion W^u does depend on u, and that the filtration we are working under is the original filtration $\{\mathcal{F}_t\}_{t\geq 0}$, rather than the (potentially smaller) filtration generated by $(W^u, \tilde{\mu}^u)$.

In general, our dynamics do not need to be Markovian. In particular, the dynamics, and the controls considered, are permitted to depend on the path of the *uncontrolled* processes W and μ. In some sense, this is natural, as, given the path of W^u *and* knowledge of the past control u, a controller can extract the path of W (and the path of μ is always observable). On the other hand, Tsirel'son's example (Example 18.1.7) shows that there is no guarantee that this can be done using only the path of W^u (without knowledge of u).

Costs: Suppose the controller faces a cost which can be decomposed into two terms:

- A running cost, determined by a function
$$c : \Omega \times [0,T] \times U \to \mathbb{R}$$
which is predictable in (ω, t) and continuous in u. We assume that c is bounded below, and that
$$E\Big[\int_{[0,T]} c(\omega,t,u_t)^2 dt\Big] < \infty$$
for all $u \in \mathcal{U}$.
- A terminal cost, determined by an $\mathcal{F}_T$-measurable random variable $\xi \in L^2(P)$.

Then, if control $u \in \mathcal{U}$ is used, the total expected cost is given by
$$J(u) = E_u\Big[\xi + \int_{[0,T]} c(\omega,t,u_t)dt\Big], \tag{21.1}$$
where E_u denotes expectation with respect to P^u. (We shall see, in the proof of Lemma 21.2.5, that the term inside the expectation is in $L^1(P^u)$, so this is well defined.) We wish to select a control u such that the total expected cost is minimized.

Remark 21.1.2. Conceptually, it is somewhat easier to think about the case where (W,μ) generates the filtration. Then, by the Doob–Dynkin lemma, the terminal cost ξ is a function of $\{W_t, \mu(\cdot,t)\}_{t\leq T}$ (the paths of W and μ for times in $[0,T]$). What a controller attempts to do is to modify the probabilities of different paths, in a dynamic way, to try and minimize the expected terminal cost $E[\xi]$. However, using a control incurs a cost c, so the controller then needs to balance the benefits from increasing the probability of less costly outcomes (low values of ξ) against the cost of controlling more actively.

21.2 The Martingale Principle

For the general problem described above, we now seek to obtain a version of the martingale principle of optimality. As in the case of a single jump, we define the value function to be the minimal cost which can be realized starting from time t. Because many of our quantities are random variables, and only defined P-a.s., this necessitates the use of the essential infimum (see Theorem 1.3.40). We recall that the essential infimum is constructed to lie in a given family of measurable functions, and is minimal up to equality almost everywhere (for a given measure).

Definition 21.2.1. *For a given control $u \in \mathcal{U}$, we define the expected remaining cost*

$$J(\omega, t, u) = E_u\Big[\xi + \int_{]t,T]} c(\omega, s, u_s)ds\Big|\mathcal{F}_t\Big].$$

The value process *is defined by*

$$V_t := \operatorname{ess\,inf}_{u \in \mathcal{U}} J(\omega, t, u) = \operatorname{ess\,inf}_{u \in U} E_u\Big[\xi + \int_{]t,T]} c(\omega, s, u_s)ds\Big|\mathcal{F}_t\Big],$$

The essential infimum is taken in the $\mathcal{F}_t$-measurable random variables, and is defined P-a.e. The cost process *is defined by*

$$M_t^u = \int_{[0,t]} c(\omega, s, u_s)ds + V_t.$$

Remark 21.2.2. As we have not assumed that $\mathcal{F}_0$ is trivial, we cannot assume that $V_0 = \operatorname{ess\,inf}_{u \in \mathcal{U}} J(\omega, 0, u)$ is deterministic. Nevertheless, this will often be the case in applications. In either case, as our control does not affect the measure on $\mathcal{F}_0$ (that is, $P^u|_{\mathcal{F}_0} = P|_{\mathcal{F}_0}$ for all u), we can see that

$$E[V_0] = \inf_{u \in \mathcal{U}} J(u).$$

Remark 21.2.3. We shall see below (Theorem 21.3.6) that the value process (and hence the cost process) has a càdlàg version, which solves a certain BSDE. Given this, we do not need to be concerned here about measurability of V with respect to time.

Lemma 21.2.4. *Suppose a control $u \in \mathcal{U}$ is built from two controls v, $w \in \mathcal{U}$ by concatenation, that is, for some $t \in\,]0, T[$,*

$$u(\omega, s) = I_{[0,t]}w + I_{]t,T]}v = \begin{cases} w(\omega, s), & 0 \le s \le t, \\ v(\omega, s), & t < s \le T. \end{cases}$$

Then $J(\omega, t, u) = J(\omega, t, v)$ a.s. and, consequently, $J(\omega, t, u)$ does not depend on the control w.

Proof. By inspecting the stochastic exponential, it is easy to verify that

$$\frac{dP^u}{dP}\Big|_{\mathcal{F}_s} = \Lambda_s^u = \begin{cases} \Lambda_t^w \big(\Lambda_s^v / \Lambda_t^v\big) & \text{if } s > t, \\ \Lambda_s^w & \text{if } s \le t. \end{cases}$$

By Bayes' rule (Exercise 5.7.1),

$$\begin{aligned} J(\omega,t,u) &= \frac{1}{\Lambda_t^u} E\Big[\Lambda_T^u\Big(\xi + \int_{]t,T]} c(s,x_s,u_s)ds\Big)\Big|\mathcal{F}_t\Big] \\ &= \frac{1}{\Lambda_t^w} E\Big[\frac{\Lambda_t^w}{\Lambda_t^v}\Lambda_T^v\Big(\xi + \int_{]t,T]} c(s,x_s,v_s)ds\Big)\Big|\mathcal{F}_t\Big] \\ &= \frac{1}{\Lambda_t^v} E\Big[\Lambda_T^v\Big(\xi + \int_{]t,T]} c(s,x_s,v_s)ds\Big)\Big|\mathcal{F}_t\Big] \\ &= J(\omega,t,v). \end{aligned}$$

□

Lemma 21.2.5. *For any $s,t \in [0,T]$ and any $u \in \mathcal{U}$,*

$$\operatorname{ess\,inf}_{w\in\mathcal{U}} E_u[J(\omega,t,w)|\mathcal{F}_s] = E_u[\operatorname{ess\,inf}_{w\in\mathcal{U}} J(\omega,t,w)|\mathcal{F}_s]$$

where the first essential infimum is taken in the $\mathcal{F}_s$-measurable random variables, while the second is taken in the $\mathcal{F}_t$-measurable random variables, both P-a.e.

Proof. From the definition of the essential infimum, it is easy to see that

$$\operatorname{ess\,inf}_{w\in\mathcal{U}} E_u[J(\omega,t,w)|\mathcal{F}_s] \ge E_u[\operatorname{ess\,inf}_{w\in\mathcal{U}} J(\omega,t,w)|\mathcal{F}_s].$$

Conversely, we can find a sequence $\{w^n\}_{n\in\mathbb{N}} \subset \mathcal{U}$ such that

$$J(\omega,t,w^n) \downarrow \operatorname{ess\,inf}_{w\in\mathcal{U}} J(\omega,t,w) \qquad \text{a.s.}$$

Using Lemma 19.2.1, for any $u \in \mathcal{U}$ we can check that Λ^u is square integrable, and as $\xi + \int_{[0,T]} c(\omega,t,u_t)dt \in L^2(P)$, it follows that

$$\xi + \int_{[0,T]} c(\omega,t,u_t)dt \in L^1(P^u).$$

As c is bounded below, we see that $\operatorname{ess\,inf}_{w\in\mathcal{U}} J(\omega,t,w) \in L^1(P^u)$. Therefore, by dominated convergence,

$$\begin{aligned} \operatorname{ess\,inf}_{w\in\mathcal{U}} E_u[J(\omega,t,w)|\mathcal{F}_s] &\le \lim_n E_u[J(\omega,t,w^n)|\mathcal{F}_s] \\ &= E[\operatorname{ess\,inf}_{w\in\mathcal{U}} J(\omega,t,w)|\mathcal{F}_s]. \end{aligned}$$

□

We now obtain the martingale optimality principle.

Theorem 21.2.6 (Martingale Optimality Principle). *For each $u \in \mathcal{U}$, the cost process M^u is a P^u-submartingale. Furthermore, M^u is a P^u-martingale if and only if the control u gives the minimum expected cost, i.e. if and only if u is optimal. In particular, for an optimal control u^*, we know*

$$E[M_0^{u^*}] = E[V_0] = \inf_{u \in \mathcal{U}} J(u),$$

and

$$M_t^{u^*} = \sup_{u \in \mathcal{U}} E_u[M_\tau^{u^*} | \mathcal{F}_t]$$

for τ any $\{\mathcal{F}_s\}_{s \geq 0}$-stopping time with $\tau \geq t$.

Proof. For any $0 \leq r \leq t \leq T$, from the definition of the essential infimum, we know

$$V_r = \operatorname{ess\,inf}_{u \in \mathcal{U}} J(\omega, r, u) = \operatorname{ess\,inf}_{u \in \mathcal{U}} E_u\Big[\int_{]r,t]} c(\omega, s, u_s) ds + J(\omega, t, u)\Big| \mathcal{F}_r\Big].$$

By considering a concatenated control, as in Lemma 21.2.4, and using Lemma 21.2.5, we see

$$\begin{aligned} V_r &= \operatorname{ess\,inf}_{w \in \mathcal{U}} E_w\Big[\int_{]r,t]} c(\omega, s, w_s) ds + \operatorname{ess\,inf}_{v \in \mathcal{U}} J(\omega, t, v)\Big| \mathcal{F}_r\Big] \\ &\leq E_u\Big[\int_{]r,t]} c(\omega, s, u_s) ds + \operatorname{ess\,inf}_{v \in \mathcal{U}} J(\omega, t, v)\Big| \mathcal{F}_r\Big], \end{aligned}$$

where in the second line u is an arbitrary element of $\mathcal{U}$. Therefore, for any $u \in \mathcal{U}$,

$$V_r \leq E_u\Big[\int_{]r,t]} c(\omega, s, u_s) ds + V_t \Big| \mathcal{F}_r\Big],$$

and hence $M_r^u \leq E_u[M_t^u | \mathcal{F}_r]$ a.s., that is, M^u is a P^u-submartingale.

From the optional stopping theorem (Theorem 5.3.1), for any bounded stopping time $\tau \geq 0$, we obtain the inequality

$$V_0 = M_0^u \leq E_u[M_\tau^u | \mathcal{F}_0] \leq E_u[M_T^u | \mathcal{F}_0] = E_u\Big[\xi + \int_{]0,T]} c(\omega, s, u_s) ds \Big| \mathcal{F}_0\Big].$$

However, if (and only if) u^* is optimal, we have

$$E_{u^*}\Big[\xi + \int_{]0,T]} c(\omega, s, u_s) ds\Big] = E[V_0],$$

so we see that M^{u^*} is a P^{u^*}-martingale (Theorem 5.4.6).

□

Exactly as in the single jump case (Theorem 20.3.5), we obtain a version of Bellman's dynamic programming principle (cf. Bellman [8]).

Theorem 21.2.7 (Dynamic Programming Principle). *We say a control $u^* \in \mathcal{U}$ is optimal at time t if*

$$J(\omega, t, u^*) = \operatorname{ess\,inf}_{u\in\mathcal{U}} J(\omega, t, u).$$

The following statements then hold.

(i) For $s < t$, if u^* is an optimal control at s, then u^* is also an optimal control at t.

(ii) For $s < t$, if u^* is an optimal control at s and v^* is an optimal control at t, then $w^* = I_{[0,t]}u^* + I_{]t,T]}v^*$ is an optimal control at s.

21.3 BSDEs and the Minimum Principle

We now seek to represent our cost equation (21.1) as the solution to a certain BSDE. This allows us to access the powerful result of the comparison theorem for BSDEs, which greatly simplifies our search for an optimal control.

Lemma 21.3.1. *For a given control $u \in \mathcal{U}$, the process*

$$J(\omega, u, t) = E_u\Big[\xi + \int_{]t,T]} c(\omega, t, u_t)dt\Big|\mathcal{F}_t\Big]$$

is the unique solution to the BSDE

$$\begin{cases} dJ(\omega, u, t) = -f(\omega, t, Z_t, \Theta_t, u_t)dt + Z_t dW_t + \displaystyle\int_{\mathcal{Z}} \Theta_t(\zeta)\tilde{\mu}(d\zeta, dt), \\ J(\omega, u, T) = \xi, \end{cases}$$

where

$$f(\omega, t, z, \theta, u) := c(\omega, t, u) + z\alpha(\omega, t, u) + \int_{\mathcal{Z}} \theta(\zeta)\big(\beta(\zeta, \omega, t, u) - 1\big)\nu(d\zeta)$$

is a balanced Lipschitz driver and is linear with respect to z and θ.

Proof. From the definition of Λ and Exercise 5.7.1, we know that

$$\begin{aligned} J(u, t) &= E_u\Big[\xi + \int_{]t,T]} c(\omega, s, u_s)ds\Big|\mathcal{F}_t\Big] \\ &= \frac{1}{\Lambda_t^u} E\Big[\Lambda_T^u \xi + \int_{]t,T]} \Lambda_{s-}^u c(\omega, s, u_s)ds\Big|\mathcal{F}_t\Big], \end{aligned}$$

which is, by Theorem 19.2.2, the solution to the stated linear BSDE.

That f is Lipschitz in z and θ follows from the boundedness of α and β. That f is balanced (Definition 19.3.1) follows from the positivity of β. □

Remark 21.3.2. Simple rearrangement in the BSDE above yields

$$dJ(\omega,u,t) = -c(\omega,t,u_t)dt + Z_t dW_t^u + \int_{\mathcal{Z}} \Theta_t(\zeta)\tilde{\mu}^u(d\zeta,dt); \qquad J(\omega,u,T)=\xi$$

so we see that Z and Θ are precisely the terms coming from applying the Martingale Representation Theorem to $\left\{E_u\left[\xi + \int_{[0,T]} c(\omega,s,u_s)ds\middle|\mathcal{F}_t\right]\right\}_{t\geq 0}$ under the measure P^u.

The driver of the BSDE essentially plays the role of the Hamiltonian in our control problem. Considering the comparison theorem, it appears that the optimal control can be obtained by taking a minimizer of $f(\omega,t,z,\theta,u)$ with respect to u. In particular, we see that the optimal value function satisfies the BSDE with driver given by the minimized Hamiltonian, defined as follows.

Lemma 21.3.3. *Define the function*

$$H(\omega,t,z,\theta) = \operatorname{ess\,inf}_{u\in U} f(\omega,t,z,\theta,u).$$

Then there is a version of H which is a balanced standard Lipschitz driver for a BSDE.

Proof. This follows directly from Lemma 19.3.8. □

Before attempting to construct an optimal control, we require a general result, which allows us to choose controls in a measurable way. The proof of this is due to Beneš [10] (extending earlier work by McShane and Warfield [130] and Filippov [78]) and can be found in Appendix A.10.

Theorem 21.3.4 (Filippov's Implicit Function Theorem). *Let U be a topological space which is the union of countably many compact metrizable subsets of itself and X be a separable metric space. Let $G : \Omega \times [0,\infty[\times X \times U \to \mathbb{R}$ be such that*

(i) $G(\cdot,\cdot,x,u)$ is Σ_p-measurable (i.e. predictable), for every $u \in U$, $x \in X$,
(ii) $G(\omega,t,\cdot,u)$ is uniformly continuous, for $dP\times dt$-almost all (ω,t) and all $u \in U$.
(iii) $G(\omega,t,x,\cdot)$ is continuous, for $dP\times dt$-almost all (ω,t) and all $x\in X$,
(iv) $\operatorname{ess\,inf}_{u\in U} G(\omega,t,x,u) > -\infty$ for $dP\times dt$-almost all (ω,t) and all $x \in X$,

where in (iv), the essential infimum is taken in the predictable processes, and defined $dP\times dt$-a.e. Then, for every $\epsilon>0$, there exists a $\Sigma_p\otimes\mathcal{B}(X)$-measurable function u^ϵ taking values in U such that, for every x,

$$G(\omega,t,x,u^\epsilon(\omega,t,x)) < \operatorname{ess\,inf}_{u\in U} G(\omega,t,x,u) + \epsilon \qquad dt\times dP - a.e.$$

If we also know that

(v) for $dP\times dt$-almost all (ω,t) and all $x\in X$, there exists $v\in U$ such that $G(\omega,t,x,v) = \operatorname{ess\,inf}_{u\in U} G(\omega,t,x,u)$,

then there exists a $\Sigma_p \otimes \mathcal{B}(X)$-measurable function u^* such that

$$G(\omega, t, x, u^*(\omega, t, x)) = \operatorname{ess\,inf}_{u \in U} G(\omega, t, x, u) \qquad dt \times dP - a.e.$$

The functions $G(\omega, t, x, u^\epsilon(\omega, t, x))$ (and $G(\omega, t, x, u^*(\omega, t, x))$, when defined) have the same modulus of continuity with respect to x as G.

Remark 21.3.5. For our applications, the above result will be applied when G is Lipschitz continuous with respect to x, in which case we have that $G(\omega, t, x, u^\epsilon(\omega, t, x))$ is also Lipschitz continuous, with the same Lipschitz constant.

This theorem will be used to select the optimal control in a predictable way. We now prove the main result of this section, which yields a representation of the value function in terms of a BSDE.

Theorem 21.3.6. *The value function V has a càdlàg modification, which is the solution to the BSDE*

$$\begin{cases} dV_t = -H(\omega, t, Z_t, \Theta_t)dt + Z_t dW_t + \displaystyle\int_{\mathcal{Z}} \Theta_t(\zeta)\tilde{\mu}(d\zeta, dt), \\ V_T = \xi, \end{cases}$$

with $H = \operatorname{ess\,inf}_{u \in U} f$ defined as in Lemma 21.3.3.

Proof. By Lemma 21.3.3 and Theorem 19.1.7, the BSDE with driver H and terminal value ξ has a (càdlàg) solution, which we denote Y. Our aim is to show that Y is a version of the value function.

As H is a balanced driver, and by definition $H(\omega, t, z, \theta) \leq f(\omega, t, z, \theta, u)$ $dP \times dt$-a.e. for all $u \in U$, we can apply the comparison theorem for BSDEs (Theorem 19.3.4) and the result of Lemma 21.3.1 to deduce that, up to indistinguishability,

$$Y_t \leq J(\omega, t, u) \text{ for all } u \in \mathcal{U}.$$

However, from Theorem 21.3.4 (applied with $f = G$ and $X = \mathbb{R}^N \times L^2(\nu)$), there exists a (predictable) control $u^\epsilon \in \mathcal{U}$ such that

$$f(\omega, t, z, \theta, u^\epsilon) \leq H(\omega, t, z, \theta) + \epsilon \qquad dP \times dt\text{-a.e.}$$

As $Y_t + \epsilon(T - t)$ solves the BSDE with driver $H(\omega, t, z, \theta) + \epsilon$, another application of the comparison theorem yields that, up to indistinguishability,

$$J(\omega, t, u^\epsilon) \leq Y_t + \epsilon(T - t).$$

Combining these results, with $\epsilon \to 0$, we see that

$$Y_t = \operatorname{ess\,inf}_{u \in \mathcal{U}} J(\omega, t, u) = V_t$$

for every t, so Y is a version of V. $\square$

Remark 21.3.7. In the course of the last proof, we obtained the existence of an ϵ-optimal control, that is, a control $u^\epsilon \in \mathcal{U}$ such that

$$J(\omega, t, u^\epsilon) \leq \operatorname{ess\,inf}_{u \in \mathcal{U}} J(\omega, t, u) + \epsilon.$$

This was obtained by approximating the minimizer of H. Naturally, we now see that a control is optimal if, and only if, it minimizes the Hamiltonian.

The following theorem gives the natural version of the minimum principle for our problem.

Theorem 21.3.8 (Minimum Principle). *Let (V, Z, Θ) be the solution to the BSDE with driver H and terminal value ξ. A control $u \in \mathcal{U}$ is optimal if and only if it satisfies*

$$f(\omega, t, Z_t, \Theta_t, u_t) = H(\omega, t, Z_t, \Theta_t) \qquad dP \times dt\text{-a.e.},$$

that is, u_t minimizes $f(\omega, t, Z_t, \Theta_t, \cdot)$ pointwise almost everywhere.

Proof. We know that $J(\omega, t, u) \geq V_t$ for all $u \in \mathcal{U}$, with equality if and only if u is optimal.

Suppose that we have a control u such that $f(\omega, t, Z_t, \Theta_t, u_t) = H(\omega, t, Z_t, \Theta_t)$ $dP \times dt$-a.e. Then the triple (V_t, Z_t, Θ_t) solves the BSDE with driver $f(\cdots, u_t)$ and, by uniqueness, $J(\omega, t, u) = V_t$. It follows that u is optimal.

Conversely, suppose u is optimal. We know, from Lemma 21.3.1, that for some Z' and Θ' the triple $(J(\cdot, \cdot, u), Z', \Theta')$ solves the BSDE with driver $f(\cdots, u_t)$, and that $f(\omega, t, z, \theta, u_t) \geq H(\omega, t, z, \theta)$ $dP \times dt$-a.e. With V as in Theorem 21.3.6, the strict part of the comparison theorem states that $J(\omega, 0, u) = V_0$ if and only if $J(\omega, s, u) = V_s$ for all $s \in [0, T]$. From the (unique) canonical semimartingale decompositions of these processes, we have

$$f(\omega, t, Z'_t, \Theta'_t, u_t) = H(\omega, t, Z_t, \Theta_t) \qquad dP \times dt\text{-a.e.}$$

and

$$Z' \bullet W + \Theta' * \tilde{\mu} = Z \bullet W + \Theta * \tilde{\mu}$$

up to indistinguishability. The uniqueness of the martingale representation theorem then implies that $\|Z - Z'\|^2 = 0$ and $\|\Theta - \Theta'\|_\nu^2 = 0$, both $dP \times dt$-a.e. As f and H are continuous with respect to these norms, the result follows. $\square$

Corollary 21.3.9. *Suppose that, for $dP \times dt$-almost all (ω, t) and all $(z, \theta) \in \mathbb{R}^N \times L^2(\nu)$, there exists $v \in U$ such that*

$$f(\omega, t, z, \theta, v) = \operatorname{ess\,inf}_{u \in U} f(\omega, t, z, \theta, u).$$

Then there exists an optimal control $u \in \mathcal{U}$.

Proof. By applying Theorem 21.3.4 under assumptions (i) – (v), we obtain a map $u : \Omega \times [0,T] \times \mathbb{R}^N \times L^2(\nu) \to U$ which is predictable in (ω, t), Borel measurable in (z, θ), and which satisfies $f(\omega, t, z, \theta, u(\omega, t, z, \theta)) = \operatorname{ess\,inf}_{u \in U} f(\omega, t, z, \theta, u)\ dP \times dt$-a.e. We also see that

$$(z, \theta) \mapsto f(\omega, t, z, \theta, u(\omega, t, z, \theta))$$

has the same modulus of continuity as f, that is, it is Lipschitz continuous. Solving the BSDE with this driver, we obtain a predictable process $\{u(\omega, t, Z_t(\omega), \Theta_t(\omega))\}_{t \geq 0}$ with the desired properties. $\square$

Remark 21.3.10. As f is continuous in u, the conditions of the corollary are immediately satisfied whenever U is compact.

Remark 21.3.11. In the above analysis, the only times we have made use of the boundedness of α and β are in Lemma 21.2.5, to show that $\xi + \int_{[0,T]} c(\omega, t, u_t) \in L^1(P^u)$, in establishing the existence of solutions and the comparison theorem for the relevant BSDEs, and in ensuring the Hamiltonian is pointwise bounded below (Lemma 21.3.3).

If we instead assume that ξ and c were bounded, then we can relax the boundedness of α and β, instead assuming that there exists a bound

$$\sup_{u \in \mathcal{U}} \left\{ \|\alpha(\omega, t, u)\|^2 + \|\beta(\omega, t, u)\|_\nu^2 \right\} \leq K(\omega, t).$$

with enough integrability assumptions on K that the relevant BSDEs have solutions (cf. Theorem A.9.20). Lemma 21.2.5 continues to hold (as $\xi + \int_{[0,T]} c(\omega, t, u_t) \in L^\infty(P^u) \subset L^1(P^u)$), and the Hamiltonian is still pointwise bounded below. This gives a wide range of problems which can be considered using these techniques.

The above results, via a non-BSDE approach, are fundamentally due to Davis [44], while the BSDE methods are due to El Karoui and Quenez (see [154] and references therein). They are much stronger than any those available in deterministic control theory, because the noise helps to “smooth out” the process. The existence of an optimal control was originally established by Beneš [11], and Duncan and Varaiya [63], under the hypothesis that the set $f(\omega, t, z, U)$ is convex.

21.4 Markovian Case

As one might expect, the difficulty in applying the above analysis is in numerically finding a solution to the BSDE with driver H. In the Markovian case, we have seen that there are close connections between BSDEs and PIDEs, so we can transfer this difficulty into the problem of solving a PIDE, for which many numerical methods are available.

For our reference filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$, suppose that we have a process X taking values in $\mathbb{R}^d$, which satisfies the SDE

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t + \int_{\mathcal{Z}} g(\zeta, t, X_t)\tilde{\mu}(d\zeta, dt) \tag{21.2}$$

with $\mathcal{F}_0$-measurable initial condition $X_0 = x_0$. These are called the reference dynamics for X. We can also define $X^{(t,x)}$ as in Section 19.4. Here W and $\tilde{\mu} = \mu - \mu_p$ are as above. We assume that b, σ and g are measurable in t and Lipschitz continuous in x, so, by Lemma 17.1.1, the SDE is well defined up to any fixed deterministic time T. As before, we assume that $(W, \tilde{\mu})$ together have the predictable representation property in the filtration $\{\mathcal{F}_t\}_{t\geq 0}$.

Remark 21.4.1. We do not require $\{\mathcal{F}_t\}_{t\geq 0}$ to be the filtration generated by $(W, \tilde{\mu})$. Therefore, in principle we allow X to be a (Markovian) weak solution of the SDE (21.2), provided, for example, the conditions of Theorem 18.2.3 are satisfied. The difficulty is that we shall need some continuity estimates on X with respect to its initial conditions, which are more easily obtained in the Lipschitz setting. We could also allow our dynamics to depend on the entire path of X. However, this will not lead to the Markovian BSDEs we study below.

The effect of a control $u \in \mathcal{U}$ is as defined above, with the assumption that the effect of the control depends on ω only through $X_t(\omega)$, that is, we can write α and β as functions of the 'state variables' $(t, X_t(\omega))$ and the control u_t. Under such a control, we see that X has P^u-dynamics

$$dX_t = \hat{b}(t, X_t, u_t)dt + \sigma(t, X_t)dW_t^u + \int_{\mathcal{Z}} g(\zeta, t, X_t)\tilde{\mu}^u(d\zeta, dt) \tag{21.3}$$

for W^u a P^u-Brownian motion, where $\tilde{\mu}^u = \mu - \mu_p^u$ and

$$\begin{aligned} \hat{b}(t, X_t, u_t) &= b(t, X_t) + \sigma(t, X_t)\alpha(t, X_t, u_t) \\ &\qquad + \int_{\mathcal{Z}} g(\zeta, t, X_t)\beta(\zeta; t, X_t, u_t)\nu(d\zeta)dt, \\ \mu_p^u(d\zeta, dt) &= \beta(\zeta; t, X_t, u_t)\nu(d\zeta)dt. \end{aligned} \tag{21.4}$$

Therefore, our control can be seen as determining the drift and jump rates of X in *a weak sense*, as $(X, W^u, \tilde{\mu}^u, P^u)$ is a weak solution to the SDE (21.3)

In this setting, the cost usually depends on ω only as a function of $X_t(\omega)$, that is, we have a running cost $c(t, X_t, u_t)$ and terminal cost $\xi(X_T)$. The aim is to choose a control which minimizes

$$J(u) = E_u\Big[\xi(X_T) + \int_{[0,T]} c(t, X_t, u_t)dt\Big].$$

Remark 21.4.2. Suppose our control problem is stated in terms of the controlled drift $\hat{b}$ and the compensator of the jump measure μ_p^u. In order to convert this into a change of measure problem, one simply needs σ to admit a right inverse and μ_p^u to be absolutely continuous with respect to some deterministic $\nu(d\zeta)dt$, in which case rearrangement of (21.4) will determine the appropriate α and β in terms of $\hat{b}$. In particular, we see that

$$\begin{aligned}&\alpha(t, X_t, u_t)\\&= \sigma(t, X_t)^{-1}\Big(\hat{b}(t, X_t, u_t) - b(t, X_t) - \int_{\mathcal{Z}} g(\zeta, t, X_t)\beta(\zeta; t, X_t, u_t)\nu(d\zeta)dt\Big).\end{aligned}$$

For the continuous case ($\nu \equiv 0$), common conditions to ensure that these equations are well behaved are then that σ and its right inverse are uniformly bounded and that $\hat{b} - b$ is of linear growth in x (cf. Condition II in Remark 21.4.5 below).

If σ does not admit a right inverse, the problem is more difficult. However, Davis [47] showed how the following important class of degenerate systems can be treated.

Example 21.4.3. Suppose $X_t = (X_t^1, X_t^2) \in \mathbb{R}^{d+d'}$ is defined by

$$\begin{aligned}dX_t^1 &= b^1(t, X_t^1, X_t^2)dt,\\ dX_t^2 &= b^2(t, X_t^1, X_t^2, u_t)dt + \widetilde{\sigma}(t, X_t^1, X_t^2)dW_t\,,\end{aligned}\tag{21.5}$$

with initial condition $x_0 \in \mathbb{R}^{d+d'}$, and where $\widetilde{\sigma}$ is bounded with bounded inverse, both of which are Lipschitz in X. Suppose b^2 is Lipschitz and bounded, and b^1 is bounded and Lipschitz in X_t^1 uniformly in (t, X_t^2). Then for each trajectory X^2 there is a unique solution $X_t^1 = \phi_t(X^2)$ of the first of the above equations, and the second equation can be written

$$dX_t^2 = b^2\big(t, \phi_t(X^2), X^2, u\big)dt + \widetilde{\sigma}\big(t, \phi_t(X^2), X_t^2\big)dW_t\,.$$

This equation is now of the form (21.3) (however, with dependence on the whole path of X^2), and so has a weak solution for each $u \in \mathcal{U}$.

This is particularly useful as, if a scalar nth-order stochastic differential equation is written as a first-order system, then a degenerate family of equations like (21.5) is obtained.

Remark 21.4.4. Another advantage of working with weak solutions for our forward process is that we do not need the dependence of $\hat{b}$ on X to be smooth (for example, requiring $\hat{b}$ to be Lipschitz is the usual requirement for strong solutions, see Theorem 16.3.11). Consequently, “bang-bang” and other discontinuous controls can be discussed.

In this setting, we can see that our value function satisfies a *Markovian* BSDE. We define, as above, the Hamiltonian

$$f(t,x,z,\theta,u) := c(t,x,u) + z\alpha(t,x,u) + \int_{\mathcal{Z}} \theta(\zeta)\big(\beta(\zeta,t,x,u)-1\big)\nu(d\zeta)$$

and hence its infimum (using Lemma 19.3.8),

$$H(t,x,z,\theta) = \operatorname{ess\,inf}_{u\in U} f(t,x,z,\theta;u).$$

Remark 21.4.5. From our results on non-Lipschitz BSDEs, we see that it is natural either to assume:

(Condition I): That $\xi(X_T) \in L^2(\mathcal{F}_T)$, $c \in L^2(dt \times dP)$ and α, β are uniformly bounded processes, in which case we have a BSDE with a uniformly Lipschitz driver f.

(Condition II): ξ and c are bounded and, uniformly in u, either α is of linear growth (with respect to x) and $\nu \equiv 0$, or α and β are both of square-root growth (with respect to x), in which case we are in the setting of Theorem A.9.20.

Our above results can then be expressed in the following way.

Theorem 21.4.6. *The value process V has a càdlàg modification, which is the solution to the BSDE*

$$\begin{cases} dV_t = -H(t,X_t,Z_t,\Theta_t)dt + Z_t dW_t + \displaystyle\int_{\mathcal{Z}} \Theta_t(\zeta)\tilde{\mu}(d\zeta,dt), \\ V_T = \xi(X_T). \end{cases}$$

Therefore, the value process V is equal to a deterministic function of (t,X_t), that is, $V_t = v(t,X_t)$, for some continuous $v : [0,T]\times\mathbb{R}^d \to \mathbb{R}$. ($v$ is naturally called the value function for the control problem.)

Proof. The connection between the value function and the BSDE is given by Theorem 21.3.6. As the BSDE is Markovian in x, the connection with a deterministic function is given by Theorem 19.4.5. □

Using the structure of BSDEs, we can then see that, if an optimal control exists, then an optimal *feedback* control exists, that is, the optimal control depends only on the current values of the state variables (t,X_t).

Theorem 21.4.7. *Suppose that, for $dP\times dt$-almost all (ω,t) and all $(z,\theta)\in \mathbb{R}^N\times L^2(\nu)$, there exists $u'\in U$ such that*

$$f(t,x,z,\theta,u') = \operatorname{ess\,inf}_{u\in U} f(t,x,z,\theta,u).$$

Then there exists an feedback control, that is, a map $u^* : [0,T]\times\mathbb{R}^d \to U$, such that $u^*(t,X_t)$ is optimal among all predictable controls.

Proof. From Corollary 21.3.9 we know that a control is optimal if and only if it minimizes $f(t,x,z,\theta,u)$ pointwise. We also know that z and θ are Borel measurable functions of (t,x), as they come from the solution of a Markovian BSDE. Using Filippov's implicit function theorem (Theorem 21.3.4), we see that there is a $\mathcal{B}([0,T]\times\mathbb{R}^d)$-measurable map u^* which minimizes $f(t,x,z(t,x),\theta(t,x),u)$ for all x, and almost all t. □

Remark 21.4.8. If the optimum is not attained then, as in the proof of Theorem 21.3.6, we can construct (feedback) controls with values arbitrarily close to the optimum.

We now derive, from the BSDE, the Hamilton–Jacobi–Bellman equation for our control problem.

Definition 21.4.9 (Hamilton–Jacobi–Bellman Equation). *The HJB (Hamilton–Jacobi–Bellman) equation for our control problem is given by*

$$\begin{cases} \qquad 0 = \dfrac{\partial v}{\partial t}(t,x) + \mathcal{L}_t v(t,x) + H(t,x,v,(\partial_x v)\sigma,\tilde{v}), \\ v(T,x) = \xi(x), \end{cases}$$

where $\tilde{v}$ denotes the element of $L^2(\nu)$ given by the map

$$\zeta \mapsto v\big(s, x+g(\zeta,s,x)\big) - v(s,x)$$

and $\mathcal{L}$ is the integro-differential operator (with $a=\sigma\sigma^\top$),

$$\begin{aligned} \mathcal{L}_s v(x) &= \sum_i b^i(s,x)\frac{\partial v}{\partial x^i}(x) + \frac{1}{2}\sum_{i,j} a^{ij}(s,x)\,\frac{\partial^2 v}{\partial x^i \partial x^j}(s,x) \\ &\quad + \int_{\mathcal{Z}} \Big(v(x+g(\zeta,s,x)) - v(x) - \sum_i \frac{\partial v}{\partial x^i}(x) g^i(\zeta,s,x)\Big)\nu(d\zeta)\,. \end{aligned}$$

Theorem 21.4.10 (Verification Theorem). *Suppose one of the conditions given in Remark 21.4.5 holds and the HJB equation admits a $C^{1,2}_\nu$ solution v, which satisfies the growth bound*

$$\|v(s,x)\|^2 + \|\partial_x v(s,x)\sigma(s,x)\|^2 + \|\tilde{v}\|_\nu^2 \le K(1+\|x\|^2)$$

(under Condition I) or $\|v(s,x)\|^2 \le K$ (under Condition II). Then $V_t = v(t,X_t)$ is the value function of our control problem.

Proof. This is simply an application of Theorem 19.5.1 or A.9.22. □

Corollary 21.4.11. *Under the conditions of this theorem, any optimal feedback control u satisfies*

$$f\big(t,x,(\partial_x v)\sigma,\tilde{v},u(t,x)\big) = H\big(t,x,(\partial_x v)\sigma,\tilde{v}\big) \qquad dt\times dP-\text{a.e.}$$

Theorem 21.4.12. *In the continuous case, given that b, σ, g and H satisfy the continuity and growth bounds of Remark 19.4.2 (where H is the BSDE driver), if H and ξ are uniformly continuous with respect to x, the value function $V_t = v(t, x)$ is a viscosity solution of the HJB equation. If H satisfies the uniqueness assumptions of Theorem 19.5.3, we see that it is the only viscosity solution of the HJB equation.*

Proof. This is an application of Theorem 19.5.3, see also Remark A.9.23. □

Remark 21.4.13. The continuity assumptions of Remark 19.4.2 can be weakened, provided one can show sufficient integrability for the forward process X under the reference dynamics.

Example 21.4.14. Consider the simple case where the controller determines the drift of a Brownian motion, so with a control $u \in \mathcal{U}$, we have the weak controlled dynamics

$$dX_t = u_t dt + dW_t.$$

Taking P to be the reference measure where $u \equiv 0$, we see that

$$f(t, x, z, u) = c(t, x, u) + zu$$

so $H(t, x, z) = \operatorname{ess\,inf}_{u \in U}\{c(t, x, u) + zu\}$, and the HJB equation becomes

$$0 = \frac{\partial v}{\partial t}(t, x) + \frac{1}{2}\frac{\partial^2 v}{\partial x^2}(t, x) + \operatorname{ess\,inf}_{u \in U}\left\{c(t, x, u) + u\frac{\partial v}{\partial x}\right\}$$

with terminal value $v(T, x) = \xi(x)$.

21.5 The Predicted Miss Problem

As an example of the above formulation for the optimal control problem we now describe a BSDE version of the treatment of Davis and Clark [42] of the "predicted miss" problem. Here the dynamics are described by a linear system, and the control values are restricted to the product interval $[-1, 1]^r$. The objective is to steer the system to a given hyperplane at the fixed terminal time $T = 1$. There is a natural candidate for the optimal control: in Beneš' [12] phrase it is full "bang" to reduce predicted miss. Because this candidate optimal control is not smooth, the classical approach of directly studying the HJB equation cannot be used. However, the above martingale/BSDE techniques can be applied.

Suppose B, σ, and Γ are, respectively, $\mathbb{R}^{d\times d}$, $\mathbb{R}^{d\times d}$ and $\mathbb{R}^{d\times r}$ valued deterministic functions of $t \in [0, 1]$, with $\sigma_t\sigma_t^\top$ strictly positive definite. Let the control set U be $[-1, 1]^r \subset \mathbb{R}^r$. An admissible control is a predictable process with values in U, and we write $\mathcal{U}$ for the family of such processes.

Reference Dynamics: We consider a filtered probability space in which there is a d-dimensional Brownian motion W, and $\{\mathcal{F}_t\}_{t\in[0,1]}$ is its completed, right-continuous natural filtration. For a given reference admissible control, which we take to be $u \equiv 0$ for simplicity, the 'forward' state process X is the unique strong solution to the equation

$$dX_t = B_t X_t dt + \sigma_t dW_t \tag{21.6}$$

with prescribed initial value $x_0 \in \mathbb{R}^d$.

Control effect: We model our control as adding a drift $\Gamma_t u_t dt$ to (21.6). Formally, for any $u \in \mathcal{U}$, define a measure P^u by

$$\frac{dP^u}{dP} = \mathcal{E}\big((\Gamma_s u_s)^\top \bullet W\big)_1,$$

so that

$$\begin{aligned} W_t^u &= \int_{[0,t]} \sigma_s^{-1} dX_s - \int_{[0,t]} \sigma_s^{-1}\big(B_s X_s + \Gamma_s u_s\big) ds \\ &= W_t - \int_{[0,t]} \sigma_s^{-1}\big(\Gamma_s u_s\big) ds \end{aligned}$$

defines an n-dimensional P^u-Brownian motion and we have dynamics

$$dX_t = B_t X_t dt + \Gamma_t u_t dt + \sigma_t dW_t^u.$$

Costs: Suppose $k \in \mathbb{R}^n$ and $\lambda : \mathbb{R} \to \mathbb{R}^+$ is such that

(i) $\lambda(\rho) = \lambda(-\rho)$ for all $\rho \in \mathbb{R}$,
(ii) $\lambda(\rho) = O\big(\exp(a|\rho|)\big)$ for some $a > 0$, so $\lambda(k^\top X_1) \in L^2(P)$ (by Lemma 15.5.7 and Grönwall's inequality).

The total expected cost corresponding to control $u \in \mathcal{U}$ is given by

$$J(u) = E_u[\lambda(k^\top X_1)],$$

That is, the objective of the controller is to minimize the distance of X_1 from the hyperplane $\{y : k^\top y = 0\}$ at the final time $T = 1$.

Note that this is purely a terminal cost $\xi = \lambda(k^\top X_1)$, and the running cost is zero.

BSDE dynamics: We define the BSDE driver

$$f(t, z, u) = z^\top \Gamma_t u_t,$$

and hence the Hamiltonian

$$H(t, z) = \inf_{u\in U}\{z^\top \Gamma_t u_t\} = \sum_i |z^\top \gamma_t^i|$$

where γ_t^i is the ith column of Γ_t. For a given control $u \in \mathcal{U}$, the expected miss distance is the solution to the BSDE with driver $f(\cdots, u)$, and the value function satisfies the BSDE

$$V_t = \Lambda(k^\top X_1) + \int_{]t,1]} H(t, Z_t)dt - \int_{]t,1]} Z_t dW_t.$$

As this is a Markovian BSDE, $V_t = v(t, X_t)$, where v is a viscosity solution of the PDE

$$0 = \frac{dv}{dt} + (\nabla v)^\top B_t x + \sum_{i,j} a_t^{ij}(x)\frac{\partial^2 v}{\partial x_i \partial x_j} + \sum_i |((\nabla v)\sigma_t)^\top \gamma_t^i|$$

with terminal value $v(1, x) = \lambda(k^\top x)$, where ∇v denotes the gradient of v with respect to x, and $a_t(x) = \sigma_t \sigma_t^\top$.

A control u is optimal if and only if $f(t, Z, u) = H(t, Z)$, which implies that (assuming v is differentiable),

$$u^i = \operatorname{sign}(Z_t^\top \gamma_t^i) = \operatorname{sign}(((\nabla v_t)\sigma_t)^\top \gamma_t^i) \qquad dt \times dP - a.e.$$

We see that the optimal control is “full bang” in the direction indicated by $(\nabla v)_t \sigma_t$.

21.6 Exercises

Exercise 21.6.1. Consider the setting where $N = 1$ and $\nu \equiv 0$ (i.e. we have only a single Brownian motion and no jumps). Let X be the controlled forward process with weak dynamics

$$dX_t = u_t dt + dW_t; \qquad X_0 = 0,$$

where u is a control in the set $U = [-1/3, 1/3]$. For the cost functions

- $c(x, t, u_t) = (1 - t) - 3x - 2(1 - t)u_t x$
- $\xi(x) = -x^3$, realized at time $T = 1$,

show that the value function is given by $v(t, x) = -x^2(x - 1 + t)$, and describe the optimal control.

Exercise 21.6.2. Let N be a Poisson process with controlled rate $u_t\lambda$, for $u_t \in U = [1/2, 2]$ and λ a fixed constant. Consider the control problem which attempts to minimize $E[(N_1 - \lambda)^2]$. Describe the optimal policy with cost $c(u) = 0$ and with cost $c(u) = u^2$.

22

Filtering

In this chapter, we suppose there is a signal process X which describes the state of a system, but which cannot be observed directly. Instead we can only observe some process Y with dynamics dependent on the value of X. Our object is to obtain an expression for the "best estimate" of X_t (or of $\phi(X_t)$ for ϕ in a large enough class of functions) given the observations up to time t, that is, given the σ-algebra

$$\mathcal{Y}_t = \sigma(Y_s : s \leq t).$$

This problem is known as 'filtering', as we attempt to filter out the state of the hidden 'signal' X given our (noisy) observations of Y.

The most successful result of this kind is that obtained for linear systems with Gaussian noise, developed by Kalman [115] and Kalman and Bucy [116] in 1960 and 1961, respectively. This has been applied in many fields and a proof is given below. Attempts have been made to extend this result to nonlinear systems and we shall describe the "innovations" and "reference probability" approaches to nonlinear filtering.

Equations giving the evolution of the conditional distribution of X were obtained in the 1960s by, for example, Bucy [28], Kushner [116], Shiryayev [167], Stratonovich [171] and Wonham [185]. In 1969, Zakai [190] showed how these results could be obtained in a simpler manner using his "reference probability" method. Kailath [112] defined the "innovations" approach to linear filtering in 1968, and it was quickly applied to the nonlinear case. It soon became clear that the filtering problem should be formulated in terms of martingales and the general theory of processes. The definitive result using the innovations approach was given by Fujisaki *et al.* [84] in 1972. Below we give new proofs of the general nonlinear filtering equation (22.3), and also the equation for the unnormalized conditional density.

S.N. Cohen, R.J. Elliott, *Stochastic Calculus and Applications*,
Probability and Its Applications, DOI 10.1007/978-1-4939-2867-5_22

Later work in nonlinear filtering was concerned, *inter alia*, with the following problems:

(i) the determination of finite dimensional nonlinear filters (Beneš [13]),
(ii) obtaining "robust" or "pathwise continuous" solutions of the filtering equations (Davis [41, 48], and Elliott and Kohlmann [70]),
(iii) developing a rigorous treatment of the theory of stochastic *partial* differential equations (Pardoux [145], Kunita [121]), which naturally arise in this setting (see (22.8)),
(iv) using Lie algebraic methods (Brockett [26]).

First note the fact that, if ϕ is some square integrable function of the history of the signal process $\{X_s\}_{s\le t}$, then the "best estimate" (in mean square) of $\phi(X_t)$ given the observations up to time t is

$$E[\phi(X_t)|\mathcal{Y}_t] = \int \phi(x)p_t(dx),$$

where p_t is the conditional probability distribution of X_t given $\mathcal{Y}_t$. Roughly speaking, the objective of our theory is to determine an expression for p_t, and to give this in a form where it is updated recursively in a memoryless manner. This will be done by expressing p_t in terms of a stochastic (partial) differential equation.

Recall, from Chapter 17, that if X is a Markov process satisfying an SDE of the form

$$dX_t = f(t,X_t)dt + \kappa(t,X_t)dB_t + \int_{\zeta\in\mathcal{Z}} g(\zeta,t,X_{t-})\tilde{\mu}(d\zeta,dt), \tag{22.1}$$

then it has an infinitesimal generator $\mathcal{L}_t$ given by

$$\begin{aligned}\mathcal{L}_s v(x) = &\sum_i f^i(s,x)\frac{\partial v}{\partial x^i}(x) + \frac{1}{2}\sum_{i,j} a^{ij}(s,x)\ \frac{\partial^2 v}{\partial x^i\partial x^j}(x)\\ &+ \int_{\mathcal{Z}}\Big(v(x+g(\zeta,s,x)) - v(x) - \sum_i \frac{\partial v}{\partial x^i}(x)g^i(\zeta,s,x)\Big)\nu(d\zeta)\,,\end{aligned}$$

where $a=\kappa\kappa^\top$ (we use κ in place of σ here to prevent notational confusion). If we write $X^{(t,x)}$ for the solution of this SDE started at time t in state x, then, in Theorem 17.4.14, we saw that, given appropriate integrability and differentiability conditions, the probability density of $X_s^{(t,x)}$, if it exists, satisfies the Kolmogorov equations

$$-\frac{\partial p}{\partial s} = \mathcal{L}_s p(s,\cdot\,;t,y), \qquad \text{(Backward equation)},$$
$$\frac{\partial p}{\partial t} = \mathcal{L}_t^* p(s,x;t,\cdot), \qquad \text{(Forward equation)},$$

where $\mathcal{L}_t^*$ is the adjoint of $\mathcal{L}_t$, and is given by, for $p \in C^{1,2}$,

$$\mathcal{L}_t^* p(t,y) = \sum_i \frac{\partial [f^i p]}{\partial y^i}(t,y) + \frac{1}{2}\sum_{i,j} \frac{\partial^2 [a^{ij} p]}{\partial y^i \partial y^j}(t,y)$$
$$+ \int_{\mathcal{Z}} \Big(p(t, y - g(\zeta, t, y)) - p(t,y) + \sum_i \frac{\partial [g^i(\zeta,\cdot,\cdot)p]}{\partial y^i}(t,y) \Big) \nu(d\zeta)\,.$$

22.1 The Innovations Approach

We shall assume that all processes are defined on a fixed probability space $(\Omega, \mathcal{F}, P)$ for time $t \in [0,T]$. Suppose there is a right continuous filtration $\{\mathcal{F}_t\}_{t\in[0,T]}$, of sub-$\sigma$-algebras of $\mathcal{F}$, and that the filtration is complete (each $\mathcal{F}_t$ contains all null sets of $\mathcal{F}$).

Within this space, we suppose we are given two processes – a signal process X and an observation process Y, both adapted to the filtration $\{\mathcal{F}_t\}_{t\in[0,T]}$. We shall say more about the signal process X later, suffice for now to say that it is a Markov process which takes values in $\mathbb{R}^d$. We first focus on the observation process Y. Writing $\{\mathcal{Y}_t\}_{t\in[0,T]}$ for the complete σ-algebra generated by Y (that is, $\mathcal{Y}_t = \sigma(Y_s : s \le t)$ up to null sets), we have $\mathcal{Y}_t \subset \mathcal{F}_t$, the inclusion being strict in general.

We shall suppose that the *observation process* Y is an m-dimensional semimartingale of the form

$$Y_t = \int_{[0,t]} c(s, X, Y)ds + \int_{[0,t]} \alpha(s, Y)dW_s, \tag{22.2}$$

where

(i) W is a standard m-dimensional Brownian motion,

(ii) $\alpha : [0,T] \times \mathcal{C}([0,T];\mathbb{R}^m) \to \mathbb{R}^{m\times m}$ is a nonanticipative nonsingular matrix valued function, such that there exists $k > 0$ with

$$\|\alpha(t,0)\| + \|\alpha(t,y)^{-1}\| \le k,$$

for all t, y, and which satisfies a uniform Lipschitz condition of the form

$$\|\alpha(t,y) - \alpha(t,y')\| \le K(y-y')_t^*,$$

(iii) $c : [0,T] \times \mathbb{R}^d \times \mathcal{C}([0,T];\mathbb{R}^m) \to \mathbb{R}^m$ and, for simplicity, we suppose that c is bounded, nonanticipative and uniformly Lipschitz in y, that is,

$$\|c(t,x,y) - c(t,x,y')\| \le K(y-y')_t^*.$$

We further suppose c is Borel measurable and nonanticipative in (t, x).

Remark 22.1.1. We allow the path Y to appear as an argument of c and α, provided it does so in a nonanticipative way and is such that we have a unique strong solution to (22.2).

Remark 22.1.2. We have assumed here that Y and W are of the same dimension and that Y is a continuous process. Neither of these assumptions is particularly significant; however, they will significantly simplify our notation.

Remark 22.1.3. Note that we have not included X in the volatility of Y. Suppose we instead modelled Y with volatility $\alpha(s, X_s, Y)$, and that the map $x \mapsto \alpha(s, x, Y)^2$ were bijective. Then we would simply calculate $\langle Y \rangle_t = \int_{[0,t]} \alpha(s, X, Y)^2 dt$, which is necessarily adapted to the filtration generated by Y. Differentiating, we see that the value of $\alpha(s, X_s, Y)^2$ is $\mathcal{Y}_s$-measurable and, therefore, the value of X_s is $\mathcal{Y}_s$-measurable. In this way, we see that the case when X appears in the volatility leads to a very different (and often simpler) filtering problem in general.

Definition 22.1.4. *If η is any process, we write $\widehat{\eta}$ for its $\{\mathcal{Y}_t\}_{t\in[0,T]}$-optional projection (cf. Section 7.6). For each t, from Theorem 7.6.5 we know that*

$$\widehat{\eta}_t = E[\eta_t | \mathcal{Y}_t] \qquad P-\text{a.s.}.$$

Similarly, as c is a function of X and Y, we define $\widehat{c} : \Omega \times [0,T] \times \mathcal{C}([0,T]; R^m) \to \mathbb{R}^m$ to be the $\{\mathcal{Y}_t\}_{t\in[0,T]}$-optional projection of the mapping c, in the sense that for any $y \in \mathcal{C}([0,T]; R^m)$,

$$\widehat{c}(\omega, t, y) = E[c(t, X, y)|\mathcal{Y}_t].$$

Note that we clearly have $\widehat{c}(\omega, t, Y) = E[c(t, X, Y)|\mathcal{Y}_t]$, from the definition of the projection.

Remark 22.1.5. We again need to be careful to define $\widehat{c}$ simultaneously for *all* paths y, as the projection is only defined up to indistinguishability. As usual, this can be done by taking a countable dense set of paths y (this exists by the Stone–Wierstrass theorem, see Royden and Fitzpatrick [160, p.247]), defining $\widehat{c}$ for these paths, and then extending $\widehat{c}$ to all y by Lipschitz continuity. We also observe that $\widehat{c}$ is nonanticipative and uniformly Lipschitz with respect to y.

Definition 22.1.6. *The process $\{V_t\}_{0\leq t\leq T}$ defined by*

$$\begin{cases} V_t = \displaystyle\int_{[0,t]} \alpha(s, Y)^{-1} dY_s - \int_{[0,t]} \alpha(s, Y)^{-1} \widehat{c}_s(\omega, s, Y) ds, \\ V_0 = 0 \in \mathbb{R}^m, \end{cases}$$

is called the innovations process.

This terminology is motivated by the observation that, formally, $V_{t+h} - V_t$ represents the 'new' information about X obtained from observations between t and $t + h$.

Lemma 22.1.7. *V is a Brownian motion with respect to the filtration $\{\mathcal{Y}_t\}_{t\in[0,T]}$.*

Proof. We first prove that V is a $\{\mathcal{Y}_t\}_{t\in[0,T]}$-martingale. For $s \leq t$, by Fubini's theorem, as $\alpha(u,Y)$ is $\mathcal{Y}_u$-measurable,

$$\begin{aligned}
&E[V_t - V_s|\mathcal{Y}_s]\\
&= E\Big[\int_{]s,t]} \alpha(u,Y)^{-1}\big(dY_u - \widehat{c}(\omega,u,Y)du\big)\Big|\mathcal{Y}_s\Big]\\
&= E\Big[\int_{]s,t]} \alpha(u,Y)^{-1}\big(c(u,X_u,Y) - \widehat{c}(\omega,u,Y)\big)du + W_t - W_s\Big|\mathcal{Y}_s\Big]\\
&= 0 \quad \text{a.s.}
\end{aligned}$$

With respect to the filtration $\{\mathcal{F}_t\}_{t\in[0,T]}$, $V_t = (V_t^1, \dots, V_t^m)$ is a continuous m-dimensional semimartingale. By Itô's rule,

$$V_t^i V_t^j = \int_{[0,t]} V_u^i dV_u^j + \int_{[0,t]} V_u^j dV_u^i + \langle W^i, W^j\rangle_t \ .$$

Therefore, with respect to the $\{\mathcal{F}_t\}_{t\in[0,T]}$ filtration, $\langle V^i, V^j\rangle_t = \langle W^i, W^j\rangle_t = \delta_{ij}t$. This process is deterministic, so with respect to the (smaller) $\{\mathcal{Y}_t\}_{t\in[0,T]}$ filtration, $\langle V^i, V^j\rangle_t = \delta_{ij}t$. We also know V is continuous, so by Corollary 14.4.2, V is an m-dimensional Brownian motion with respect to the filtration $\{\mathcal{Y}_t\}_{t\in[0,T]}$. □

Theorem 22.1.8. *V has the predictable representation property in the filtration $\{\mathcal{Y}_t\}_{t\in[0,T]}$. That is, if M is a local martingale with respect to the filtration $\{\mathcal{Y}_t\}_{t\in[0,T]}$, then there is a $\{\mathcal{Y}_t\}_{t\in[0,T]}$-predictable, V-integrable m-dimensional process H such that $M = E[M_0] + H^\top \bullet V$ up to indistinguishability.*

Proof. By assumption, we know that $\int_{[0,T]} \|\alpha(s,Y)^{-1}c(s,X_s,Y)\|^2 ds$ is bounded. Let Q be the measure with density

$$\frac{dQ}{dP} = \mathcal{E}\Big(-\int_{[0,t]} \big(\alpha(s,Y)^{-1}c(s,X,Y)\big)^\top dW_s\Big)_T.$$

Then, by Girsanov's theorem (Theorem 15.2.6), it is easy to check that $Y = \alpha(s,Y)\bullet\tilde{Y}$, for a Q-Brownian motion

$$\tilde{Y}_t := W_t - \int_{[0,t]} \alpha(s,Y)^{-1}c(s,X,Y)ds = \int_{[0,t]} \alpha(s,Y)^{-1}dY_s.$$

Clearly, $\tilde{Y}$ is $\{\mathcal{Y}_t\}_{t\in[0,T]}$-adapted. Consider the filtration generated by $\tilde{Y}$, denoted $\{\tilde{\mathcal{Y}}_t\}_{t\in[0,T]}$. By Theorem 14.5.1, $\tilde{Y}$ has the predictable representation property in $\{\tilde{\mathcal{Y}}_t\}_{t\in[0,T]}$.

As $y \mapsto \alpha(s,y)$ is uniformly Lipschitz, from Theorem 16.3.11 we see that Y is the unique solution (in any filtration containing $\tilde{Y}$) to the SDE

$dY_s = \alpha(s,Y)d\tilde{Y}_s$. It follows that Y is $\{\tilde{\mathcal{Y}}_t\}_{t\in[0,T]}$-adapted. Therefore, $\tilde{\mathcal{Y}}_t = \mathcal{Y}_t$, and $\tilde{Y}$ has the predictable representation property in $\{\mathcal{Y}_t\}_{t\in[0,T]}$, under the measure Q.

Finally, we have that

$$V_t = \tilde{Y}_t - \int_{[0,t]} \alpha(s,Y)^{-1}\widehat{c}(\omega,s,Y)ds,$$

and so, as the predictable representation property is appropriately preserved under changes of measure (Theorem 15.2.8), we see that V also has the predictable representation property in $\{\mathcal{Y}_t\}_{t\in[0,T]}$, under the measure P. □

To obtain the general filtering equation, we shall now consider a real $\{\mathcal{F}_t\}_{t\in[0,T]}$-semimartingale ξ and obtain a stochastic differential equation satisfied by $\widehat{\xi}$. The kind of semimartingale we have in mind is some real valued function ϕ of the signal process, that is, $\xi_t = \phi(X_t)$. The differential equation we obtain will provide the recursive and memoryless filter for $\widehat{\xi}$.

Theorem 22.1.9. *Suppose ξ is a real $\{\mathcal{F}_t\}_{t\in[0,T]}$-semimartingale of the form*

$$\xi_t = \xi_0 + \int_{[0,t]} \beta_s ds + N_t.$$

We assume $E[\xi_0^2] < \infty$, $E\big[\int_{[0,T]}\beta_s^2 ds\big] < \infty$ and N is a square integrable $\{\mathcal{F}_t\}_{t\in[0,T]}$-martingale with $\langle N, W^i\rangle_t = \int_{[0,t]}\lambda_s^i ds$, for λ a $\mathbb{R}^m$-valued $\{\mathcal{F}_t\}_{t\in[0,T]}$-predictable process.

Suppose our observation process Y is of the form discussed above, and V is the associated innovations process. Then $\widehat{\xi}$ solves the stochastic differential equation

$$\widehat{\xi}_t = \widehat{\xi}_0 + \int_{[0,t]} \widehat{\beta}_s ds + \int_{[0,t]} \big(\widehat{\lambda}_s + \alpha^{-1}(s,Y)\big(R_s - \widehat{\xi}_s\,\widehat{c}(\omega,s,Y)\big)\big)^\top dV_s, \quad (22.3)$$

where R is the $\{\mathcal{Y}_t\}_{t\in[0,T]}$-optional projection of $\{\xi_s c(s,X,Y)\}_{s\in[0,T]}$.

Proof. The proof is an extension of an idea of Wong [184]. Define

$$M_t = \widehat{\xi}_t - \widehat{\xi}_0 - \int_{[0,t]} \widehat{\beta}_u du.$$

Then, for $0 \le s \le T$,

$$E[M_t - M_s|\mathcal{Y}_s] = E\Big[\widehat{\xi}_t - \widehat{\xi}_s - \int_{[s,t]} \widehat{\beta}_u du\Big|\mathcal{Y}_s\Big].$$

However, as $\widehat{\xi}$ is the projection of ξ,

$$\begin{aligned}E[\widehat{\xi}_t - \widehat{\xi}_s|\mathcal{Y}_s] &= E[\xi_t - \xi_s|\mathcal{Y}_s] \\ &= E\Big[\int_{]s,t]} \beta_u du\Big|\mathcal{Y}_s\Big] + E[N_t - N_s|\mathcal{Y}_s] \\ &= E\Big[\int_{]s,t]} E[\beta_u|\mathcal{Y}_u]du\Big|\mathcal{Y}_s\Big] + E\big[E[N_t - N_s|\mathcal{F}_s]\big|\mathcal{Y}_s\big] \\ &= E\Big[\int_{]s,t]} \widehat{\beta}_u du\Big|\mathcal{Y}_s\Big],\end{aligned}$$

because N is an $\{\mathcal{F}_t\}_{t\in[0,T]}$-martingale. Therefore, M is a local martingale, so, by Theorem 22.1.8, there is a $\{\mathcal{Y}_t\}_{t\in[0,T]}$-predictable process H such that

$$M_t = \int_{[0,t]} H_u^\top \, dV_u \quad \text{a.s.}$$

and we can write

$$\widehat{\xi}_t = \widehat{\xi}_0 + \int_{[0,t]} \widehat{\beta}_u du + \int_{[0,t]} H_u^\top dV_u \,. \tag{22.4}$$

(As usual, the $^\top$ denotes the transpose of column vector H_u to row vector $H_u^\top$.)

We now wish to determine H. By Itô's rule,

$$\begin{aligned}\xi_t Y_t = \xi_0 Y_0 &+ \int_{[0,t]} \xi_u \big(c(u,X,Y)du + \alpha(u,Y)dW_u\big) \\ &+ \int_{[0,t]} Y_u(\beta_u du + dN_u) + \int_{[0,t]} \alpha(u,Y)\lambda_u du \,.\end{aligned} \tag{22.5}$$

The integrals

$$J_t^1 := \int_{[0,t]} \xi_u \alpha(u,Y)dW_u \qquad \text{and} \qquad J_t^2 := \int_{[0,t]} Y_u dN_u$$

are locally square integrable $\{\mathcal{F}_t\}_{t\in[0,T]}$-local martingales. The conditions on the components of ξ imply that $E[\xi_t^2] < \infty$ for all $t \in [0,T]$. For each $n \in \mathbb{N}$, define

$$S_n := \sup\{u : \|\alpha(u,Y_u)\| \le n\}, \qquad T_n := \sup\{u : |Y_u| \le n\}$$

so that S_n and T_n are $\{\mathcal{Y}_t\}_{t\in[0,T]}$-stopping times. Then, for $t \in [0,T]$,

$$E\Big[\Big(\int_{[0,t\wedge S_n]} \xi_u \alpha(u,Y)dW_u\Big)^2\Big] \le n^2 \int_{[0,t]} E[\xi_u^2]du < \infty$$

and

$$E\Big[\Big(\int_{[0,t\wedge T_n]} Y_u dN_u\Big)^2\Big] \leq n^2 E\big[\langle N\rangle_T\big] < \infty.$$

Consequently,

$$\begin{aligned} E\big[\widehat{J}^1_{t\wedge S_n} - \widehat{J}^1_{s\wedge S_n}\big|\mathcal{Y}_s\big] &= E\big[E[J^1_{t\wedge S_n}|\mathcal{Y}_t] - E[J^1_{s\wedge S_n}|\mathcal{Y}_s]\big|\mathcal{Y}_s\big] \\ &= E\big[E[J^1_{t\wedge S_n} - J^1_{s\wedge S_n}|\mathcal{F}_s]\big|\mathcal{Y}_s\big] = 0, \end{aligned}$$

so $\widehat{J}^1$ is a locally square integrable $\{\mathcal{Y}_t\}_{t\in[0,T]}$-local martingale, and similarly for $\widehat{J}^2$.

Consider the processes defined by

$$\begin{aligned} K^1_t &= \int_{[0,t]} \xi_u c(u,X,Y)du, \\ K^2_t &= \int_{[0,t]} Y_u \beta_u du, \\ K^3_t &= \int_{[0,t]} \alpha(u,Y)\lambda_u du. \end{aligned}$$

Then, by a calculation similar to that for M above, the processes

$$\begin{aligned} \widetilde{K}^1_t &= \widehat{K}^1_t - \int_{[0,t]} R_u du, \\ \widetilde{K}^2_t &= \widehat{K}^2_t - \int_{[0,t]} Y_u \widehat{\beta}_u du, \\ \widetilde{K}^3_t &= \widehat{K}^3_t - \int_{[0,t]} \alpha(u,Y)\widehat{\lambda}_u du, \end{aligned}$$

are local martingales with respect to the filtration $\{\mathcal{Y}_t\}_{t\in[0,T]}$. Therefore, from (22.5),

$$\begin{aligned} (\widehat{\xi Y})_t &= \widehat{\xi}_t Y_t \\ &= \widehat{\xi}_0 Y_0 + \widehat{J}^1_t + \widehat{J}^2_t \\ &\quad + \Big(\widetilde{K}^1_t + \int_{[0,t]} R_u du\Big) + \Big(\widetilde{K}^2_t + \int_{[0,t]} Y_u \widehat{\beta}_u\, du\Big) \\ &\quad + \Big(\widetilde{K}^3_t + \int_{[0,t]} \alpha(u,Y)\widehat{\lambda}_u\, du\Big). \end{aligned} \tag{22.6}$$

Because this represents $(\widehat{\xi\, Y})$ as the sum of local martingales plus continuous (and so predictable) finite variation processes, we see $(\widehat{\xi\, Y})$ is a special semimartingale with respect to the filtration $\{\mathcal{Y}_t\}_{t\in[0,T]}$. However, using (22.4) and Itô's rule,

$$\begin{aligned}\widehat{\xi}_t Y_t = \widehat{\xi}_0 Y_0 &+ \int_{[0,t]} \widehat{\xi}_u(\widehat{c}(\omega,u,Y)du + \alpha(u,Y)dV_u) \\ &+ \int_{[0,t]} Y_u(\widehat{\beta}_u du + H_u^\top\, dV_u) + \int_{[0,t]} \alpha(u,Y)H_u\, du.\end{aligned} \tag{22.7}$$

The integrals with respect to V are again local martingales, and the remaining integrals give continuous, and so predictable, processes. The two canonical decompositions of the special semimartingale $(\widehat{\xi\, Y})$ must be the same, so equating the integrands in the finite variation terms,

$$R_u + \alpha(u,Y)\widehat{\lambda}_u = \widehat{\xi}_u\widehat{c}(\omega,u,Y) + \alpha(u,Y)H_u \quad dt \times dP\text{-a.e.}$$

Therefore,

$$H_u = \alpha(u,Y)^{-1}(R_u - \widehat{\xi}_u\widehat{c}(\omega,u,Y)) + \widehat{\lambda}_u.$$

Substituting in (22.4), the result follows. □

22.1.1 A More Concrete Equation

By specifying the dynamics of X and assuming that only the current state of X affects the dynamics of Y, we can reduce this equation to a concrete form. Suppose that the signal process X is the unique strong solution of the stochastic differential equation (22.1), that is,

$$dX_t = f(t,X_t)dt + \kappa(t,X_t)dB_t + \int_{\zeta\in\mathcal{Z}} g(\zeta,t,X_{t-})\tilde{\mu}(d\zeta,dt),$$

with $\mathcal{F}_0$-measurable initial condition $X_0 = x \in \mathbb{R}^d$. Suppose that ϕ is a twice continuously differentiable function on $\mathbb{R}^d$ with bounded first and second derivatives (i.e. $\phi \in C_b^2$, or more generally, $\phi \in C_\nu^2$, where ν is the compensator of the random measure μ, cf. Definition 17.4.9). Then Itô's rule implies

$$\phi(X_t) = \phi(X_0) + \int_{[0,t]} \mathcal{L}_u\phi(X_u)du + N_t,$$

where

$$N_t = \int_{[0,t]} \nabla\phi\cdot\kappa(u,X_u)dB_u + \int_{\mathcal{Z}\times[0,t]} \big(\phi(u,X_{u-}+\zeta) - \phi(u,X_{u-})\big)\tilde{\mu}(d\zeta,du).$$

Furthermore, suppose $E\big[\int_{[0,T]}(\mathcal{L}_u\phi(X_u))^2du\big] < \infty$, N is a square integrable martingale and $\langle B^i, W^j\rangle_t = \int_{[0,t]}\rho_u^{ij}du$ for some $\{\mathcal{F}_t\}_{t\in[0,T]}$-predictable processes ρ^{ij}. The observation process Y will be as above, with the restriction that the drift can be written $c(t,X_t,Y)$, so only the current state of X is relevant. We then obtain the following corollary to Theorem 22.1.9.

Corollary 22.1.10. *For $\phi \in C_b^2$, we write $\pi_t(\phi) = \widehat{\phi(X_t)}$, so that $\pi_t(\phi) = E[\phi(X_t)|\mathcal{Y}_t]$. Similarly, we write $\pi_t(\phi) = \phi\widehat{(X_t, Y)}$ for functions depending continuously on Y. Then*

$$\begin{aligned}\pi_t(\phi) = \pi_0(\phi) &+ \int_{[0,t]} \pi_u(\mathcal{L}_u\phi)du \\ &+ \int_{[0,t]} \left\{\pi_u(\nabla\phi\cdot\kappa\cdot\rho) + \alpha^{-1}(u,Y)\big(\pi_u(\phi c) - \pi_u(\phi)\pi_u(c)\big)\right\}^\top dV_u\,.\end{aligned} \tag{22.8}$$

Proof. $\phi(X)$ plays the role of the semimartingale ξ in Theorem 22.1.9. We know that N is a martingale and, for $1 \le j \le m$,

$$\langle N, W^j\rangle_t = \int_{[0,t]} \nabla\phi\,\kappa(u, X_{u-})\,\rho_u^j\,du,$$

where $\rho_u^j = (\rho_u^{1j}, \ldots, \rho_u^{dj})^\top$ is the jth column of the matrix ρ with entries ρ^{ij}. Therefore, in the notation of Theorem 22.1.9,

$$\lambda_u^j = \nabla\phi\cdot\kappa(u, X_{u-})\cdot\rho_u^j\,.$$

Given the notational equivalence $\widehat{c}(\omega, t, Y) = \pi_t(c)$, substituting in the formula of Theorem 22.1.9 gives the stated result.

□

Remark 22.1.11. As the right-hand side of equation (22.3) involves $R = \widehat{\xi_u c(\ldots)}$ (as well as $\widehat{\xi}_u$), it is not recursive in $\widehat{\xi}_u$. However, the formula of Corollary 22.1.10 can be considered as a recursive stochastic differential equation for π_t, the conditional probability distribution of X_t given $\mathcal{Y}_t$, because

$$\pi_t(\phi) = E[\phi(X_t)|\mathcal{Y}_t] = \int_{\mathbb{R}^d} \phi(x)\pi_t(dx),$$

for $\phi \in C_b^2$.

However, this is then a stochastic differential equation with a variable in the infinite dimensional space of probability measures. Only in certain special cases is it possible to obtain finite dimensional recursive filters, even for the conditional mean $\widehat{X}_t$.

Remark 22.1.12. Suppose now that the signal and observation noise are independent, so that $\langle B^i, W^j\rangle_t = 0$, and that $\alpha(u, Y) = aI$, where $a > 0$ and I is the $d \times d$ identity matrix. Then, as $a \to \infty$ the observations become infinitely noisy, so give no information about the signal, and equation (22.8) reduces to

$$\pi_t(\phi) = \pi_0(\phi) + \int_{[0,t]} \pi_u(\mathcal{L}_u\phi)du,$$

the same equation as given by Dynkin's formula (Remark 17.4.5) for the unconditional expectation $E[\phi(X_t)]$.

Remark 22.1.13. As we do not depend at all on the fact that X is a strong solution to (22.1), or on the structure of $\{\mathcal{F}\}_{t\geq 0}$, it is equally valid to consider the setting where X is an $\{\mathcal{F}\}_{t\geq 0}$-Markov process which solves the SDE

$$dX_t = f(t, X_t)dt + \kappa(t, X_t)dB_t + \int_{\mathbb{R}^d} \zeta\, \tilde{\mu}^X(d\zeta, dt),$$

where $\tilde{\mu}^X(d\zeta, dt) = \mu^X(d\zeta, dt) - \nu(t, X_t; d\zeta)dt$. This is of the form considered in Chapter 18. Given these changes, the only difference is in the form of the infinitesimal generator $\mathcal{L}$, which is instead as given in (18.4). On the other hand, it is significant that Y is a *strong* solution to (22.2), and is also adapted to the filtration $\{\mathcal{F}\}_{t\geq 0}$, as this was used to establish that the innovations process has the predictable representation property.

Corollary 22.1.14. *Suppose that the signal and observation noise are independent and that the conditional distribution of X_t given $\mathcal{Y}_t$ has a density $\widehat{p}(\omega, t, x)$ which satisfies suitable differentiability hypotheses, similar to those in Theorem 17.4.14 (see Lipster and Shiryayev [126, Section 8.6] for details). Then, with the hypotheses and notation of Corollary 22.1.10,*

$$\pi_t(\phi) = \int_{\mathbb{R}^d} \phi(x)\widehat{p}(\omega, t, x)dx.$$

Using integration by parts,

$$\pi_t(\mathcal{L}_t\phi) = \int_{\mathbb{R}^d} \mathcal{L}_t\phi(x)\widehat{p}(\omega, t, x)dx = \int_{\mathbb{R}^d} \phi(x)\mathcal{L}_t^*\widehat{p}(\omega, t, x)dx.$$

Equation (22.8) holds for all twice continuously differentiable functions with compact support, so we obtain the following recursive, infinite dimensional equation for $\widehat{p}$:

$$d\widehat{p} = \mathcal{L}_t^*\widehat{p}\,dt + \alpha^{-1}\big(c - \pi_t(c)\big)\widehat{p}\,dV_t\,. \tag{22.9}$$

Given the presence of the differential operator $\mathcal{L}^*$, this equation can be thought of as a stochastic *partial* differential equation. It is the analog of the Kolmogorov forward equation (Theorem 17.4.14) and, if $\alpha = aI$ as above, it converges to it as $a \to \infty$. Unfortunately (22.9) is further complicated by the term $\pi_t(c) = \int_{\mathbb{R}^d} c(t, x, Y)\widehat{p}(t, x)dx$, and so is a nonlinear equation in $\widehat{p}$. We shall not further consider the theory of these equations, the interested reader should consult the introductory lecture notes of Martin Hairer [91], or the classic works of Da Prato and Zabczyk [39] and Walsh [182]. For a more specialized approach relevant for filtering, see Bain and Crisan [2].

Remark 22.1.15. Suppose the signal and observation noise are independent, that $d = 1$ and $\phi(x) = x$. Substituting in (22.8) we obtain the following equation for the conditional mean $\widehat{X}_t = \pi_t(X)$:

$$\widehat{X}_t = \widehat{X}_0 + \int_{[0,t]} \pi_u(f)du + \int_0^t \big(\pi_u(Xc) - \widehat{X}_u\pi_u(c)\big)dV_u\,. \tag{22.10}$$

Therefore, to calculate $\widehat{X}_t$ we need to know $\pi_u(f)$, $\pi_u(Xc)$ and $\pi_u(h)$, so the equation is not, in general, recursive. One situation where a recursive, finite dimensional filter for $\widehat{X}_t$ is obtained is when the equations for the signal and observation are both linear, with Brownian motion noise. The Kalman–Bucy filter is then obtained, and we describe its derivation in the next result.

Theorem 22.1.16 (Kalman–Bucy Filter). *Suppose, for simplicity of exposition, that both the signal process X and the observation process Y are one-dimensional and given by the following linear equations,*

$$\begin{cases} X_t = X_0 + \int_{[0,t]} aX_u du + b\,B_t, \\ Y_t = \int_{[0,t]} cX_u du + W_t, \end{cases}$$

where

- *B and W are independent Brownian motions, so $\langle B, W\rangle_t = 0$,*
- *a, b and c are constants, and*
- *X_0 is an $\mathcal{F}_0$-measurable Gaussian random variable (and so independent of B).*

Then $\widehat{X}$ is given by the following finite dimensional recursive filtering equation

$$\widehat{X}_t = \widehat{X}_0 + \int_0^t a\widehat{X}_u du + c\int_0^t P_u dV_u\,. \tag{22.11}$$

Here V is the innovations process given by $dV_u = dY_u - c\widehat{X}_u du$ and $P_t = E[(X_t - \widehat{X}_t)^2|\mathcal{Y}_t]$ is the conditional variance of the error. Furthermore, P_t is the solution of the (deterministic) Ricatti equation

$$\frac{dP_t}{dt} = 2aP_t + b^2 - c^2P_t^2\,.$$

Proof. Substituting in (22.10), we have immediately

$$\widehat{X}_t = \widehat{X}_0 + \int_{[0,t]} a\widehat{X}_u du + c\int_{[0,t]} [\pi_u(X_u^2) - \widehat{X}_u^2]dV_u\,, \tag{22.12}$$

where $dV_u = dY_u - c\widehat{X}_u du$ describes the innovations process. Write

$$P_t = \pi_t(X_t^2) - \widehat{X}_t^2 = E[(X_t - \widehat{X}_t)^2|\mathcal{Y}_t].$$

We wish to show that P has the specified dynamics.

First note that the joint process $\{X_t, Y_t\}_{t\in[0,T]}$ is Gaussian. To see this, note that we can explicitly express X_t as

$$X_t = e^{at}\Big(X_0 + b\int_{[0,t]} e^{-au}dB_u\Big).$$

Then, as

$$Y_t = \int_0^t cX_u du + W_t,$$

both X_t and Y_t are limits of sums of jointly Gaussian random variables, and so are jointly Gaussian.

For $t \in [0,T]$ consider an increasing sequence $\{\Pi_n\}_{n\in\mathbb{N}}$ of partitions of $[0,t]$, where

$$\Pi_n = \{0 = t_0^n < t_1^n < \cdots < t_{N_n}^n = t\}$$

and $\cup_n \Pi_n$ is dense in $[0,t]$. By standard arguments for Gaussian distributions, we see that the finitely conditioned expectation $E[X_t|\sigma(Y_\tau : \tau \in \Pi_n)]$ is Gaussian. Because Y is continuous, $\mathcal{Y}_t$ is, up to null sets, the limit of the increasing family of σ-algebras $\sigma(Y_\tau : \tau \in \Pi_n)$. Therefore, by the martingale convergence of Corollary 4.4.5,

$$E[X_t|\mathcal{Y}_t] = \widehat{X}_t = \lim_{n\to\infty} E[X_t|\sigma(Y_\tau : \tau \in \Pi_n)]$$

and so $\widehat{X}_t$ is Gaussian.

Write $\mathcal{K}(Y,t)$ for the subspace of $L^2(\Omega,\mathcal{F},P)$ spanned by the random variables $\{Y_s\}_{s\in[0,t]}$. Then $\widehat{X}_t$ is the projection of X_t onto $\mathcal{K}(Y,t)$ and $X_t - \widehat{X}_t$ is orthogonal to $\mathcal{K}(Y,t)$. Because we are dealing with Gaussian random variables, $X_t - \widehat{X}_t$ is, therefore, independent of the random variables $\{Y_s\}_{s\in[0,t]}$, that is, $X_t - \widehat{X}_t$ is independent of $\mathcal{Y}_t$.

Consequently,

$$E[(X_t - \widehat{X}_t)^2|\mathcal{Y}_t] = E[(X_t - \widehat{X}_t)^2] = P_t\,,$$

so P_t is deterministic. Furthermore,

$$E[(X_t - \widehat{X}_t)^3|\mathcal{Y}_t] = E[(X_t - \widehat{X}_t)^3] = 0. \tag{22.13}$$

From (22.12), we have that

$$\begin{aligned}(\widehat{X}_t)^2 &= (\widehat{X}_0)^2 + 2\int_{[0,t]} \widehat{X}_u d\widehat{X}_u + c^2 \int_{[0,t]} P_u^2 du \\ &= (\widehat{X}_0)^2 + \int_{[0,t]} \big(2a(\widehat{X}_u)^2 + c^2 P_u^2\big) du + 2c\int_{[0,t]} \widehat{X}_u P_u dV_u\,.\end{aligned} \tag{22.14}$$

Substituting in the general filtering equation (22.8) for $\phi(x) = x^2$,

$$\begin{aligned}\pi_t(X^2) &= E[X_t^2|\mathcal{Y}_t] \\ &= \pi_0(X^2) + \int_{[0,t]} \big(2a\pi_u(X^2) + b^2\big) du \\ &\quad + \int_{[0,t]} c\big(\pi_u(X^3) - \widehat{X}_u \pi_u(X^2)\big) dV_u\,.\end{aligned} \tag{22.15}$$

Subtracting (22.14) from (22.15),

$$\begin{aligned} P_t &= \pi_t(X^2) - (\widehat{X}_t)^2 \\ &= P_0 + \int_{[0,t]} (2aP_u + b^2 - c^2 P_u^2)du \\ &\quad + c\int_{[0,t]} \big(\pi_u(X^3) + 2(\widehat{X}_u)^3 - 3\widehat{X}_u\pi_u(X^2)\big)dV_u\,. \end{aligned} \tag{22.16}$$

However, we have observed above that P_t is deterministic, so the integrand in the above stochastic integral must be zero. This can also be seen directly, as

$$E[(X_u - \widehat{X}_u)^3|\mathcal{Y}_u] = \pi_u(X^3) + 2(\widehat{X}_u)^3 - 3\widehat{X}_u\pi(X^2),$$

which is zero by (22.13). Therefore,

$$P_t = P_0 + \int_{[0,t]} (2aP_u + b^2 - c^2P_u^2)du. \tag{22.17}$$

□

Remark 22.1.17. The quantity P_t represents the 'tracking error' or 'gain'. The fact that it is deterministic relies very heavily on the Gaussian nature of the X and Y processes. The nonlinear (quadratic) equation (22.17) satisfied by P_t is known as a Ricatti equation. Standard results for ordinary differential equations imply it has a unique solution. We see, from equations (22.12) and (22.15), how, in order to compute conditional moments of X_t, a knowledge of higher conditional moments is required. (For example, in (22.15) to obtain $\pi_t(X_t^2)$ we need $\pi_t(X^3)$.) However, in the (conditionally) Gaussian case all higher moments $\pi_t(X^n)$ can be expressed in terms of $\pi_t(X^2)$ and $\pi_t(X) = \widehat{X}_t$. This situation is extensively investigated in the books of Lipster and Shirayev [126] and Kallianpur [114], where the analogous Kalman–Bucy formulae for $\widehat{X}_t$ and P_t are given, when X and Y are described by linear vector equations with deterministic coefficients.

22.2 The Reference Probability Method

In this section, we give a different approach to the filtering problem. This approach was first considered by Zakai [190] and has the advantage that it often yields simpler calculations, as the SPDE obtained for the density is linear. Fundamentally, the idea is to avoid working with the innovations process V, by focussing instead on the measure Q constructed in the course of proving Theorem 22.1.8. Under this measure, Y no longer has any X-dependent drift, and so the estimation problem can be significantly simplified. The challenge is then to estimate the change of measure term dQ/dP.

Again, we assume all processes are defined on a fixed probability space $(\Omega, \mathcal{F}, P)$ for time $t \in [0,T]$. We suppose there is a right continuous filtration $\{\mathcal{F}_t\}_{t\in[0,T]}$ of sub-σ-algebras of $\mathcal{F}$, and that each $\mathcal{F}_t$ contains all null sets of $\mathcal{F}$.

The signal process X will be, as before, a d-dimensional Markov process, which is the unique strong solution of the SDE (22.1), that is,

$$dX_t = f(t, X_t)dt + \kappa(t, X_t)dB_t + \int_{\mathcal{Z}} g(\zeta, t, X_{t-})\tilde{\mu}(d\zeta, dt),$$

with $\mathcal{F}_0$-measurable initial condition $X_0 \in \mathbb{R}^d$. The infinitesimal generator of X is denoted $\mathcal{L}_t$.

We shall suppose the observation process Y is defined by the m-dimensional system of equations

$$dY_t = c(t, X_t, Y)dt + \alpha(t, Y)dW_t \tag{22.18}$$

with initial condition $Y_0 \in \mathbb{R}^m$. Here

- $\alpha : [0,T] \times \mathcal{C}([0,T];\mathbb{R}^m) \to \mathbb{R}^{m\times m}$ is a nonsingular matrix, Lipschitz in y and with $\|\alpha(t,0)\| + \|\alpha(t,y)^{-1}\|$ uniformly bounded,
- c is bounded, measurable in (t,x) and uniformly Lipschitz in y, and
- W is a standard m-dimensional Brownian motion.

We assume that, in $\{\mathcal{F}_t\}_{t\geq 0}$, the Brownian motions are correlated by $d\langle B^i, W^j\rangle_t = \rho_t^{ij}dt$. As before, we write $\{\mathcal{Y}_t\}_{t\geq 0}$ for the completed filtration generated by Y.

For $\phi \in C_b^2(\mathbb{R}^d)$ (or more generally C_ν^2), we have

$$\phi(X_t) = \phi(X_0) + \int_{[0,t]} \mathcal{L}_u\phi(X_u)du + N_t, \tag{22.19}$$

with N an $\{\mathcal{F}_t\}_{t\in[0,T]}$-adapted P-local martingale, explicitly given in Corollary 22.1.10.

Write $\widetilde{Y}$ for the process given by

$$\widetilde{Y}_t = \int_{0,t]} \alpha^{-1}(u, Y)dY_u.$$

As in the proof of Theorem 22.1.8, as α is Lipschitz in Y, we know $\widetilde{Y}$ and Y generate the same filtration. We define the 'reference' probability measure Q by

$$\frac{dQ}{dP} = \mathcal{E}\Big(- \int_{[0,\cdot]} \big(\alpha(s,Y)^{-1}c(s,X_t,Y)\big)^\top dW_s\Big)_T.$$

By Girsanov's theorem (Theorem 15.2.6), it is easy to check that, under the reference measure Q,

- $\widetilde{B}_t := B_t - \int_{[0,t]} \rho_u^\top \alpha^{-1}(u,Y)c(u,X_s,Y)du$ is a Brownian motion (where ρ_u is the matrix with entries ρ_u^{ij}),
- $\tilde{\mu}$ is a martingale random measure,
- $\widetilde{Y}$ is a Brownian motion, and $d\langle \widetilde{B}^i, \widetilde{Y}^j\rangle_t = \rho_t^{ij}dt$.

For future convenience, we define

$$\Lambda := \mathcal{E}\Big(\int_{[0,\cdot]} \big(\alpha(s,Y)^{-1}c(s,X_s,Y)\big)^\top d\tilde{Y}_s\Big), \tag{22.20}$$

and one can verify that the inverse measure change is given by $dP/dQ = \Lambda_T$ and that Λ is a Q-martingale.

As $\widetilde{Y}$ is a Brownian motion under Q its future increments are independent of the past. The result of Exercise 12.6.8 (see also Hajek and Wong [92]) indicates that, for any $\{\mathcal{F}_t\}_{t\geq 0}$-predictable process H, we have

$$E[(H \bullet \widetilde{Y})_t|\mathcal{Y}_t] = \int_{[0,t]} \hat{H}_s d\widetilde{Y}_s,$$

where $\hat{H}$ is the $\{\mathcal{Y}_t\}_{t\geq 0}$-predictable version of $\{E[H_t|\mathcal{Y}_{t-}]\}_{t\geq 0}$. As Y is continuous, $\hat{H}$ is also the optional version of $\{E[H_t|\mathcal{Y}_t]\}_{t\geq 0}$.

Definition 22.2.1. *For any $\{\mathcal{F}_t\}_{t\geq 0}$-adapted process Z, write $\sigma(Z)$ for the $\{\mathcal{Y}_t\}_{t\geq 0}$-optional process defined by*

$$\sigma_t(Z) := E_Q[\Lambda_t Z_t|\mathcal{Y}_t].$$

The advantage of this notation as follows. For any integrable function ϕ, from Bayes' rule we know that

$$\pi_t(\phi) = E[\phi(X_t)|\mathcal{Y}_t] = \frac{E_Q[\Lambda_t\, \phi(X_t)|\mathcal{Y}_t]}{E_Q[\Lambda_t|\mathcal{Y}_t]} = \frac{\sigma_t(\phi)}{\sigma_t(1)}, \tag{22.21}$$

Therefore, to calculate $\pi_t(\phi)$, it is sufficient for us to calculate $\sigma_t(\phi)$ and $\sigma_t(1)$, which may have a simpler form.

To derive an equation for $\sigma_t(\phi)$, we first note that

$$\sigma_t(\phi) = \sigma_t(1)\cdot \pi_t(\phi).$$

As we already have an expression (22.8) for $\pi_t(\phi)$, we shall derive an equation for $\sigma_t(\phi)$ by obtaining an equation for $\sigma_t(1) = E_0[\Lambda_t|\mathcal{Y}_t]$ and using the differentiation rule for the product.

Theorem 22.2.2. *Write $\widehat{\Lambda}$ for the $\{\mathcal{Y}_t\}_{t\geq 0}$-optional process defined by*

$$\widehat{\Lambda}_t = E_Q[\Lambda_t|\mathcal{Y}_t].$$

(Note the expectation here is with respect to the measure Q.) Then

$$\widehat{\Lambda} = \mathcal{E}\Big(\int_{[0,\cdot]} \big(\alpha^{-1}(u,Y)\pi_u(c)\big)^\top d\widetilde{Y}_u\Big).$$

Here, as before,

$$\pi_u(c) = E[c(u,X_u,Y)|\mathcal{Y}_u],$$

where the expectation is with respect to measure P.

Proof. As defined in (22.20), under Q, Λ is an $\{\mathcal{F}_t\}_{t\geq 0}$-martingale and

$$\Lambda = \mathcal{E}\Big(\int_{[0,\cdot]} \big(\alpha^{-1}(u,Y)c(u,X_u,Y)\big)^\top d\widetilde{Y}_u\Big).$$

Equivalently, Λ satisfies the SDE

$$\Lambda_t = 1 + \int_{[0,t]} \Lambda_u\big(\alpha^{-1}(u,Y)c(u,X_u,Y)\big)^\top d\widetilde{Y}_u. \tag{22.22}$$

Now

$$\widehat{\Lambda}_t = E_Q[\Lambda_t|\mathcal{Y}_t] = E_Q[\Lambda_T|\mathcal{Y}_t] = E_Q[\widehat{\Lambda}_T|\mathcal{Y}_t].$$

Therefore, $\widehat{\Lambda}$ is a $\{\mathcal{Y}_t\}_{t\geq 0}$-martingale under Q. As $\widetilde{Y}$ is a Brownian motion generating the filtration $\{\mathcal{Y}_t\}_{t\geq 0}$, it has the predictable representation property, so there is a $\{\mathcal{Y}_t\}_{t\geq 0}$-predictable m-dimensional process $\{\eta_t\}_{t\geq 0}$, such that, for all $t \in [0,T]$,

$$\widehat{\Lambda}_t = 1 + \int_{[0,t]} \eta_u^\top d\widetilde{Y}_u\,. \tag{22.23}$$

As in Theorem 22.1.9, we shall identify $\{\eta_t\}_{t\geq 0}$ by using the unique decomposition of special semimartingales. Using Itô's differentiation rule, from (22.18) and (22.22) we have

$$\begin{aligned}\Lambda_t Y_t = &\int_{[0,t]} \Lambda_u dY_u + \int_{[0,t]} Y_u \Lambda_u\big(\alpha^{-1}(u,Y)c(u,X_u,Y)\big)^\top d\widetilde{Y}_u \\ &+ \int_{[0,t]} \Lambda_u\big(\alpha^{-1}(u,Y)c(u,X_u,Y)\big)^\top \alpha(u,Y)du\,.\end{aligned} \tag{22.24}$$

Applying the product rule to (22.22), because c and α^{-1} are bounded there exists a constant K such that (after a localization and monotone convergence argument)

$$E_Q[\Lambda_t^2] \leq 2 + 2K\int_{[0,t]} E_Q[\Lambda_u^2]du$$

for all $t \in [0,T]$. Therefore, by Grönwall's inequality (Corollary 15.1.7), Λ is a square integrable $\{\mathcal{F}_t\}_{t\geq 0}$-martingale under Q. Consider the $\{\mathcal{Y}_t\}_{t\geq 0}$-stopping times $\{S_n\}_{n\in\mathbb{N}}$ and $\{T_n\}_{n\in\mathbb{N}}$ of Theorem 22.1.9, and the processes

$$J_t^1 = \int_{[0,t]} \Lambda_u dY_u,$$
$$J_t^2 = \int_{[0,t]} Y_u \Lambda_u \big(\alpha^{-1}(u,Y)c(u,X_u,Y)\big)^\top d\widetilde{Y}_u\,.$$

Then, using Doob's inequality (Theorem 5.1.3), as in Theorem 22.1.9 we have $E[(J^1_{t\wedge S_n})^2] < \infty$ and $E[(J^2_{t\wedge T_n})^2] < \infty$, so J^1 and J^2 are locally square integrable martingales under measure Q with respect to the filtration $\{\mathcal{F}_t\}_{t\geq 0}$. Because the stopping times are $\{\mathcal{Y}_t\}_{t\geq 0}$-measurable, the optional processes defined by

$$\widehat{J}_t^1 = E_Q[J_t^1|\mathcal{Y}_t]$$
$$\widehat{J}_t^2 = E_Q[J_t^2|\mathcal{Y}_t]$$

are locally square integrable martingales under Q with respect to the filtration $\{\mathcal{Y}_t\}_{t\geq 0}$. Write

$$K_t = \int_{[0,t]} \Lambda_u \big(\alpha^{-1}(u,Y)c(u,X_u,Y)\big)^\top \alpha(u,Y)du$$

and $\widehat{K}$ for the $\{\mathcal{Y}_t\}_{t\geq 0}$-optional process

$$\widehat{K}_t = E_Q[K_t|\mathcal{Y}_t].$$

Then, if $R_t = E_Q[\Lambda_t c(t,X_t,Y)|\mathcal{Y}_t]$, we know

$$\widetilde{K}_t = \widehat{K}_t - \int_{[0,t]} \big(\alpha^{-1}(u,Y)\,R_u\big)^\top \alpha(u,Y)du$$

is a Q-local martingale with respect to the filtration $\{\mathcal{Y}_t\}_{t\geq 0}$. From (22.24), it follows that

$$E_Q[\Lambda_t Y_t|\mathcal{Y}_t] = \widehat{\Lambda}_t Y_t = \widehat{J}_t^1 + \widehat{J}_t^2 + \widetilde{K}_t + \int_{[0,t]} \big(\alpha^{-1}(u,Y)R_u\big)^\top \alpha(u,Y)du. \tag{22.25}$$

This represents $\widehat{\Lambda}_t Y_t$ as the sum of a Q-local martingale and a continuous (and so predictable) process of finite variation. Consequently, $\widehat{\Lambda}_t Y_t$ is a Q-special semimartingale and this representation is unique. However, from (22.23) and Definition 22.1.6,

$$\widehat{\Lambda}_t Y_t = \int_{[0,t]} \widehat{\Lambda}_u dY_u + \int_{[0,t]} Y_u \eta_u^\top d\widetilde{Y}_u + \int_{[0,t]} \eta_u^\top \alpha(y_u)du.$$

Again, the first two integrals are $\{\mathcal{Y}_t\}_{t\geq 0}$-local martingales under measure Q. By the uniqueness of the decomposition of special semimartingales,

$$\eta_u = \alpha^{-1}(u,Y)R_u = \alpha^{-1}(u,Y)E_Q[\Lambda_u c(u,X_u,Y)|\mathcal{Y}_u].$$

However, from (22.21), this simplifies to

$$\eta_u = \alpha^{-1}(u,Y)\widehat{\Lambda}_u \pi_u(c).$$

Substituting in (22.23) gives the result. □

In what follows, recall that κ refers to the volatility of the process X, while σ denotes the unnormalized $\{\mathcal{Y}_t\}_{t\geq 0}$-optional projection.

Theorem 22.2.3 (Zakai Equation). *For any $\phi \in C_b^2(\mathbb{R}^d)$, the projection $\sigma_t(\phi)$ satisfies the equation*

$$\begin{aligned}\sigma_t(\phi) = \sigma_0(\phi) &+ \int_{[0,t]} \sigma_u(\mathcal{L}_u\phi)du \\ &+ \int_{[0,t]} \big(\sigma_u(\nabla\phi\cdot\kappa\cdot\rho) + \alpha^{-1}(u,Y)\sigma_u(\phi c)\big)^{\top} d\widetilde{Y}_t\,,\end{aligned} \tag{22.26}$$

where $\sigma_u(\phi c) = E_Q[\Lambda_u \phi(X_u)c(u,X_u,Y)|\mathcal{Y}_u]$.

Proof. We have seen that

$$\sigma_t(\phi) = \widehat{\Lambda}_t \pi_t(\phi),$$

so, from (22.8), (22.21), the dynamics of $\widehat{\Lambda}$ and the fact $dV_t = d\widetilde{Y}_t - \alpha(t,Y)^{-1}\pi(c)dt$,

$$\begin{aligned}\widehat{\Lambda}_t\pi_t(\phi) &= \sigma_0(\phi) + \int_{[0,t]} \widehat{\Lambda}_u \pi_u(\mathcal{L}_u\phi)du \\ &\quad + \int_{[0,t]} \widehat{\Lambda}_u\big\{\pi_u(\nabla\phi\cdot\kappa\cdot\rho) + \alpha^{-1}(u,Y)\big(\pi_u(\phi c) - \pi_u(\phi)\pi_u(c)\big)\big\}dV_u \\ &\quad + \int_{[0,t]} \pi_u(\phi)\widehat{\Lambda}_u\big(\alpha^{-1}(u,Y)\pi_u(c)\big)^{\top} d\widetilde{Y}_u \\ &\quad + \int_{[0,t]} \widehat{\Lambda}_u\Big(\pi_u(\nabla\phi\cdot\kappa\cdot\rho) \\ &\qquad + \alpha^{-1}(u,Y)\big(\pi_u(\phi c) - \pi_u(\phi)\pi_u(c)\big)\Big)^{\top}\big(\alpha^{-1}(u,Y)\pi_u(c)\big)du \\ &= \sigma_0(\phi) + \int_{[0,t]} \sigma_u(\mathcal{L}_u\phi)du \\ &\quad + \int_{[0,t]} \big(\sigma_u(\nabla\phi\cdot\kappa\cdot\rho) + \alpha^{-1}(u,Y)\sigma_u(\phi c)\big)^{\top} d\tilde{Y}_u\,.\end{aligned}$$

□

Remark 22.2.4. Note the much simpler form of the equation (22.26) for $\sigma_t(\phi)$ compared with (22.8) for $\pi_t(\phi)$: (22.26) is linear in σ_t, whereas (22.8) is quadratic in π_t. In particular, when the signal noise B is independent of the observation noise W in the observation, so that the predictable quadratic covariation matrix $\rho = (\rho^{ij}) = (\langle B^i, W^i \rangle)$ is zero, the unnormalized density $\sigma_t(\phi)$ satisfies the equation

$$\sigma_t(\phi) = \sigma_0(\phi) + \int_{[0,t]} \sigma_u(\mathcal{L}_u \phi) du + \int_{[0,t]} \big(\alpha^{-1}(u, Y)\sigma_u(\phi c)\big)^\top \alpha^{-1}(u, Y) dY_u \,. \tag{22.27}$$

Example 22.2.5. Suppose the homogeneous Markov process X is such that, as in Corollary 22.1.14, the conditional distribution of X_t given $\mathcal{Y}_t$ has a smooth density $\widehat{p}(t, x)$. Then we can define the *unnormalized conditional density* as

$$q(t, x) = \widehat{\Lambda}_t \widehat{p}(t, x),$$

so that

$$\widehat{p}(t, x) = \frac{q(t, x)}{\int_{\mathbb{R}^d} q(t, x')dx'} \,.$$

Similarly to Corollary 22.1.14, substituting this equation in (22.27) and integrating by parts we obtain the following stochastic partial differential equation for q,

$$dq(t, x) = \mathcal{L}_t^* q(t, x)dt + q(t, x)\big(\alpha(t, Y)^{-1} c(t, x, Y)\big)^\top \alpha(t, Y)^{-1} dY_t \,, \tag{22.28}$$

where again $\mathcal{L}^*$ is the adjoint of the infinitesimal generator $\mathcal{L}$. Equation (22.28) is much simpler than (22.9) obtained for $\widehat{p}$. It is linear in q, it does not involve terms such as $\pi_t(c)$ and it has the observation process Y as input.

22.3 The Wonham Filter for Markov Chains

In applications, a key difficulty in working with equations such as (22.28) is that they are infinite dimensional. We are, therefore, particularly interested in considering those cases where it is possible to reduce this to a 'finite-dimensional filter', that is, a filter in which sufficient statistics of the signal process are given by a finite-dimensional system of equations.

We shall see that one setting in which this is possible is when the underlying process X is a continuous-time finite-state Markov chain. This is unsurprising, as in this setting the space of distributions over the states of X form a finite dimensional space. Alternatively, one can see this from the fact that the state process is, in effect, an indicator function and indicator functions are idempotent. Consequently, the square of the state process can be expressed in terms of the process itself and no higher order terms arise.

Exploiting this idea, in this section we determine some finite-dimensional filters related to the Wonham filter, which will be given in (22.36). In particular, we obtain finite-dimensional filters and smoothers for the following processes:

(i) the state of the Markov chain,
(ii) the number of jumps N_t^{ij} of the chain from state i to state j,
(iii) the occupancy time J_t^i of the Markov chain in state i,
(iv) a stochastic integral G_t^i related to the observation process.

The filtered estimate of the state is the Wonham filter [13]. The smoothed estimate of the state is given in Clements and Anderson [9]. A finite-dimensional filter for the number of jumps N_t^{ij} was obtained by Zeitouni and Dembo [10, 26], and used to estimate the parameters of the Markov chain and the observation process. However, this estimation also involves J_t^i and G_t^i for which finite-dimensional filters are not given in [26].

Given these quantities, we shall also outline the application of the EM (Expectation-Maximization) algorithm of Dempster, Laird and Rubin [57] for estimation of the parameters of the model. In this setting, the EM algorithm can be seen as an extension of the discrete-time Baum–Welch algorithm (see [10, 26]). Unlike the Baum–Welch method, our equations are recursive and can be implemented by the usual methods of discretization; no forward-backward estimates are required.

We begin by describing the formal dynamics of a general finite-state Markov chain X. For further details on this approach, see [71] (where question related to filtering are explored in more detail and generality), or the introductory text of Norris [139].

22.3.1 Markov Chain Dynamics

For any finite set $\Sigma = \{s_1, s_2, \ldots, s_N\}$ consider the functions $\{\phi_i\}_{1\le i\le N}$ defined by $\phi_i(s_j) := \delta_{ij}$, and the corresponding vector function $\phi(s) := \big(\phi_1(s), \phi_2(s), \ldots, \phi_N(s)\big)^\top$. Then ϕ is a bijection of Σ and the set $\mathbb{S} = \{e_i, e_2, \ldots, e_N\}$ of unit vectors $e_j = (0, 0, \ldots, 1, \ldots, 0)^\top$ of $\mathbb{R}^N$. Using such a bijection, the state space of a finite-state space Markov chain can, without loss of generality, be taken to be the set $\mathbb{S}$.

Suppose, therefore, that $\{X_t\}_{t\ge 0}$ is a time homogeneous Markov chain defined on a probability space $(\Omega, \mathcal{F}, P)$ with state space $\mathbb{S}$. Associated with X is the Q-matrix $A^\top$, which is the infinitesimal generator of X. For $1 \le i \le N$, writing $p_t^i = P(X_t = e_i)$, we know that the probability distribution $p_t = (p_t^1, p_t^2, \ldots, p_t^N)^\top$ satisfies the forward equation

$$\frac{dp_t}{dt} = Ap_t\,. \tag{22.29}$$

As $A^\top$ is a Q-matrix, we know

$$\sum_{i=1}^N a_{ij} = 0 \text{ and } a_{ij} \geq 0 \text{ for all } i \neq j. \tag{22.30}$$

The process X is not observed directly; rather we suppose there is a (scalar) observation process Y given by

$$Y_t = \int_{[0,t]} c(X_s)ds + W_t. \tag{22.31}$$

(The extension to vector processes Y is straightforward, as is the inclusion of Y-dependent volatility and drift.) Here, W is a standard Brownian motion on $(\Omega, \mathcal{F}, P)$, which is independent of X. Because X takes values in $\mathbb{S}$, the function c is given by a vector $c = (c_1, c_2, \dots, c_N)^\top$, so that $c(X_t) = X_t^\top c$.

As in the previous section, let $\{\mathcal{F}_t\}_{t\geq 0}$ be the completed, right-continuous filtration generated by (X, Y) and $\{\mathcal{Y}_t\}_{t\geq 0}$ that generated by Y alone. Note $\mathcal{Y}_t \subset \mathcal{F}_t$ for all t.

Write $\Phi(t,s) = \exp(A(t-s))$ for the transition matrix associated with A, (where exp here denotes the matrix exponential) so that

$$\frac{d}{dt}\Phi(t,s) = A\Phi(t,s) \tag{22.32}$$

and, for $s \leq t$,

$$E[X_t|\mathcal{F}_s] = E[X_t|X_s] = \Phi(t,s)X_s\,.$$

Lemma 22.3.1. *The process M defined by*

$$M_t := X_t - X_0 - \int_{[0,t]} AX_s ds$$

is a (vector) $\{\mathcal{F}_t\}_{t\geq 0}$-martingale under P.

Proof. As X and A are bounded, the integrability of M is trivial. For $s \leq t$, using (22.32) we see

$$\begin{aligned} E[M_t - M_s|\mathcal{F}_s] &= E\Big[X_t - X_s - \int_{]s,t]} AX_u du | X_s\Big] \\ &= \Phi(t,s)X_s - X_s - \int_{]s,t]} A\Phi(u,s)X_s du = 0. \end{aligned}$$

□

Remark 22.3.2. The semimartingale representation of the Markov chain X is, therefore,

$$X_t = X_0 + \int_{]0,t]} AX_s ds + M_t. \tag{22.33}$$

Note $\int_{[0,t]} AX_s ds = \int_{[0,t]} AX_{s-} ds$, because $X_s(\omega) = X_{s-}(\omega)$ except for countably many s, almost surely. We shall make this and similar identifications for the sake of notational simplicity.

We shall consider the Zakai equation. For this, we introduce the probability measure Q by putting

$$\frac{dQ}{dP} = \mathcal{E}\big(- (X^\top c)\bullet W\big)_T.$$

It is also convenient to define the process

$$\Lambda = \mathcal{E}\big((X^\top c)\bullet Y\big), \tag{22.34}$$

so that $dP/dQ = \Lambda_T$. Note that Λ is an $\{\mathcal{F}\}_{t\geq 0}$-martingale under Q.

As in the previous section, given these definitions we proceeded to determine the behaviour of $E_Q[\Lambda_t|\mathcal{Y}_t]$. In this context, an alternative approach follows from the following lemma.

Lemma 22.3.3. *Under Q, the following hold.*

(i) X is a Markov chain with the same dynamics as under P and Y is a Brownian motion,
(ii) X and Y are independent processes,
(iii) for any $\{\mathcal{F}_t\}_{t\geq 0}$-predictable process $H \in L^2(M)$ of appropriate dimension, we know

$$E_Q[(H\bullet M)_t|\mathcal{Y}_t] = 0 \quad \text{a.s.},$$

(iv) for any $\mathcal{F}_s$-measurable random variable K and any $t \geq s$,

$$E_Q[K|\mathcal{Y}_t] = E_Q[K|\mathcal{Y}_s] \quad \text{a.s.},$$

(v) for any $\{\mathcal{F}_t\}_{t\geq 0}$-predictable process H admitting a $\{\mathcal{Y}_t\}_{t\geq 0}$-predictable projection $\tilde{H}$ under Q,

$$E_Q[(H\bullet Y)_t|\mathcal{Y}_t] = (\tilde{H}\bullet Y)_t \quad \text{a.s.}$$

In (v) we also know $\tilde{H}_t = E_Q[H_t|\mathcal{Y}_{t-}] = E_Q[H_t|\mathcal{Y}_t]$ almost surely, for each t.

Proof. From Girsanov's theorem, it is easy to verify that Y and M are both Q-martingales. As $dX_s = AX_{s-}ds + dM_s$, and X takes values in $\mathbb{S}$, it follows that X is a Markov chain with the same dynamics as under P. As $\langle Y\rangle_t = t$, Lévy's characterization guarantees that Y is a Brownian motion. Hence (i) holds.

To see that X and Y are independent, we first notice that, from the structure of the measure change, the pair (X, Y) is a Markov process under Q. The generator of (X, Y) is the same as the generator of (X', Y'), where X' is a Markov chain with the same dynamics as X and Y' is an independent Brownian motion. As independence depends only on the joint law of the processes, and from the Kolmogorov forward equation the law of a Markov process depends only on the generator, it follows that X and Y are independent, so (ii) holds.

As we know Y is independent of X, it follows that M is independent of $\mathcal{Y}_t$ for all t. For any $\{\mathcal{F}_t\}_{t\geq 0}$-stopping times $S \leq T$ and any t, we deduce that $E_Q[M_T - M_S|\mathcal{Y}_t] = E_Q[M_T - M_S] = 0$. Consequently, for any simple $\{\mathcal{F}_t\}_{t\geq 0}$-predictable function H, the result holds. Statement (iii) then follows by approximation and the Itô isometry.

To show (iv), observe that, as Y is a Brownian motion in $\{\mathcal{F}_t\}_{t\geq 0}$, the increments $Y_t - Y_s$ are independent of $\mathcal{F}_s$. As K is $\mathcal{F}_s$ measurable, it follows that K is independent of $Y_t - Y_s$. Finally, we know that, up to null sets, $\mathcal{Y}_t = \mathcal{Y}_s \vee \sigma(\{Y_u - Y_s\}_{u\in[s,t]})$, and (iv) follows.

Finally, we recall that (v) is the result of Exercise 12.6.8 (see also Hajek and Wong [92]). As Y is continuous, that $\tilde{H}_t = E[H_t|\mathcal{Y}_{t-}] = E[H_t|\mathcal{Y}_t]$ follows immediately. □

Definition 22.3.4. *If ϕ is an integrable, measurable process,*

- *for $t \in [0, T]$, we write $\widehat{\phi}$ for the $\{\mathcal{Y}_t\}_{t\geq 0}$-optional projection $\widehat{\phi}_t = E[\phi_t|\mathcal{Y}_t]$, and call $\widehat{\phi}_t$ the 'filtered' estimate of ϕ_t,*
- *for $0 \leq s \leq t \leq T$, we write $\pi_t(\phi_s)$ for $\{\mathcal{Y}_t\}_{t\geq 0}$-optional projection of the constant process ϕ_s, so $\pi_t(\phi_s) = E[\phi_s|\mathcal{Y}_t]$, and call $\pi_t(\phi_s)$ the 'smoothed' estimate of ϕ_s given $\mathcal{Y}_t$,*

where all projections are under the measure P. Of course, $\pi_t(\phi_t) = \widehat{\phi}_t$.

If ϕ is an $\{\mathcal{F}_t\}_{t\geq 0}$-adapted, integrable process, we shall write $\sigma_t(\phi)$ for the $\{\mathcal{Y}_t\}_{t\geq 0}$-optional projection of $(\Lambda\phi)$ under Q, so that

$$\sigma_t(\phi) = E_Q[\Lambda_t\phi_t|\mathcal{Y}_t] \qquad \text{a.s.}$$

Bayes' theorem states that

$$\pi_t(\phi) = \widehat{\phi}_t = E[\phi_t|\mathcal{Y}_t] = \frac{E_Q[\Lambda_t\phi_t|\mathcal{Y}_t]}{E_Q[\Lambda_t|\mathcal{Y}_t]} = \frac{\sigma_t(\phi)}{\sigma_t(1)}, \tag{22.35}$$

where E_Q denotes expectation with respect to Q. Consequently, $\sigma_t(1) = E_Q[\Lambda_t|\mathcal{Y}_t] =: \widehat{\Lambda}_t$, defines the $\{\mathcal{Y}_t\}_{t\geq 0}$-optional projection of Λ_t under Q.

By applying (22.26), we can now proceed to write down the Zakai equation in this context. The Markov chain X satisfies an SDE of the form considered in Remark 22.1.13, with no volatility term, and where the random measure can only jump between basis vectors. We write $\phi(X) = X^\top\phi$, for ϕ a vector in $\mathbb{R}^N$, and hence $\mathcal{L}_u\phi = A\phi$. Simplifying (22.26) (by removing terms related to the diffusion B and using the fact $XX^\top c = \mathrm{diag}(c)X$, as X is a basis vector) we then obtain

$$\begin{aligned}\phi^\top\sigma_t(X) = \phi^\top\sigma_0(X) &+ \int_{[0,t]} \phi^\top A\sigma_u(X)du \\ &+ \int_{[0,t]} \phi^\top \mathrm{diag}(c)\sigma_u(X)\, dY_t\,,\end{aligned}$$

Writing $q_t = \sigma_t(X)$ for the unnormalized density, we immediately obtain the recursive finite-dimensional filtering equation (the unnormalized Zakai form of the Wonham filter)

$$q_t = q_0 + \int_{[0,t]} Aq_u du + \int_{[0,t]} \mathrm{diag}(c) q_u \, dY_u. \tag{22.36}$$

Comparing this with the forward equation for the unconditional density (22.29), we see that the only difference is the addition of the stochastic term $\mathrm{diag}(c)q_u \, dY_u$, which represents the information gained from observing Y. By comparing with the innovations approach, we can also obtain the normalized equation,

$$\widehat{p}_t = \widehat{p}_0 + \int_{[0,t]} A\widehat{p}_u du + \int_{[0,t]} (\mathrm{diag}(c) - (c^\top \widehat{p}_u)I)\widehat{p}_u \, dV_t,$$

where $dV_t = dY_t - (p_t^\top c)dt$ is the innovations process.

22.3.2 A General Finite-Dimensional Filter

For the sake of estimating parameters, it is convenient to also derive recursive formulae for more general quantities, using Lemma 22.3.3.

Consider a scalar process H of the form

$$H_t = H_0 + \int_{]0,t]} \alpha_s ds + \int_{]0,t]} \beta_s^\top dM_s + \int_{]0,t]} \delta_s dW_s, \tag{22.37}$$

where α, β and δ are $\{\mathcal{F}_t\}_{t\geq 0}$-predictable, square-integrable processes of appropriate dimensions. That is, α and δ are real and β is an N-dimensional real vector.

Using the product rule for semimartingales, as M is of finite variation and W is continuous (hence M and W are orthogonal),

$$\begin{aligned} H_t X_t = H_0 X_0 &+ \int_{]0,t]} \alpha_s X_s ds + \int_{]0,t]} \beta_s^\top X_{s-} dM_s \\ &+ \int_{]0,t]} \delta_r X_{s-} dW_s + \int_{[0,t]} H_s A X_s ds + \int_{]0,t]} H_{s-} dM_s \\ &+ \sum_{0<s\leq t} (\beta_s^\top \Delta X_s)\Delta X_s. \end{aligned} \tag{22.38}$$

From the structure of X, the final term can be written

$$\begin{aligned} \sum_{0<s\leq t} (\beta_s^\top \Delta X_s)\Delta X_s = &\sum_{i,j=1}^N \int_{]0,t]} (\beta_s^j - \beta_s^i)\langle X_{s-}, e_i\rangle \, \langle e_j, dM_s\rangle (e_j - e_i) \\ &+ \sum_{i,j=1}^N \int_{]0,t]} \langle \beta_s^j X_s - \beta_s^i X_s, e_i\rangle \, a_{ji} \, ds \, (e_j - e_i), \end{aligned}$$

Here $\langle X, e_i \rangle = X^\top e_i$ is the Euclidean inner product (not to be confused with the predictable quadratic variation). Substituting in (22.38), we have

$$\begin{aligned} H_t X_t = H_0 X_0 &+ \int_{]0,t]} \alpha_s X_s ds + \int_{[0,t]} \beta_s X_{s-} dM_s \\ &+ \int_{]0,t]} \delta_s X_{s-} dW_s + \int_{[0,t]} H_s A X_s ds + \int_{]0,t]} H_{s-} dM_s \\ &+ \sum_{i,j=1}^{N} \int_{]0,t]} (\beta_s^j - \beta_s^i) \langle X_{s-}, e_i \rangle \langle e_j, dM_s \rangle (e_j - e_i) \\ &+ \sum_{i,j=1}^{N} \int_{]0,t]} \langle \beta_s^j X_s - \beta_s^i X_s, e_i \rangle a_{ji} ds (e_j - e_i). \end{aligned}$$

For simplicity, we shall write

$$C = \mathrm{diag}(c).$$

Theorem 22.3.5. *If H satisfies (22.37), then the recursive equation for the unnormalized estimate $\sigma_t(HX)$ is given by the following linear equation:*

$$\begin{aligned} \sigma_t(HX) = \sigma_0(HX) &+ \int_{]0,t]} \sigma_s(\alpha X) ds + \int_{]0,t]} A \sigma_s(HX) ds \\ &+ \int_{]0,t]} \sum_{i,j=1}^{N} \langle \sigma_s(\beta^j X - \beta^i X), e_i \rangle a_{ji} \, ds (e_j - e_i) \\ &+ \int_{]0,t]} \big(\sigma_s(\delta X) + C \sigma_s(HX)\big) dY_s \,. \end{aligned} \tag{22.39}$$

Proof. Using the product rule, we have

$$\begin{aligned} \Lambda_t H_t X_t = H_0 X_0 &+ \int_{]0,t]} \Lambda_s \alpha_s X_s ds + \int_{]0,t]} \delta_s \Lambda_{s-} X_{s-} dW_s + \int_{]0,t]} \Lambda_s H_s A X_s ds \\ &+ \sum_{i,j=1}^{N} \int_{]0,t]} \langle \beta_s^j \Lambda_s X_s - \beta_s^i \Lambda_s X_s, e_i \rangle a_{ji} ds \, (e_j - e_i) \\ &+ \sum_{i=1}^{N} \int_{]0,t]} \langle \Lambda_s X_s, e_i \rangle c_i H_s dY_s e_i + \sum_{i=1}^{N} \int_{]0,t]} \langle \Lambda_s X_s, e_i \rangle \delta_s ds \, c_i e_i \\ &+ \tilde{M}_t \end{aligned}$$

where $\tilde{M}$ is a vector martingale which can be written as a linear combination of stochastic integrals with respect to M. Under Q, Y is a standard Brownian motion and

$$dY_t = dW_t + X_t^\top c \, dt = dW_t + \sum_{i=1}^{N} \langle X_t, e_i \rangle dt \, c_i e_i.$$

We also have the simplification $\sum_{i=1}^{N}\langle X_s, e_i\rangle c_i e_i = CX_s$. Conditioning each side of our expansion on $\mathcal{Y}_t$ under $Q = \Lambda^{-1}\cdot P$, we have

$$\begin{aligned}\sigma_t(HX) = E[H_0X_0|\mathcal{Y}_t] &+ \int_{]0,t]} E_Q[\Lambda_s\alpha_sX_s|\mathcal{Y}_t]ds \\ &+ E_Q\Big[\int_{]0,t]}\big(\delta_s\Lambda_sX_s + C\Lambda_sH_sX_s\big)dY_s\Big|\mathcal{Y}_t\Big] \\ &+ \int_{]0,t]} A\,E_Q[\Lambda_sH_sX_s|\mathcal{Y}_t]ds \\ &+ \sum_{i,j=1}^{N}\int_{]0,t]}\langle E_Q[\Lambda_s(\beta_s^j-\beta_s^i)X_s|\mathcal{Y}_t], e_i\rangle a_{ji}ds\,(e_j-e_i) \\ &+ E_Q[\tilde{M}_t|\mathcal{Y}_t].\end{aligned}$$

and using the results of Lemma 22.3.3, the result follows. □

The Zakai equation is recursive, so for $s \le t$ we have the following form.

Corollary 22.3.6.

$$\begin{aligned}\sigma_t(HX) = \sigma_s(HX) &+ \int_{]s,t]}\sigma_u(\alpha X)du + \int_{]s,t]}A\sigma_u(HX)du \\ &+ \sum_{i,j=1}^{N}\int_{]s,t]}\big\langle\sigma_u\big((\beta^j-\beta^i)X\big), e_i\big\rangle\, a_{ji}du(e_j-e_i) \qquad (22.40) \\ &+ \int_{]s,t]}\big(\sigma_u(\delta X) + C\sigma_u(HX)\big)dY_u\,.\end{aligned}$$

Here, the initial condition is $E_Q[\Lambda_sH_sX_s|\mathcal{Y}_s]$, again a $\mathcal{Y}_s$-measurable random variable.

22.3.3 States, Transitions and Occupation Times

We now obtain particular finite-dimensional filters and smoothers for various quantities, in their unnormalized (Zakai) form, by specializing the result of Section 22.3.2.

The State

Take $H_0 := 1$, $\alpha_s := 0$, $\beta_s := 0 \in \mathbb{R}^N$, and $\delta_s := 0$, so $H_t \equiv 1$. Applying Theorem 22.3.5 we obtain a single, finite-dimensional equation for the unnormalized conditional distribution $\sigma_t(X) = E_Q[\Lambda_tX_t|\mathcal{Y}_t]$, namely

$$\sigma_t(X) = \widehat{X}_0 + \int_{]0,t]} A\sigma_s(X)ds + \int_{]0,t]} C\sigma_s(X)dY_s. \tag{22.41}$$

This is the same as (22.36), which we obtained using the results of Section 22.2. For the 'smoothed' estimates of X_s given $\mathcal{Y}_t$, for s,t fixed and $r \in]s,t]$, take

$$H_s = \langle X_s, e_i \rangle, \quad \alpha_r = 0, \quad \beta_r = 0 \quad \text{and} \quad \delta_r = 0,$$

so that $H_r \equiv \langle X_s, e_i \rangle$. Substituting in (22.40) gives[1]

$$\begin{aligned} \sigma_t(\langle X_s, e_i \rangle X) = \sigma_s(\langle X_s, e_i \rangle X_s) &+ \int_{]s,t]} A\sigma_u(\langle X_s, e_i \rangle X)du \\ &+ \int_{]s,t]} C\sigma_u(\langle X_s, e_i \rangle X)dY_u\,. \end{aligned} \tag{22.42}$$

This is a single-equation finite-dimensional filter for

$$\sigma_t(\langle X_s, e_i \rangle X) = E_Q[\langle X_s, e_i \rangle X_t | \mathcal{Y}_t] = E\left[\Lambda_t \langle X_s, e_i \rangle | \mathcal{Y}_t\right].$$

Taking the inner product with $\mathbf{1} = (1, 1, \ldots, 1)^\top \in \mathbb{R}^N$ gives $\sigma_t(\langle X_s, e_i \rangle)$.

The Number of Jumps

For $i \neq j$, the number of jumps from e_i to e_j in the period $]0,t]$ is

$$\begin{aligned} \sum_{0<s\leq t} \langle X_{s-}, e_i \rangle \langle X_s, e_j \rangle &= \sum_{0<s\leq t} \langle X_{s-}, e_i \rangle \langle X_s - X_{s-}, e_j \rangle \\ &= \sum_{0<s\leq t} \langle X_{s-}, e_i \rangle \langle \Delta z_s, e_j \rangle \\ &= \int_{]0,t]} \langle X_{s-}, e_i \rangle \langle dX_s, e_j \rangle \\ &= \int_{]0,t]} \langle X_{s-}, e_i \rangle \langle AX_s ds + dM_s, e_j \rangle \\ &= \int_{]0,t]} \langle X_{s-}, e_i \rangle\, a_{ji} ds + V_t^{ij}, \end{aligned}$$

where

$$V_t^{ij} = \int_{]0,t]} \langle X_{s-}, e_i \rangle \langle dM_s, e_j \rangle,$$

so the number of jumps from e_i to e_j in the time interval $]0,t]$ is given by

[1] Rather than taking $H_r = \langle X_s, e_i \rangle$ and estimating $P(X_s = e_i | \mathcal{Y}_t) = \pi_t(\langle X_s, e_i \rangle) = \sigma_t(\langle X_s, e_i \rangle)/\sigma_t(1)$ we could consider all states of X_s simultaneously by taking $H_r = X_s$, for all $s \leq r \leq t$. However, the product $H_t X_t$ would then have to be interpreted as a tensor, or Kronecker, product $H_t X_t^\top$.

$$\mathcal{J}_t^{ij} = \int_{]0,t]} \langle X_{r^-}, e_i \rangle \, a_{ji} dr + V_t^{ij} \,. \tag{22.43}$$

To obtain the unnormalized filter equation, take

$$H_0 = 0, \quad \alpha_r = \langle X_r, e_i \rangle \, a_{ji}, \quad \beta_r = \langle X_r, e_i \rangle \, e_j \quad \text{and} \quad \delta_r = 0,$$

so that $H_t = \mathcal{J}_t^{ij}$. The Zakai equation for $\sigma_t(\mathcal{J}^{ij}X)$ is then obtained by substituting in Theorem 22.3.5,

$$\begin{aligned} \sigma_t(\mathcal{J}^{ij}X) = &\int_{]0,t]} \langle \sigma_u(X), e_i \rangle \, a_{ji} e_j du + \int_{]0,t]} A\sigma_u(\mathcal{J}^{ij}X) du \\ &+ \int_{]0,t]} C\sigma_u(\mathcal{J}^{ij}X) dY_u \,. \end{aligned} \tag{22.44}$$

The smoothed estimate of $\mathcal{J}_s^{ij}$ given $\mathcal{Y}_t$, for $s \le t$, is obtained from Corollary 22.3.6 by taking, for all $r \in \,]s,t]$,

$$H_s = \mathcal{J}_s^{ij}, \quad \alpha_r = 0, \quad \beta_r = 0 \quad \text{and} \quad \delta_r = 0.$$

so $H_r = \mathcal{J}_s^{ij}$. Then, from (22.40) we have the finite-dimensional Zakai form of the smoother

$$\sigma_t(\mathcal{J}_s^{ij}X) = \sigma_s(\mathcal{J}_s^{ij}X) + \int_{]s,t]} A\sigma_u(\mathcal{J}_s^{ij}X) du + \int_{]s,t]} C\sigma_u(\mathcal{J}_s^{ij}X) dY_u. \tag{22.45}$$

Occupation Times

The time spent by the process in state e_i is given by

$$\mathcal{O}_t^i = \int_{[0,t]} \langle X_r, e_i \rangle dr, \quad 1 \le i \le N.$$

Take

$$H_0 = 0, \quad \alpha_r = \langle X_r, e_i \rangle, \quad \beta_r = 0 \in \mathbb{R}^N, \quad \text{and} \quad \delta_r = 0,$$

so $H_t = \mathcal{O}_t^i$. Substituting again in Theorem 22.3.5 gives the Zakai equation

$$\begin{aligned} \sigma_t(\mathcal{O}^iX) = &\int_{]0,t]} \langle \sigma_u(X), e_i \rangle e_i du + \int_{]0,t]} A\sigma_u(\mathcal{O}^iX) du \\ &+ \int_{]0,t]} C\sigma_u(\mathcal{O}^iX) dY_u \,. \end{aligned} \tag{22.46}$$

The finite-dimensional unnormalized smoother is obtained for $\mathcal{O}_s^i$ by taking, for $r \in \,]s,t]$,

$$H_s = \mathcal{O}_s^i \quad \alpha_r = 0, \quad \beta_r = 0 \quad \text{and} \quad \delta_r = 0,$$

so $H_r = \mathcal{O}_s^i$. Applying Corollary 22.3.6 gives

$$\sigma_t(\mathcal{O}_s^iX) = \sigma_s(\mathcal{O}_s^iX) + \int_{]s,t]} A\sigma_u(\mathcal{O}_s^iX) du + \int_{]s,t]} C\sigma_u(\mathcal{O}_s^iX) dY_u. \tag{22.47}$$

The Drift Coefficient

Below, we shall see that the estimation of the drift coefficient $c = (c_1, c_2, \dots, c_N)^\top$ of the observation process involves the filtered estimate of the processes

$$\mathcal{T}_t^i = \int_{]0,t]} \langle X_u, e_i \rangle dY_u = \int_{]0,t]} c_i \langle X_u, e_i \rangle du + \int_{]0,t]} \langle X_u, e_i \rangle dW_u.$$

This process gives the total change in the observation Y over periods when X is in state e_i. Taking

$$H_0 = 0, \quad \alpha_r = c_i \langle X_r, e_i \rangle, \quad \beta_r = 0 \quad \text{and} \quad \delta_r = \langle X_r, e_i \rangle,$$

so $H_t = \mathcal{T}_t^i$, we shall apply Theorem 22.3.5, again noting that

$$X_r \alpha_r = c_i \langle X_r, e_i \rangle X_r = c_i \langle X_r, e_i \rangle e_i$$

and

$$X_r \delta_r = X_r \langle X_r, e_i \rangle = \langle X_r, e_i \rangle e_i.$$

The Zakai equation here is

$$\begin{aligned} \sigma_t(\mathcal{T}^i X) = c_i \int_{]0,t]} \langle \sigma_u(X), e_i \rangle e_i du + \int_{]0,t]} A \sigma_u(\mathcal{T}^i X) du \\ + \int_{]0,t]} \big(C \sigma_u(\mathcal{T}^i X) + \langle \sigma_u(X), e_i \rangle e_i \big) dY_u. \end{aligned}$$

For the smoother, taking, for $r \in]s,t]$,

$$H_s = \mathcal{T}_s^i, \quad \alpha_r = 0, \quad \beta_r = 0 \quad \text{and} \quad \delta_r = 0,$$

so $H_r = \mathcal{T}_s^i$, we obtain from Corollary 22.3.6

$$\sigma_t(\mathcal{T}_s^i X) = \sigma_s(\mathcal{T}_s^i X) + \int_{]s,t]} A \sigma_u(\mathcal{T}_s^i X) du + \int_{]s,t]} C \sigma_u(\mathcal{T}_s^i X) dY_u. \quad (22.48)$$

Remark 22.3.7. In all the above smoothing equations, when we take an inner product with $\mathbf{1}$ the integral involving A will vanish, because $\sum_{j=1}^N a_{ji} = 0$.

22.3.4 Parameter Estimation

This problem is nicely discussed in Dembo and Zeitouni [56] and Zeitouni and Dembo [191]. We review their formulation in our setting.

Suppose, as above, that X is a Markov chain with representation (22.29). Again, suppose X is observed indirectly through the process Y, which satisfies (22.30).

The above model, therefore, is determined by the set of parameters

$$\theta := \big(\{a_{ij}\}_{1\le i,j\le N},\ \{c_i\}_{1\le i\le N}\big).$$

The question is, how can we find the parameters θ which maximize the likelihood of observed data.

The trouble is that we do not know the state of the Markov chain X at any point, so this class of problem falls into the general setting of "incomplete data" estimation, as considered by Dempster, Laird and Rubin [57].

Suppose we knew the values of X on the observation interval $[0, T]$. Then we would be able to obtain estimates by maximizing the 'complete data' likelihood. As the time between two jumps of X has an exponential distribution, and between two jumps Y is simply a Brownian motion with constant drift, straightforward calculations yield the maximum likelihood estimates

$$\widehat{a}_{ji} = \frac{\mathcal{J}_T^{ij}}{\mathcal{O}_T^i}, \qquad \widehat{c}_i = \frac{\mathcal{T}_T^i}{\mathcal{O}_T^i}$$

and hence we see that $\Gamma(X,Y) := \{\mathcal{J}_T^{ij}, \mathcal{T}_T^i, \mathcal{O}_T^i\}_{1\le i,j\le N}$ are sufficient statistics for θ given the complete data.

The EM algorithm exploits these quantities to deal with the 'incomplete data' estimation problem. It consists of two steps:

(i) (Expectation) For a given parameter set θ^n, calculate the expected value

$$\hat{\Gamma}^n = E[\Gamma(X,Y)|\mathcal{Y}_T;\theta^n]$$

(ii) (Maximization) Given the estimates of the sufficient statistics, calculate the new parameter values θ^{n+1} to solve the equation

$$E[\Gamma(X,Y)|\theta^{n+1}] = \hat{\Gamma}^n.$$

(iii) Repeat steps (i) and (ii), generating a sequence $\theta^0, \theta^1, \dots$ of estimates.

Note that step (ii) is equivalent to the usual formula for the maximum likelihood estimator from a regular exponential family in terms of sufficient statistics.

The convergence of the sequence $\{\theta^n\}_{n\in\mathbb{N}}$ is discussed in Dembo and Zeitouni [56] and Zeitouni and Dembo [191], or more generally in Dempster, Laird and Rubin [57].

In our context, we can now make these steps explicit.

(i) (Expectation) Given parameter estimates θ^n, calculate $\sigma_T(\mathcal{J}^{ij}), \sigma_T(\mathcal{T}^i)$ and $\sigma_T(\mathcal{O}^i)$, using the formulae of the previous section.

(ii) (Maximization) Calculate the new parameter estimates θ^{n+1} by

$$\widehat{a}_{ji} = \frac{\sigma_T(\mathcal{J}^{ij})}{\sigma_T(\mathcal{O}^i)}, \qquad \widehat{c}_i = \frac{\sigma_T(\mathcal{T}^i)}{\sigma_T(\mathcal{O}^i)}.$$

The required iteration is then apparent – we use estimated parameters to calculate the necessary filtered values, use these to reestimate the parameters, and repeat. This provides an explicit method for calculating the maximum likelihood estimates of the parameters of our model.

22.4 Exercises

Exercise 22.4.1. Let T be a random variable, with known density $f(t)$ supported on $[0, T]$. Let X be the Markov process $X_t = I_{\{t \geq T\}}$. Given observations Y satisfying

$$dY_t = cX_t dt + dW_t,$$

where $c \in \mathbb{R}$ is a known constant and W is a Brownian motion independent of X, find the optimal filter for X.

Exercise 22.4.2. Using the same techniques as in Section 22.3, derive the optimal filters for a Markov chain X given observations Y with dynamics

$$dY = c^\top X_t dt + d\tilde{N}_t,$$

where N is a Poisson process independent of X, and c is a vector with components $c^i > -1$.

Exercise 22.4.3. Let X be a Poisson process, and Y satisfy the dynamics

$$dY_t = -\alpha(Y_t - X_t)dt + dW_t,$$

for W a Brownian motion independent of X. Derive a filter for X given observation of Y.

Exercise 22.4.4. Derive the Wonham filter for a Markov chain observed in two-dimensional (correlated) Brownian noise.

A

Appendix

A.1 Outer Measure and Carathéodory's Extension Theorem

Unfortunately, it is often not possible to create 'natural' measure spaces using *all* subsets of a given set. For example, the Lebesgue measure λ on the real line, which associates every interval $]a, b]$ with the length $\lambda(]a, b]) = b - a$, can be well defined on the Borel σ-algebra $\mathcal{B}(\mathbb{R})$, but does not extend to the σ-algebra of all subsets $2^{\mathbb{R}}$. See Gelbaum and Olmsted [86, p.92] for further details. In this appendix, our aim is to show how we can, in a fairly general way, extend a 'measure' defined on a small family of sets (e.g. the intervals) to a larger family (e.g. the Borel sets $\mathcal{B}(\mathbb{R})$).

We begin with a useful general lemma.

Lemma A.1.1. *Consider a finitely additive, nonnegative set function μ defined on a σ-algebra of sets Σ. Then for a sequence $\{A_n\}_{n\in\mathbb{N}} \subset \Sigma$ of disjoint sets*

$$\sum_{n=1}^{\infty} \mu(A_n) \leq \mu\left(\bigcup_{n=1}^{\infty} A_n\right).$$

Proof. To see this note that for $m, n \in \mathbb{N}$ with $m \leq n$, we can write

$$\sum_{k=1}^{m} \mu(A_k) \leq \sum_{k=1}^{n} \mu(A_k) = \mu\left(\bigcup_{k=1}^{n} A_k\right).$$

The result follows by first letting $n \to \infty$ and then letting $m \to \infty$. □

S.N. Cohen, R.J. Elliott, *Stochastic Calculus and Applications*,
Probability and Its Applications, DOI 10.1007/978-1-4939-2867-5

Definition A.1.2. *A* measure on an algebra Σ *is a* $[0,\infty]$*-valued set function* $\bar{\mu}$ *on a (Boolean) algebra* Σ *such that*

(i) (Triviality) $\bar{\mu}(\emptyset) = 0$*,*
(ii) (Countable additivity) for every disjoint sequence $\{A_i\}_{i\in\mathbb{N}} \subset \Sigma$ *with* $\bigcup_{i\in\mathbb{N}} A_i \in \Sigma$ *we know*

$$\bar{\mu}\Big(\bigcup_{i\in\mathbb{N}} A_i\Big) = \sum_{i\in\mathbb{N}} \bar{\mu}(A_i).$$

Remark A.1.3. It is clear that a 'measure on an algebra', $\bar{\mu}$, is a 'true' measure if and only if Σ is a σ-algebra. To extend the domain of $\bar{\mu}$ to $\sigma(\Sigma)$, we shall first use it to construct an 'outer measure' on $2^S \supset \sigma(\Sigma)$, which is only required to be subadditive, and then show that the restriction of our outer measure to $\sigma(\Sigma)$ is a measure.

Lemma A.1.4. *Let* μ *be a finitely additive,* $\mathbb{R}$*-valued set function on an algebra* Σ *of subsets of a set* S*. Then* μ *is a measure on the algebra* Σ *(in particular, it is countably additive) if and only if* $\lim_n \mu(A_n) = 0$ *for all decreasing sequences* $\{A_n\}_{n\in\mathbb{N}} \subset \Sigma$ *(that is,* $A_{n+1} \subseteq A_n$ *for all* n*) with* $\bigcap_{n=1}^{\infty} A_n = \emptyset$*.*

Proof. Suppose μ is a measure on the algebra Σ, and $\cap_n A_n = \emptyset$. Then $\{A_n \setminus A_{n+1}\}_{n\in\mathbb{N}}$ is a sequence of disjoint sets and $\bigcup_{n\geq m}(A_n \setminus A_{n+1}) = A_m \in \Sigma$ for all $m \in \mathbb{N}$. By countable additivity

$$\mu(A_m) = \mu\bigg(\bigcup_{n\geq m}(A_n \setminus A_{n+1})\bigg) = \sum_{n\geq m} \mu(A_n \setminus A_{n+1}).$$

Since $\mu(A_m) \in \mathbb{R}$, this series converges, and thus

$$\lim_{m\to\infty} \mu(A_m) = \lim_{m\to\infty} \sum_{n\geq m} \mu(A_n \setminus A_{n+1}) = 0.$$

For the converse, first note that, as we could take $A_n = \emptyset$ for all n, we must have $\mu(\emptyset) = 0$. Now consider any sequence of disjoint sets $\{B_n\}_{n\in\mathbb{N}} \subset \Sigma$ such that $B = \cup_n B_n \in \Sigma$. Let $A_m = B \setminus (\bigcup_{n\leq m} B_n)$. Then $A_m \in \Sigma$, $A_m \subseteq A_{m+1}$ and $\cap_m A_m = \emptyset$, so $\lim_m \mu(A_m) = 0$. Since μ is finitely additive

$$\mu(B) = \mu(A_m) + \sum_{n\leq m} \mu(B_n),$$

and letting $m \to \infty$, we have

$$\mu(B) = \sum_{n=1}^{\infty} \mu(B_n).$$

$\square$

Example A.1.5 (Lebesgue Measure on the Intervals). Recall from Example 1.1.4 and Remark 1.1.10, $\Sigma_{\mathfrak{J}}$ is the algebra made up of finite unions of intervals of $\mathbb{R}$ the form $]a_i, b_i]$ or $]a_i, \infty[$. Define

$$\bar{\mu}(A) = \sum_i (b_i - a_i)$$

where $\{a_i, b_i\}$ are the endpoints of the intervals comprising A. Then $\bar{\mu}$ is a measure on an algebra (the algebra being $\Sigma_{\mathfrak{J}}$).

Definition A.1.6. *For an arbitrary set S, an* outer measure μ^* *on S is a $[0, \infty]$-valued set function defined on 2^S satisfying*

(i) $\mu^*(\emptyset) = 0$,
(ii) $\mu^*(A) \le \mu^*(B)$ *whenever* $A \subseteq B \subseteq S$,
(iii) $\mu^*\left(\bigcup_{n=1}^\infty A_n\right) \le \sum_{n=1}^\infty \mu^*(A_n)$ *for all* $\{A_n\}_{n\in\mathbb{N}} \subset 2^S$.

Definition A.1.7. *Let $\mathcal{E}$ be a collection of sets. We write $\mathfrak{C}_{\mathcal{E}}(A)$ for the collection of all finite covers of A by elements of $\mathcal{E}$, that is, collections $\{E_i\}_{i=1}^n \subset \mathcal{E}$ with $A \subseteq \cup_{i=1}^n E_i$.*

Example A.1.8 (Lebesgue Outer Measure on $\mathbb{R}$). For a set $A \subseteq \mathbb{R}$, let

$$\mu^*(A) = \inf_{\mathfrak{C}_{\mathfrak{J}}(A)} \sum_i (b_i - a_i)$$

where $\mathfrak{J}$ is the collection of intervals, so each $\left\{]a_i, b_i]\right\}_{i=1}^n \in \mathfrak{C}_{\mathfrak{J}}(A)$ is a finite collection of intervals with $A \subseteq \bigcup_{i=1}^n]a_i, b_i]$. Then μ^* is an outer measure on $\mathbb{R}$.

Definition A.1.9. *Let μ^* be a $\overline{\mathbb{R}}$-valued set function defined on 2^S such that $\mu^*(\emptyset) = 0$. A set $A \in 2^S$ is called a* μ^*-set *if*

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c)$$

for all $E \in 2^S$. The collection of μ^*-sets will be denoted Σ_{μ^*}.

In general the μ^*-sets will be the 'nice' members of 2^S, and it is using these that we can say where our *outer* measure will behave like a (true) measure. The requirement $\mu^*(\emptyset) = 0$ is so that both $\emptyset$ and S will be μ^*-sets.

Lemma A.1.10. *Let μ^* be an $\overline{\mathbb{R}}$-valued set function defined on 2^S, such that $\mu^*(\emptyset) = 0$. Then Σ_{μ^*}, the collection of μ^*-sets, is an algebra of sets, and μ^* is finitely additive on Σ_{μ^*}.*

Proof. Clearly $\emptyset, S \in \Sigma_{\mu^*}$ and if $A \in \Sigma_{\mu^*}$ then $A^c \in \Sigma_{\mu^*}$. Now suppose $A, B \in \Sigma_{\mu^*}$ and $E \in 2^S$. Then

$$\begin{aligned}\mu^*(E) &= \mu^*(E \cap A) + \mu^*(E \cap A^c)\\ &= \mu^*((E \cap A) \cap B) + \mu^*((E \cap A) \cap B^c) + \mu^*(E \cap A^c)\\ &= \mu^*(E \cap (A \cap B)) + \mu^*(E \cap (A \cap B)^c),\end{aligned}$$

where in passage to the last line we used

$$\begin{aligned}\mu^*(E \cap (A \cap B)^c) &= \mu^*(E \cap (A \cap B)^c \cap A) + \mu^*(E \cap (A \cap B)^c \cap A^c)\\ &= \mu^*((E \cap A) \cap B^c) + \mu^*(E \cap A^c).\end{aligned}$$

Therefore, Σ_{μ^*} is closed under complements and pairwise intersections, and hence, by the simple identity $A \cup B = (A^c \cap B^c)^c$, we know Σ_{μ^*} is closed under pairwise unions. That Σ_{μ^*} is closed under finite unions follows by induction.

Finite additivity follows from the fact that, for any disjoint $A, B \in \Sigma_{\mu^*}$,

$$\mu(A \cup B) = \mu\big((A \cup B) \cap A\big) + \mu\big((A \cup B) \cap A^c\big) = \mu(A) + \mu(B).$$

$\square$

Lemma A.1.11. *Let μ^* be an outer measure. If $\{A_n\}_{n=1}^m \subseteq \Sigma_{\mu^*}$ is a finite disjoint collection of sets, then, for $E \in 2^S$,*

$$\mu^*\Big(E \cap \Big(\bigcup_{n=1}^m A_n\Big)\Big) = \sum_{n=1}^m \mu^*(E \cap A_n).$$

Proof. We prove the result for $m = 2$, the general case follows by induction. Let $E \in 2^S$ and suppose $A_1, A_2 \in \Sigma_{\mu^*}$ are disjoint. Since $A_1 \in \Sigma_{\mu^*}$,

$$\begin{aligned}\mu^*(E \cap (A_1 \cup A_2)) &= \mu^*(E \cap (A_1 \cup A_2) \cap A_1) + \mu^*(E \cap (A_1 \cup A_2) \cap A_1^c)\\ &= \mu^*(E \cap A_1) + \mu^*(E \cap A_2).\end{aligned}$$

$\square$

Theorem A.1.12. *If μ^* is an outer measure, then Σ_{μ^*}, the collection of μ^*-sets, forms a σ-algebra.*

Proof. We know that Σ_{μ^*} is an algebra of sets, by Lemma A.1.10. Hence we only need to show that Σ_{μ^*} is closed under countable unions of disjoint sets.

Suppose $\{A_n\}_{n\in\mathbb{N}}$ is a sequence of disjoint sets in Σ_{μ^*}. From the definition of an outer measure, we know that for any $E \in 2^S$

$$\begin{aligned}\mu^*\Big(E \cap \Big(\bigcup_{n\in\mathbb{N}} A_n\Big)\Big) &+ \mu^*\Big(E \cap \Big(\bigcup_{n\in\mathbb{N}} A_n\Big)^c\Big)\\ &\le \sum_{n\in\mathbb{N}} \mu^*(E \cap A_n) + \mu^*\Big(E \cap \Big(\bigcup_{n\in\mathbb{N}} A_n\Big)^c\Big).\end{aligned} \tag{A.1}$$

For $E \in 2^S$ and $m \in \mathbb{N}$, by Lemma A.1.11,

$$\mu^*(E) = \mu^*\Big(E \cap \Big(\bigcup_{n=1}^m A_n\Big)\Big) + \mu^*\Big(E \cap \Big(\bigcup_{n=1}^m A_n\Big)^c\Big)$$
$$\geq \sum_{n=1}^m \mu^*(E \cap A_n) + \mu^*\Big(E \cap \Big(\bigcup_{n\in\mathbb{N}} A_n\Big)^c\Big).$$

Letting $m \to \infty$ and using (A.1), we have

$$\begin{aligned}
&\mu^*\Big(E \cap \Big(\bigcup_{n\in\mathbb{N}} A_n\Big)\Big) + \mu^*\Big(E \cap \Big(\bigcup_{n\in\mathbb{N}} A_n\Big)^c\Big)\\
&\quad\leq \sum_{n\in\mathbb{N}} \mu^*(E \cap A_n) + \mu^*\Big(E \cap \Big(\bigcup_{n\in\mathbb{N}} A_n\Big)^c\Big)\\
&\quad\leq \mu^*(E)\\
&\quad\leq \mu^*\Big(E \cap \Big(\bigcup_{n\in\mathbb{N}} A_n\Big)\Big) + \mu^*\Big(E \cap \Big(\bigcup_{n\in\mathbb{N}} A_n\Big)^c\Big).
\end{aligned}$$

As the first and last terms of this inequality are equal, we see that $\bigcup_{n\in\mathbb{N}} A_n$ is a μ^*-set. □

Definition A.1.13. *For any measure $\bar{\mu}$ on an algebra Σ, define a set function μ^* on 2^S by*

$$\mu^*(A) := \inf_{\{A_n\}_{n\in\mathbb{N}} \in \mathfrak{C}_\Sigma(A)} \sum_{n=1}^\infty \bar{\mu}(A_n).$$

Remark A.1.14. μ^* is called the outer measure induced by $\bar{\mu}$. We shall see in Theorem A.1.17 that μ^* is indeed an outer measure.

The following lemma allows us to approximate arbitrary sets 'from above' by sets in $\sigma(\Sigma)$.

Lemma A.1.15. *Let $\bar{\mu}$ be a measure on an algebra Σ, μ^* the outer measure induced by $\bar{\mu}$. Then for any $B \in 2^S$, there exists a set $A \in \sigma(\Sigma)$ with $B \subseteq A$ and $\mu^*(A) = \mu^*(B)$.*

Proof. From the definition of μ^*, for every $n \in \mathbb{N}$ there exists a sequence $\{A_i^n\}_{i\in\mathbb{N}} \subseteq \Sigma$ such that $B \subseteq \bigcup_i A_i^n$ and $\sum_i \bar{\mu}(A_i^n) \leq \mu^*(B) + n^{-1}$. Defining $A := \bigcap_n \bigcup_i A_i^n \in \sigma(\Sigma)$, we have $B \subseteq A$ and

$$\mu^*(B) \leq \mu^*(A) \leq \sum_i \mu^*(A_i^n) \leq \mu^*(B) + n^{-1}$$

for all n, so $\mu^*(A) = \mu^*(B)$. □

Remark A.1.16. Note that $\mu^*(A) = \mu^*(B)$ does not imply that $\mu^*(A \setminus B) = 0$, unless B is a μ^*-set.

The next theorem is fundamental, as it allows us to extend a measure on an algebra to a measure on a σ-algebra, typically in a unique manner. While we do not frequently refer to it, this result arguably underlies large parts of measure theory, as it is by means of this result that the measures we consider can be shown to exist.

Theorem A.1.17 (Carathéodory's Extension Theorem). *Let $\bar{\mu}$ be a measure on an algebra Σ. Then μ^*, the outer measure induced by $\bar{\mu}$, satisfies*

(a) μ^* is an outer measure,
(b) every set in Σ is a μ^*-set,
(c) $\mu^*|_\Sigma = \bar{\mu}$,
(d) μ^* is a measure on $\sigma(\Sigma)$.

Furthermore, if $\bar{\mu}$ is σ-finite on Σ, then $\mu^*|_{\sigma(\Sigma)}$ is the unique measure on $\sigma(\Sigma)$ satisfying $\mu^*|_\Sigma = \bar{\mu}$.

Proof. (a) That μ^* satisfies properties (i) and (ii) in Definition A.1.6 is immediate. For property (iii) let $\{A_n\}_{n=1}^\infty$ be any sequence of sets and put $A := \bigcup_{n=1}^\infty A_n$. Fix $\epsilon > 0$ and for each $n \in \mathbb{N}$, choose a sequence $\{A_{n,m}\}_{m=1}^\infty$ satisfying

- $A_{n,m} \in \Sigma$,
- $A_n \subset \bigcup_{m=1}^\infty A_{n,m}$, and
- $\sum_{m=1}^\infty \bar{\mu}(A_{n,m}) \le \mu^*(A_n) + \epsilon 2^{-(n+1)}$.

Then $A \subset \bigcup_{n=1}^\infty \bigcup_{m=1}^\infty A_{n,m}$ so that

$$\mu^*(A) \le \sum_{n=1}^\infty \sum_{m=1}^\infty \bar{\mu}(A_{n,m}) \le \sum_{n=1}^\infty \mu^*(A_n) + \epsilon.$$

The countable subadditivity of μ^* follows since ϵ is arbitrary.

(b) Given $\epsilon > 0$ and $E \in 2^S$, $A \in \Sigma$, let $\{E'_n\}_{n\in\mathbb{N}} \subset \Sigma$ be a sequence such that

$$\sum_{n=1}^\infty \bar{\mu}(E'_n) \le \inf_{\{E_n\}\in\mathfrak{C}_\Sigma(E)} \sum_{n=1}^\infty \bar{\mu}(E_n) + \epsilon.$$

Then

$$\begin{aligned}
&\mu^*(E\cap A) + \mu^*(E\cap A^c)\\
&\quad= \inf_{\{B_n\}\in\mathfrak{C}_\Sigma(E\cap A)} \sum_{n=1}^\infty \bar{\mu}(B_n) + \inf_{\{C_n\}\in\mathfrak{C}_\Sigma(E\cap A^c)} \sum_{n=1}^\infty \bar{\mu}(C_n)\\
&\quad\le \sum_{n=1}^\infty \bar{\mu}(E'_n\cap A) + \sum_{n=1}^\infty \bar{\mu}(E'_n\cap A^c) = \sum_{n=1}^\infty \bar{\mu}(E'_n)\\
&\quad\le \inf_{\{E_n\}\in\mathfrak{C}_\Sigma(E)} \sum_{n=1}^\infty \bar{\mu}(E_n) + \epsilon = \mu^*(E) + \epsilon.
\end{aligned}$$

The result follows as ϵ is arbitrary.

(c) This is straightforward.

(d) By parts (a)–(c) and Theorem A.1.12, since μ^* is an outer measure and $\sigma(\Sigma) \subset \Sigma_{\mu^*}$, we have that μ^* is countably additive on $\sigma(\Sigma)$. To see uniqueness, suppose that $\bar{\mu}$ is σ-finite on Σ and that $\tilde{\mu}$ is another extension of $\bar{\mu}$ to $\sigma(\Sigma)$. As $\bar{\mu}$ is σ-finite, uniqueness will follow if we show that $\mu^*(A) = \tilde{\mu}(A)$ for every $A \in \sigma(\Sigma)$ such that there exists $E \in \Sigma$ with $A \subset E$ and $\mu(E) < \infty$.

For $\epsilon > 0$, choose $\{E_n\}_{n\in\mathbb{N}} \subset \Sigma$ such that $A \subset \bigcup_n E_n$ and $\mu^*(A) + \epsilon \geq \sum_n \bar{\mu}(E_n)$. Then, since $\tilde{\mu}(A) \leq \sum_n \tilde{\mu}(E_n) = \sum_n \bar{\mu}(E_n)$, it follows that $\tilde{\mu}(A) \leq \mu^*(A)$. Similarly $\tilde{\mu}(E \setminus A) \leq \mu^*(E \setminus A)$. On the other hand,

$$\tilde{\mu}(A) + \tilde{\mu}(E \setminus A) = \tilde{\mu}(E) = \mu^*(E) = \mu^*(A) + \mu^*(E \setminus A),$$

and so $\tilde{\mu}(A) = \mu^*(A)$. □

Corollary A.1.18. *Let Σ be an algebra and μ, ν be σ-finite measures on $\sigma(\Sigma)$. If $\mu(S) = \nu(S)$ for all $S \in \Sigma$, it follows that $\mu = \nu$ everywhere.*

Proof. Considering μ, ν as measures on the algebra Σ, as they agree on Σ they have the same corresponding outer measures, and hence the same (unique) extension to $\sigma(\Sigma)$. Therefore they are equal. □

The following lemma allows us to approximate the measures of sets in $\sigma(\Sigma)$ by sequences in Σ.

Lemma A.1.19. *Let μ be a σ-finite measure on an algebra Σ. Then for any $A \in \sigma(\Sigma)$ and any $\epsilon > 0$ there exists*

(i) a sequence $\{B_n\}_{n\in\mathbb{N}} \subset \Sigma$ with $A \subseteq \bigcup_n B_n := B$ and $\mu^*(B \setminus A) \leq \epsilon$,
(ii) a sequence $\{C_n\}_{n\in\mathbb{N}} \subset \Sigma$ with $C := \bigcap_n C_n \subseteq A$ and $\mu^*(A \setminus C) \leq \epsilon$.

Proof. First assume μ is a finite measure. From the definition of the outer measure, we can find a sequence $\{B_n\}_{n\in\mathbb{N}}$ with $A \subseteq \bigcup_n B_n := B$ and

$$\mu^*(B) \leq \sum_n \mu^*(B_n) \leq \mu^*(A) + \epsilon.$$

As A is a μ^*-set, we have $\mu^*(B \setminus A) = \sum_i \mu^*(B) - \mu^*(A)$ and the result follows.

To find C_n, use (i) to take a sequence $\{D_n\}_{n\in\mathbb{N}}$ with $A^c \subseteq \bigcup_n D_n$ and $\sum_n \mu^*(D_n) \leq \mu^*(A^c) + \epsilon$. Defining $C_n := D_n^c$, we see $C := \bigcap_n C_n \subseteq A$ and $\mu^*(A \setminus C) < \epsilon$, as desired.

If μ is only σ-finite, decompose S into a countable partition $\{S_j\}_{j\in\mathbb{N}} \subset \Sigma$ with $\mu(S_j) < \infty$. Then repeat the above constructions on each S_j with ϵ replaced by $\epsilon 2^{-j}$. For part (i), we have a doubly indexed sequence $B_{j,n}$ with the desired properties. For part (ii), we find a sequence $C_{n,j}$, and define a sequence $C'_{n,j} = C_{n,j} \cup (S \setminus S_j)$, which will have the desired properties. A simple reordering gives the result. □

These results allow us to prove, for example Theorem 1.2.16, which we repeat for the reader's convenience.

Theorem A.1.20. *There is a one-to-one correspondence between distribution functions (up to addition by a constant) and Lebesgue–Stieltjes measures on $\mathcal{B}(\mathbb{R})$, given by*

$$\mu(]a,b]) = F(b) - F(a).$$

and the requirement $F(0) = 0$.

Proof. Suppose that μ is a Lebesgue–Stieltjes measure on $\mathcal{B}(\mathbb{R})$. Set

$$F(x) = \begin{cases} \mu\big(]0,x]\big) & \text{if } x \geq 0, \\ -\mu\big(]x,0]\big) & \text{if } x < 0. \end{cases}$$

It is easy to check that F is a distribution function.

Given a distribution function F, we show that F generates a Lebesgue–Stieltjes measure on $\Sigma_{\mathfrak{I}}$, the algebra made from finite (disjoint) unions of intervals of the form $]a,b]$ and $]a,\infty[$.

Let $F(\infty) := \lim_{x\to\infty} F(x)$, $F(-\infty) := \lim_{x\to-\infty} F(x)$. For each element of $\Sigma_{\mathfrak{I}}$, we can define a nonnegative, finitely additive set function μ by

$$\mu(]a,b]) = F(b) - F(a),$$

$$\mu(]a,\infty[) = F(\infty) - F(a)$$

and

$$\mu\left(\bigcup_{n=1}^{m} I_n\right) = \sum_{n=1}^{m} \mu(I_n)$$

for disjoint $I_n \in \Sigma_{\mathfrak{I}}$. Although straightforward, a proof that μ is well defined is left as an exercise.

First suppose $F(\infty) - F(-\infty) < \infty$. We verify the sufficient condition of Theorem A.1.17 to show that μ extends uniquely to a measure on $\mathcal{B}(\mathbb{R})$.

We must first establish that μ is a measure on the algebra $\Sigma_{\mathfrak{I}}$. Let $\{A_n\}_{n\in\mathbb{N}}$ be a sequence of sets in $\Sigma_{\mathfrak{I}}$ such that for each n, $A_{n+1} \subseteq A_n$ and $\bigcap_{n=1}^{\infty} A_n = \emptyset$. For each n, there exists $m_n \in \mathbb{N}$ such that A_n has the form

$$A_n =]a_{n,1}, b_{n,1}] \cup \cdots \cup]a_{n,m_n}, b_{n,m_n}] \cup]a_{n,m_n+1}, \infty[.$$

Since

$$\begin{aligned} \mu(A_n) &= \sum_{i=1}^{m_n} \mu\big(]a_{n,i}, b_{n,i}]\big) + \mu\big(]a_{n,m_n+1}, \infty[\big) \\ &= \sum_{i=1}^{m(n)} \big(F(b_{n,i}) - F(a_{n,i})\big) + F(\infty) - F(a_{n,m_n+1}), \end{aligned}$$

and F is right-continuous, we have $\lim_{n\to\infty} \mu(A_n) = 0$. Thus by Lemma A.1.4, μ is countably additive and hence a measure on the algebra $\Sigma_{\mathfrak{J}}$. By Theorem A.1.17, μ has a unique extension to $\sigma(\Sigma_{\mathfrak{J}}) = \mathcal{B}(\mathbb{R})$ (see Exercise 1.8.2).

Since any bounded interval I is contained in an interval of the form $]a, b]$ and $\mu(]a, b]) = F(b) - F(a) < \infty$, μ is a Lebesgue–Stieltjes measure.

Now suppose $F(\infty) - F(-\infty) = \infty$. We define

$$F_r(x) := \begin{cases} F(r) & \text{if } x > r, \\ F(x) & \text{if } |x| \leq r, \\ F(-r) & \text{if } x < -r. \end{cases}$$

Let μ_r be the set function defined by F_r and let $\{A_n\}_{n\in\mathbb{N}}$ be a sequence of disjoint sets in $\Sigma_{\mathfrak{J}}$ such that $A := \bigcup_{n=1}^{\infty} A_n \in \Sigma$. Since μ, the set function defined by F, is nonnegative and finitely additive, by Lemma A.1.1 we have $\mu(A) \geq \sum_{n=1}^{\infty} \mu(A_n)$. Hence, if $\sum_{n=1}^{\infty} \mu(A_n) = \infty$, the result is proven. If $\sum_{n=1}^{\infty} \mu(A_n) < \infty$, then

$$\mu(A) = \lim_{r\to\infty} \mu_r(A) = \lim_{r\to\infty} \sum_{n=1}^{\infty} \mu_r(A_n)$$

since the μ_r are finite. Now $\sum_{n=1}^{\infty} \mu(A_n) < \infty$ and $\mu_r \leq \mu$, so we may write

$$0 \leq \mu(A) - \sum_{n=1}^{\infty} \mu(A_n) = \lim_{r\to\infty} \sum_{n=1}^{\infty} (\mu_r(A_n) - \mu(A_n)) \leq 0.$$

By Theorem A.1.17, μ again has a unique extension to $\sigma(\Sigma_{\mathfrak{J}}) = \mathcal{B}(\mathbb{R})$. □

Exercise A.1.21. Prove that the measure constructed in Example A.1.5 is a measure on the algebra $\Sigma_{\mathfrak{J}}$. Hence or otherwise, show that any countable set in $\sigma(\Sigma_{\mathfrak{J}})$ must have (outer) measure zero.

A.2 Kolmogorov's Extension Theorem

In this section, we give useful fundamental results for constructing stochastic processes. These allow us to define a process through its behaviour on finite collections of time points, and then extend this behaviour to continuous time. In some sense, if Carathéodory's extension theorem (Theorem A.1.17) allows us to extend measures over measurable spaces, Kolmogorov's extension theorem allows us to extend these measures over time as well.

To establish these results, we first need to be more precise about the nature of the space of real functions on $[0, \infty[$ and $\{0, 1, \ldots\}$.

Definition A.2.1. *Let $\mathbb{T} = [0, \infty]$ or $[0, \infty[$. Then $(\mathbb{R}^d)^{\mathbb{T}}$ denotes the space of real functions $x : \mathbb{T} \to \mathbb{R}^d$. The 'cylinder topology' on $(\mathbb{R}^d)^{\mathbb{T}}$ is given by finite intersections of sets of the form $\{x_t \in B\}$, where $t \in \mathbb{T}$ and B is an open set in $\mathbb{R}^d$. This in turn defines the Borel cylinder σ-algebra, denoted $\mathcal{B}((\mathbb{R}^d)^{\mathbb{T}})$.*

We recall that $\mathcal{B}((\mathbb{R}^d)^\infty)$ denotes the Borel σ-algebra on sequences in $\mathbb{R}^d$, and is given by the product $\otimes_{n\in\mathbb{N}}\mathcal{B}(\mathbb{R}^d)$.

Lemma A.2.2. *A set $A \in \mathcal{B}((\mathbb{R}^d)^\mathbb{T})$ if and only if there are a countable collection of points $t_1, t_2, \dots$ and a Borel set $B \in \mathcal{B}((\mathbb{R}^d)^\infty)$ with*

$$A = \{x : (x_{t_1}, x_{t_2}, \dots) \in B\}.$$

A set of this form where the collection of points $t_1, t_2 \dots$ is finite is called a 'cylinder set'.

Proof. Let $\mathcal{A}$ denote the collection of sets A with the stated representation. Clearly $\mathcal{A}$ contains finite intersections of sets of the form $\{x_t \in B\}$ and is a monotone class. Hence, by the monotone class theorem (Theorem 1.1.14), we have $\mathcal{B}((\mathbb{R}^d)^\mathbb{T}) \subseteq \mathcal{A}$. Conversely, every set $A \in \mathcal{A}$ can be written as a countable union of Borel measurable rectangles

$$A = \bigcup_{j\in\mathbb{N}} \{x_{t_{1,j}} \in B_{1,j}\} \times \{x_{t_{2,j}} \in B_{2,j}\} \times \dots$$

and so $\mathcal{A} \subseteq \mathcal{B}((\mathbb{R}^d)^\mathbb{T})$, and the result is proven. □

A particularly important property of the real numbers, without which our extension of measures from cylinder sets to general Borel measurable sets would not be possible, is given by the following lemma. This is closely related to Lemma A.1.19, and to the more general considerations in Section 2.6, but involves the topology of $\mathbb{R}^d$ directly, rather than simply through the Borel σ-algebra.

Lemma A.2.3. *A set $A \in \mathcal{B}(\mathbb{R}^d)$ is* regular*, that is, for every finite measure μ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ and every $\epsilon > 0$ we can find a compact set C and an open set B such that $C \subseteq A \subseteq B$ and $\mu(B \setminus C) < \epsilon$.*

Proof. We use a monotone class argument. The result is easy for a set of the form $A = \big(]a_1, b_1] \times \dots \times]a_n, b_n]\big) \cap \mathbb{R}^d$ (where $a_i, b_i = \infty$ is permitted). These sets form an algebra generating $\mathcal{B}(\mathbb{R}^d)$. Now suppose $\{A_n\}_{n\in\mathbb{N}}$ is a sequence of regular sets with associated $\{B_n\}_{n\in\mathbb{N}}, \{C_n\}_{n\in\mathbb{N}}$ such that $\mu(B_n \setminus C_n) \leq \epsilon 2^{-n}$. If $\{A_n\}_{n\in\mathbb{N}}$ is increasing, then $B = \cup_n B_n$ is an open set containing $\cup_n A_n$. As the measure is finite, we know $\mu\big(B \setminus (\cup_{m\leq n} B_m)\big) \downarrow 0$ and so, for any n,

$$\begin{aligned}\mu\big(B \setminus (\cup_{m\leq n} C_m)\big) &\leq \mu\big(B \setminus (\cup_{m\leq n} B_m)\big) + \sum_{m\leq n} \mu(B_m \setminus C_m) \\ &\leq \mu\big(B \setminus (\cup_{m\leq n} B_m)\big) + \epsilon\end{aligned}$$

which can be made arbitrarily small. As $\cup_{m\leq n} C_m$ is compact for any finite m, we see that $\cup_n A_n$ is regular. Conversely, if A_n is a decreasing sequence, then we see that $C = \cap_n C_n$ is a compact contained within $\cap_n A_n$. Again, by finiteness of μ we have $\mu(C_n \setminus C) \downarrow 0$.

For any n,

$$\mu\big((\cap_{m\le n}B_m)\setminus C\big)\le\mu(B_n\setminus C_n)+\mu(C_n\setminus C)\le\epsilon 2^{-n}+\mu(C_n\setminus C).$$

As this can also be made arbitrarily small and $\cap_{m\le n}B_m$ is open for finite m, we see that $\cap_n A_n$ is regular. Hence the regular sets form a monotone class, and the result holds for all Borel measurable sets. □

Definition A.2.4. *Let $\mathbb{T}=[0,\infty]$ or $[0,\infty[$. For each finite subset T of $\mathbb{T}$, consider P_T, a probability measure on $((\mathbb{R}^d)^T,\mathcal{B}((\mathbb{R}^d)^T))$. We say the family $\{P_T\}$ is* consistent *if, for all finite (unordered) collections*

$$S=\{s_1,s_2,\dots,s_M\}\subseteq T=\{s_1,s_2,\dots,s_M,t_1,\dots t_N\}\subseteq\mathbb{T},$$

the probability measure on $(\mathbb{R}^d)^M=\{(x_{s_1},x_{s_2},\dots,x_{s_M})\}$ given by P_S is the same as the probability given by P_T restricted to its first M components,

$$\begin{aligned}&P_S(\{x\in(\mathbb{R}^d)^M:(x_{s_1},x_{s_2},\dots,x_{s_M})\in B\})\\&\qquad=P_T(\{x\in(\mathbb{R}^d)^{M+N}:(x_{s_1},x_{s_2},\dots,x_{s_M},x_{t_1},\dots,x_{t_N})\in B\times(\mathbb{R}^d)^N\})\end{aligned}$$

Remark A.2.5. One can often skirt the full generality of this condition by first defining the probabilities on the *ordered* values, and then stating that if $o(t_1,t_2,...,t_N)$ is the ordered rearrangement of $t_1,t_2,\dots t_N$, then

$$P_{\{t_1,t_2,\dots,t_N\}}:=P_{\{o(t_1,t_2,\dots,t_N)\}}.$$

In this case, one needs only show that if $B_{s_i}\in\mathcal{B}(\mathbb{R}^d)$ for each i, then, for ordered sequences $\{s_1,s_2,\dots,s_M\}\subseteq\{t_1,t_2,\dots,t_N\}$

$$\begin{aligned}&P_S(B_{s_1}\times B_{s_2}\times\dots\times B_{s_M})\\&\qquad=P_T(B_{s_1}\times B_{s_2}\times\dots\times B_{s_i}\times\mathbb{R}^d\times\mathbb{R}^d\times B_{s_{i+1}}\times\dots\times B_{s_M}),\end{aligned}$$

where $\mathbb{R}^d$ is inserted wherever there is no element of $\{s_i\}$ equal to the corresponding t_i.

Theorem A.2.6 (Extension of Measures). *Let $\{P_T\}$ be a consistent family of probability measures, where P_T is defined for all finite sets $T\subseteq\mathbb{T}$. Then there is a unique probability measure P on $((\mathbb{R}^d)^{\mathbb{T}},\mathcal{B}((\mathbb{R}^d)^{\mathbb{T}}))$ such that $P(\pi_T(B))=P_T(B)$ for all $B\in\mathcal{B}((\mathbb{R}^d)^T)$, where π_T is the projection $\pi_T(B)=\{x\in(\mathbb{R}^d)^{\mathbb{T}}:(x_{t_1},x_{t_2},\dots,x_{t_N})\in B\}$.*

Proof. Let $\mathcal{A}$ denote the subalgebra of $\mathcal{B}(\mathbb{R}^{\mathbb{T}})$ given by sets of the form

$$\{x\in(\mathbb{R}^d)^{\mathbb{T}}:(x_{t_1},x_{t_2},\dots,x_{t_n})\in B,B\in\mathcal{B}((\mathbb{R}^d)^n)\}$$

for finite sequences $\{t_1,t_2,\dots,t_n\}$. Note that, as we restrict to finite sequences, $\mathcal{A}$ is an algebra, but not a σ-algebra. We can define a measure P on the algebra $\mathcal{A}$ by $P(\pi_T(B))=P_T(B)$ for all $B\in\mathcal{B}((\mathbb{R}^d)^T)$.

We need to show that P is countably additive. By Lemma A.1.4, it is enough to show that $\lim_n P(A_n) = 0$ for $\{A_n\}_{n\in\mathbb{N}}$ a nonincreasing sequence of sets in $\mathcal{A}$ with $\cap_n A_n = \emptyset$. As the sequence $\{A_n\}_{n\in\mathbb{N}}$ is nonincreasing, we know that $P(A_n) = P(A_n \setminus A_{n+1}) + P(A_{n+1}) \geq P(A_{n+1})$, so the limit $\lim_n P(A_n)$ exists, and is within $[0,1]$ by construction.

Suppose that $\lim_n P(A_n) = \epsilon > 0$. As $A_n \in \mathcal{A}$, we know that there exists a sequence $t_1, t_2, \dots$ such that we can write

$$A_n = \{x \in (\mathbb{R}^d)^{\mathbb{T}} : (x_{t_1}, x_{t_2}, \dots, x_{t_{k(n)}}) \in B_n, B_n \in \mathcal{B}((\mathbb{R}^d)^{k(n)})\}$$

for some function $k : \mathbb{N} \to \mathbb{N}$. By Lemma A.2.3, we can find compact sets $C_n \subseteq B_n$ such that the corresponding events

$$D_n = \big\{x \in (\mathbb{R}^d)^{\mathbb{T}} : (x_{t_1}, x_{t_2}, \dots, x_{t_{k(n)}}) \in C_n\big\} \subseteq A_n$$

satisfy $P(A_n \setminus D_n) \geq \epsilon 2^{-n}$. Taking an intersection, we see that

$$\cap_{i\leq n} D_i = \big\{x \in (\mathbb{R}^d)^{\mathbb{T}} : (x_{t_1}, \dots, x_{t_{k(n)}}) \in \tilde{C}_n := \cap_{i\leq n}(C_i \times (\mathbb{R}^d)^{(t_{k(n)} - t_{k(i)})})\big\}.$$

Therefore,

$$\begin{aligned} P(\cap_{i\leq m} D_i) &= P(A_m) - P\big(A_m \setminus (\cap_{n\leq m} D_n)\big) = P(A_m) - P\big(\cup_{n\leq m} (A_m \setminus D_n)\big) \\ &\geq P(A_m) - P\big(\cup_{n\leq m} (A_n \setminus D_n)\big) \geq \epsilon - \epsilon \sum_{n\leq m} 2^{-m} > 0. \end{aligned}$$

Therefore, for each m, $\cap_{i\leq m} D_i$ is nonempty, and so $\tilde{C}_n$ is also nonempty. Take an arbitrary $c_m = (c_m^1, c_m^2, \dots, c_m^k(m)) \in \tilde{C}_m$. As the sets $\tilde{C}_m$ are nonincreasing, the sequence $\{(c_m^1, \dots, c_m^{k(1)})\}_{m\in\mathbb{N}} \subset (\mathbb{R}^d)^{k(1)}$ is within the compact set $\tilde{C}_1$, and hence has a convergent subsequence with limit $(c^1, \dots, c^{k(1)})$. Similarly, taking subsequences, the sequence $\{(c_m^1, \dots, c_m^{k(2)})\}_{m\in\mathbb{N}} \subset (\mathbb{R}^d)^{k(2)}$ is within the compact set $\tilde{C}_2$, and so we have a limit $(c^1, \dots, c^{k(2)})$.

We therefore obtain a sequence $(c^1, c^2, \dots)$ with $(c^1, c^2, \dots, c^{k(n)}) \in \cap_{i\leq n} \tilde{C}_i$ for every n. It follows that the event $(\cap_{i=1}^{\infty}\{x_{t_i} = c^i\}) \in D_n$ for every n, which implies that $\cap_n D_n$ is nonempty. This is a contradiction with the fact that $\cap_n D_n \subseteq \cap_n A_n = \emptyset$. Therefore, we must have $\lim_n P(A_n) = 0$, and so P is countably additive.

By Carathéodory's extension theorem (Theorem A.1.17) we can now extend P uniquely to a measure on $\sigma(\mathcal{A})$, which is equal to $\mathcal{B}((\mathbb{R}^d)^{\mathbb{T}})$, by Lemma A.2.2. □

A fundamental application of this result is to show the existence of processes with predefined law. For example, this gives a construction of Brownian motion (Theorem 5.5.4).

Theorem A.2.7 (Existence of Processes). *Let* $(\Omega, \mathcal{F}) = ((\mathbb{R}^d)^{\mathbb{T}}, \mathcal{B}((\mathbb{R}^d)^{\mathbb{T}}))$, *and consider a consistent family of probability measures* $\{P_T\}$ *on* $(\mathbb{R}^d)^{\mathbb{T}}$. *Then there exists a measure* P *such that the canonical process* $X_t(\omega) = \omega_t$ *has the joint probability laws*

$$P\big(\{(X_{t_1}, X_{t_2}, \dots, X_{t_N}) \in B\}\big) = P_{\{t_1, t_2, \dots, t_N\}}(B) \qquad \textit{for all } B \in (\mathbb{R}^d)^N.$$

Proof. This is simply an application of Theorem A.2.6. □

A.3 Regular Conditional Probability

We here prove Theorem 2.6.7 on the existence of regular conditional probabilities, which we reproduce here for convenience. This proof is taken from [21].

Theorem A.3.1. *Consider a (countably additive signed) finite measure* μ *on a measurable space* $(\Omega, \mathcal{F})$.

(i) Suppose that $\mathcal{F}$ *is countably generated (that is, there exists a sequence of sets* A_n *such that* $\mathcal{F} = \sigma(\{A_n\}_{n\in\mathbb{N}})$*) and that* μ *has a compact approximating class in* $\mathcal{F}$. *Then, for any sub-*σ*-algebra* $\mathcal{G}$ *of* $\mathcal{F}$, *there exists a regular conditional measure* $\mu|_{\mathcal{G}}$ *on* $\mathcal{F}$.

(ii) More generally, let $\tilde{\mathcal{F}}$ *be a sub-*σ*-algebra of* $\mathcal{F}$ *generated by a countable algebra of sets* $\mathcal{U}$. *Suppose that there is a compact class* $\mathcal{K}$ *such that for every* $A \in \mathcal{U}$ *and* $\epsilon > 0$, *there exist* $K_\epsilon \in \mathcal{K}$ *and* $A_\epsilon \in \mathcal{F}$ *with* $A_\epsilon \subseteq K_\epsilon \subseteq A$ *and* $|\mu|(A \setminus A_\epsilon) < \epsilon$. *Then, for every sub-*$\sigma$*-algebra* $\mathcal{G} \subseteq \mathcal{F}$, *there exists a regular conditional measure* $\mu|_{\mathcal{G}}$ *on* $\tilde{\mathcal{F}}$ *given* $\mathcal{G}$ *(which can be taken to be a probability measure if* μ *is nonnegative). In addition, for every* $\tilde{\mathcal{F}}$*-measurable* μ*-integrable function* f, *one has*

$$\int_\Omega f\, d\mu = \int_\Omega \int_\Omega f(\omega')\mu|_{\mathcal{G}}(d\omega', \omega)|\mu|(d\omega).$$

Proof. We shall prove the more general second assertion. Suppose first that μ is a probability measure.

We know that $\mathcal{U} = \{A_n\}_{n\in\mathbb{N}}$ is a countable algebra of sets, which generates $\mathcal{F}$. For each n, we find sets $C_{n,k} \in \mathcal{K}$ and $A_{n,k} \in \mathcal{F}$ such that

$$A_{n,k} \subseteq C_{n,k} \subseteq A_n \quad \text{ and } \quad \mu(A_n \setminus A_{n,k}) < 1/k.$$

The sets $A_{n,k}$, along with the sets A_n, generate a countable algebra $\mathcal{U}_0 \subseteq \mathcal{F}$. By the Radon–Nikodym theorem, for every set $A \in \mathcal{U}_0$ there exists a nonnegative $\mathcal{G}$-measurable function $\omega \mapsto p_0(A, \omega)$ such that $p_0(\Omega, \omega) = 1$ and $p_0(\emptyset, \omega) = 0$ for all ω, and for all $B \in \mathcal{G}$,

$$\mu(A \cap B) = \int_B p_0(A, \omega)\mu(d\omega).$$

In particular, by uniqueness of the Radon–Nikodym derivative, we have that $p_0(A \cup B, \omega) = p_0(A, \omega) + p_0(B, \omega)$ for almost all ω, for any disjoint $A, B \in \mathcal{U}_0$. As the set of pairs $(A, B) \in \mathcal{U}_0$ is countable, it follows that there exists a set $N \in \mathcal{G}$ of μ-measure zero such that for all $\omega \notin N$ the function $A \mapsto p_0(A, \omega)$ is (finitely) additive on $\mathcal{U}_0$.

Write $q_n(\omega) := \sup_k p_0(A_{n,k}, \omega)$. It is clear that q_n is $\mathcal{G}$-measurable, and the inclusion $A_{n,k} \subseteq A_n$ implies that there exist measure zero sets $N_{n,k} \in \mathcal{G}$ such that $p_0(A_{n,k}, \omega) \leq p_0(A_n, \omega)$ for all $\omega \notin N_{n,k}$. Therefore, $q_n(\omega) \leq p_0(A_n, \omega)$ for all $\omega \notin \cup_{n,k} N_{n,k}$. On the other hand, we clearly have $p_0(A_{n,k}, \omega) \leq q_n(\omega)$, so

$$\mu(A_{n,k}) = \int_\Omega p_0(A_{n,k}, \omega)\mu(d\omega) \leq \int_\Omega q_n(\omega)\mu(d\omega).$$

As we know that $\cup_{n,k} N_{n,k}$ is a μ-null set, using the approximation assumption of the theorem we obtain

$$\mu(A_n) = \sup_k \mu(A_{n,k}) \leq \int_\Omega q_n(\omega)\mu(d\omega) \leq \int_\Omega p_0(A_n, \omega)\mu(d\omega) = \mu(A_n),$$

that is, $\sup_k p_0(A_{n,k}, \omega) = q_n(\omega) = p_0(A_n, \omega)$ except on some null set N_1.

Define $N = N_0 \cup N_1$. Then for all $\omega \notin N$, the additive set function $p_0(\cdot, \omega)$ has the property that the compact class $\mathcal{K}$ approximates $p_0(\cdot, \omega)$ on $\mathcal{U}$. The same argument as used in Theorem A.2.6 shows that $p_0(\cdot, \omega)$ is countably additive, that is, it is a measure on the algebra $\mathcal{U}_0$. By Carathéodory's extension theorem (Theorem A.1.17), we can extend $p_0(\cdot, \omega)$ uniquely to a measure $\mu|_\mathcal{G}$ on $\tilde{\mathcal{F}} = \sigma(\mathcal{U}) = \sigma(\mathcal{U}_0)$. As $p_0(\Omega, \omega) = 1$, we obtain a probability measure. For all $\omega \in N$, we simply define $\mu|_\mathcal{G} = \mu$.

We can now verify that the object $\mu|_\mathcal{G}$ has the appropriate measurability properties. We know that if $A \in \mathcal{U}$, then $\mu|_\mathcal{G}(A, \cdot)$ is $\mathcal{G}$-measurable, and by dominated convergence, sets A with this measurability form a monotone class. Therefore, we have the desired measurability for all $A \in \sigma(\mathcal{U}) = \tilde{\mathcal{F}}$.

Suppose now that μ is a general finite nonnegative measure. Then we can apply the above result to $\mu/(\mu(\Omega))$, which is a probability measure. This defines the conditional measures $\mu|_\mathcal{G}$, which are probability measures, and simple calculation shows that they have the desired properties for both $\mu/(\mu(\Omega))$ and μ. Finally, suppose that μ is a general finite signed measure. Then we can apply the above results to μ^+ and μ^- respectively, and take the difference of the resulting conditional measures. □

A.4 Continuity Results

We here prove the Kolmogorov–Čentsov theorem (Theorem 5.5.9), which is used to establish the Hölder continuity of Brownian motion. This proof is adapted from the presentation in [155]. We restate the result for the reader's convenience.

Theorem A.4.1 (Kolmogorov–Čentsov Theorem). *Let X be a Banach space valued measurable process such that, for some positive α, β, c, for all $s < t$,*

$$E\big[\|X_t - X_s\|^\alpha\big] \leq c|t-s|^{1+\beta}.$$

Then there exists a modification $\tilde{X}$ of X which is almost surely locally Hölder γ-continuous for all $\gamma \in]0, \beta/\alpha[$. In particular, for each T, there exists a constant $k > 0$ such that for all $\delta > 0$,

$$P\Big(\sup_{\{s<t<T\}}\Big\{\frac{\|\tilde{X}_t - \tilde{X}_s\|}{|t-s|^\gamma}\Big\} > \delta\Big) \leq k\delta^{-\alpha}.$$

Proof. Fix $T = 1$ for simplicity, the general result will hold by induction. For $n \in \mathbb{N}$, let D_n be the dyadic rationals of the form $D_n = \{k2^{-n}\}_{k\in\mathbb{Z}^+} \subset [0,1[$, let $\Delta_n = 2^{-n}$ and let $\lfloor t \rfloor_n = \max\{s \in D_n : s \leq t\}$, $\lceil t \rceil_n = \min\{s \in D_n : s \geq t\}$, as in Lévy's construction of Brownian motion. Then for $t \in D_n$, the assumption states that

$$E\big[\sup_{t\in D_n} |X_{t+\Delta_n} - X_t|^\alpha\big] \leq \sum_{t\in D_n} E[|X_{t+\Delta_n} - X_t|^\alpha] \leq c2^{n+1}\Delta_n^{1+\beta} = c2^{1-n\beta}.$$

Define $\mu_n = \sup_{t\in D_n} |X_{t+\Delta_n} - X_t|$.

For $s, t \in \cup_n D_n$, with $s < t$, the sequences $\{\lceil s \rceil_n\}_{n\in\mathbb{N}}$, $\{\lfloor t \rfloor_n\}_{n\in\mathbb{N}}$ will equal s, t respectively after at most finitely many steps. Then we can write

$$X_t - X_s = \Big(X_{\lfloor t\rfloor_m} + \sum_{i=m}^{\infty}(X_{\lfloor t\rfloor_{i+1}} - X_{\lfloor t\rfloor_i})\Big) - \Big(X_{\lceil s\rceil_m} + \sum_{i=m}^{\infty}(X_{\lceil s\rceil_{i+1}} - X_{\lceil s\rceil_i})\Big)$$

where all the sums are in fact finite. Therefore, as $\lfloor t \rfloor_{i+1} = \lfloor t \rfloor_i + \Delta_{i+1}$ unless $\lfloor t \rfloor_i = t$,

$$\begin{aligned}\|X_t - X_s\| &\leq \|X_{\lfloor t\rfloor_m} - X_{\lceil s\rceil_m}\| + \sum_{i=m}^{\infty}\|X_{\lfloor t\rfloor_{i+1}} - X_{\lfloor t\rfloor_i}\| \\ &\quad + \sum_{i=m}^{\infty}\|X_{\lceil s\rceil_{i+1}} - X_{\lceil s\rceil_i}\| \\ &\leq \|X_{\lfloor t\rfloor_m} - X_{\lceil s\rceil_m}\| + 2\sum_{i=m+1}^{\infty}\mu_i.\end{aligned}$$

In particular, if $|t-s| < 2^{-m}$, then $\lfloor t \rfloor_m = \lceil s \rceil_m$, and so $\|X_t - X_s\| \leq 2\sum_{i=m+1}^{\infty}\mu_i$. From this, we see

$$M_\gamma := \sup_{\{s,t \in \cup_n D_n\}} \frac{\|X_t - X_s\|}{|t-s|^\gamma} \leq \sup_{\substack{m,n\in\mathbb{Z}^+ \\ m<n}} \left\{ \sup_{\substack{s,t\in D_n \\ |t-s|<2^{-m}}} \frac{\|X_t - X_s\|}{|t-s|^\gamma} \right\}$$
$$\leq \sup_{\substack{m,n\in\mathbb{Z}^+ \\ m<n}} \frac{2\sum_{i=m+1}^\infty \mu_i}{2^{-\gamma n}} = \sup_{m\in\mathbb{Z}^+} \left\{ 2^{1+\gamma(m+1)} \sum_{i=m+1}^\infty \mu_i \right\}$$
$$\leq 2^{1+\gamma} \sum_{i=1}^\infty 2^{\gamma i} \mu_i.$$

If $\alpha \geq 1$, by Minkowski's inequality

$$E\big[|M_\gamma|^\alpha\big]^{1/\alpha} \leq 2^{1+\gamma} \sum_{i=1}^\infty 2^{\gamma i} E\big[|\mu_i|^\alpha\big]^{1/\alpha} \leq 2^{1+\gamma} \sum_{i=1}^\infty 2^{\gamma i} c^{1/\alpha} 2^{(1-i\beta)/\alpha}$$
$$= c^{1/\alpha} 2^{1+\gamma+1/\alpha} \sum_{i=1}^\infty 2^{(\gamma-\beta/\alpha)i} = \frac{c^{1/\alpha} 2^{1+\gamma+1/\alpha}}{2^{(\gamma-\beta/\alpha)} - 1} < \infty.$$

For $\alpha < 1$, the same bound holds for $E\big[|M_\gamma|^\alpha\big]$.

Therefore, as M_γ is almost surely finite, for almost all ω, X is uniformly continuous on $\cup_n D_n$. Defining $\tilde{X}_t := \lim_n X_{\lfloor t \rfloor_n}$, by Fatou's inequality we see $X_t = \tilde{X}_t$ almost surely, and

$$\sup_{\{s<t<T\}} \frac{\|\tilde{X}_t - \tilde{X}_s\|}{|t-s|^\gamma} = M_\gamma,$$

so $\tilde{X}$ is a uniformly continuous modification of X. Finally, by Markov's inequality, $P(M_\gamma > \delta) \leq E\big[|M_\gamma|^\alpha\big]\delta^{-\alpha}$, so the stated inequality holds. The Borel–Cantelli lemma (with $\delta = n^{2/\alpha}$) shows $\tilde{X}$ is almost surely Hölder γ-continuous. □

A.5 A Progressive But Not Optional Set

We here give an example of a set which is progressive but not optional, taken from Dellacherie [53, p.128]. Let X be a one-dimensional Brownian motion, starting at zero, in some probability space with a complete right-continuous filtration.

Define the set

$$G = \bigcup_{r\in\mathbb{Q}:r>0} \big\{ \sup\{s < r : X_s = 0\} \big\}.$$

As this is a countable union of random times $\sup\{s < r : X_s = 0\}$, G is a measurable set. (Note, however, that these times are not stopping times.)

By the same argument, for every t, as $\mathcal{F}_t = \mathcal{F}_{t+}$, the set

$$G \cap [\![0,t]\!] = \bigcap_{\epsilon>0} \Big\{ \bigcup_{\{r\in\mathbb{Q}: r<t+\epsilon\}} \big\{ \sup\{s<r : X_s = 0\}\big\}\Big\} \in \mathcal{B}([0,t]) \otimes \mathcal{F}_t$$

so, by Exercise 7.7.2, G is a progressive set.

By Exercise 5.7.8, as $X_0 = 0$, for almost all ω there exists no $\epsilon > 0$ such that $X_s \neq 0$ for all $s \in \,]0,\epsilon]$. Therefore, we know that G does not contain the time 0, or more precisely, $G \cap [\![0]\!]$ is evanescent. By the strong Markov property for Brownian motion (Exercise 14.7.7), this argument also holds for the process $\tilde{X}_t = X_{t-T} - X_T$, whenever T is a stopping time with $X_T = 0$. Therefore, we see that $G \cap [\![T]\!]$ is an evanescent set for any stopping time T with $X_T = 0$. By construction $G \subset \{s : X_s = 0\}$. Therefore, $G \cap [\![T]\!]$ is evanescent for any stopping time T. If G were optional, this would contradict the optional section theorem (Theorem 7.3.17). Therefore, we see that G is a progressive but not optional set.

A.5.1 Zeros of Brownian Motions

We can also use the setting of the above example to see that a Brownian motion has almost surely uncountably many zeros. Let $\mathcal{Z}(\omega) = \{t : X_t(\omega) = 0\}$. As X is continuous, $\mathcal{Z}(\omega)$ is closed.

By the above argument, for any stopping time T with $X_T = 0$, there exists almost surely a sequence of points in $\mathcal{Z}(\omega)$ decreasing to T. Therefore, any zero of X which is at a stopping time almost surely cannot be an isolated point in $\mathcal{Z}(\omega)$. This includes all points of the form $\inf\{t > q : X_t = 0\}$ for q a rational number.

For a random time τ such that $X_{\tau(\omega)}(\omega) = 0$, if we know that, for every $q \in \mathbb{Q}$, $\tau(\omega) \neq \inf\{s > q : X_s(\omega) = 0\}$, then there exists a sequence of rationals q_n increasing to $\tau(\omega)$ such that $q_n \leq s_n = \inf\{s > q_n : X_s(\omega) = 0\} < \tau(\omega)$. Therefore, $s_n \uparrow \tau(\omega)$ and $X_{s_n}(\omega) = 0$ for all n. Hence $\tau(\omega)$ is not isolated in $\mathcal{Z}(\omega)$. It follows that $\mathcal{Z}(\omega)$ cannot contain any islolated points, and so is almost surely a perfect set (a closed set with no isolated points). From basic properties of perfect sets (see, for example, [160, p.53]), we know $\mathcal{Z}(\omega)$ is uncountable.

Finally, from Fubini's theorem, as $X_t \sim N(0,t)$, we also know

$$E[I_{\{t\in\mathcal{Z}\}}] = P(t \in \mathcal{Z}) = P(X_t = 0) = 0$$

and so, by Fubini's theorem,

$$E\Big[\int_{[0,\infty]} I_{\{t\in\mathcal{Z}\}} dt\Big] = \int_{[0,\infty]} E[I_{\{t\in\mathcal{Z}\}}]dt = 0$$

which implies $\mathcal{Z}(\omega)$ is almost surely a Lebesgue null set.

The interested reader should see Rogers [158] for further exploration of these and related issues.

A.6 Results on Semimartingales

In this section, we prove three fundamental results related to the theory of semimartingales. The first is that the definition of the class of X-integrable processes (where X is a semimartingale) is exhaustive, in the sense of Theorem 12.3.22. The second gives an alternative characterization of semimartingales as the class of 'good integrators'. The third is that the set of stochastic integrals $\{H\bullet X\}_{H\in L(X)}$ is complete in the semimartingale topology. Together they demonstrate the generality of the theory of stochastic integration with respect to semimartingales. Results are stated in Chapter 12.

A.6.1 A Description of X-Integrable Processes

Recall that we defined a predictable process H to be X-integrable (Definition 12.3.10) if there exists a decomposition $X = M + A$, where M is a local martingale and A is of finite variation, such that $(H^2\bullet[M])^{1/2}$ is locally P-integrable and $H(\omega)$ is almost surely $|dA(\omega)|$ integrable. Under these conditions, we defined the semimartingale $H\bullet X$ and called this the stochastic integral of H with respect to X. We seek to prove the following theorem (Theorem 12.3.22 in the main text).

Theorem A.6.1. *The class of X-integrable processes is a vector space, equal to the largest class (denoted $L'(X)$) of predictable processes H such that one can define a bilinear map $\mathfrak{I}$, with $\mathfrak{I}(H,X)$ defined for all semimartingales X and all $H \in L'(X)$, which satisfies*

(i) $\mathfrak{I}(H,X)$ *is a semimartingale,*
(ii) $\mathfrak{I}(H,X)^c = H\bullet X^c$, *where X^c is the continuous martingale part of X, (with $\bullet$ denoting the Itô integral of Theorem 12.3.3 with respect to a continuous local martingale)*
(iii) $H\Delta X = \Delta\mathfrak{I}(H,X)$ *and*
(iv) $\mathfrak{I}(H,X) = H\bullet X$ *whenever X has finite variation and H is $|dX|$-integrable, in the sense of Stieltjes integrals.*

Furthermore, $\mathfrak{I}$ is uniquely defined by these properties and is given by the stochastic integral $\mathfrak{I}(H,X) = H\bullet X$ as defined in Definition 12.3.10.

Recall that $\mathfrak{I}$ is bilinear only up to evanescent sets, which may depend on the arguments.

The problem we face is that Definition 12.3.10 depends on the abstract statement 'there exists a decomposition', and this decomposition may differ for different integrands. To deal with this, we give a proof adapted from Jacod [107, Chapter 2f], which depends on a careful consideration of the jumps of a semimartingale. We consider decomposing the jumps into two sets, one of which is summable, the other of which is square summable. By looking at the possible decompositions with this property, we see that for any finite collection of integrands we can find a common decomposition of X with which we can define the integral simultaneously. This fact can then be used to restrict the class of possible integrands to the X-integrable processes in Definition 12.3.10.

In order to analyse the jumps of X, we use the following notation.

Definition A.6.2. *A measurable process Y is called a* thin process *if it is zero except on a thin set (see Definition 7.5.1). For a thin process, we define $\mathfrak{S}(Y)$ and $\mathfrak{T}(Y)$ by*

$$\mathfrak{S}(Y)_t = \sum_{u\leq t} Y_u, \qquad \mathfrak{T}(Y) = \mathfrak{S}(Y^2)^{1/2}$$

provided these sums are absolutely convergent.

Remark A.6.3. Note that, if $\mathfrak{S}(Y) \in \mathcal{A}_{\mathrm{loc}}$, then it follows that $\mathfrak{T}(Y) \in \mathcal{A}_{\mathrm{loc}}$. Also, from Exercise 7.7.1, for any càdlàg adapted process X, the jump process $Y = \Delta X$ is thin. Furthermore, if X is a semimartingale, then $\mathfrak{T}(\Delta X)$ is well defined, by Lemma 10.3.6.

Lemma A.6.4. *Let Y be a thin process. Then*

(i) $\mathfrak{S}(Y) \in \mathcal{A}$ *implies* $\mathfrak{S}(\Pi_p Y) \in \mathcal{A}$,
(ii) $\mathfrak{T}(Y) \in \mathcal{A}_{\mathrm{loc}}$ *implies* $\mathfrak{T}(\Pi_p Y) \in \mathcal{A}_{\mathrm{loc}}$.

Proof. (i) The set $\{Y \neq 0\}$ is a thin set, and $\Pi_p Y$ is zero outside the predictable projection of this set. Decomposing $\{Y \neq 0\}$ into a union of the graphs of accessible and totally inaccessible stopping times $\{Y \neq 0\} = \bigcup_n (\llbracket T_n^a\rrbracket \cup \llbracket T_n^i \rrbracket)$, we see that $\{\Pi_p Y \neq 0\} \subseteq \bigcup_n \llbracket T_n^a \rrbracket$, so $\Pi_p Y$ is a predictable thin process. However, for any predictable stopping time T, from Theorem 7.6.5 we know

$$|\Pi_p Y|_T I_{\{T<\infty\}} \leq E\big[|Y_T| I_{\{T<\infty\}}\,\big|\mathcal{F}_{T-}\big],$$

and it follows that $\mathfrak{S}(\Pi_p Y) \in \mathcal{A}$.

(ii) Let $X = Y I_{\{|Y|\leq 1\}}$ and $Z = Y - X = Y I_{\{|Y|>1\}}$. By Lemma 10.3.5 we know that $\mathfrak{S}(Z) \in \mathcal{A}_{\mathrm{loc}}$, and hence $\mathfrak{S}(\Pi_p Z) \in \mathcal{A}_{\mathrm{loc}}$ by (i). Applying Lemma 10.3.5 again we see that $\mathfrak{T}(\Pi_p Z) \in \mathcal{A}_{\mathrm{loc}}$.

As in Lemma 2.4.11, we can prove Jensen's inequality for the order preserving projection Π_p, in particular we see that $(\Pi_p X)^2 \leq \Pi_p(X^2)$. We know $\mathfrak{T}(X) \leq \mathfrak{T}(Y) \in \mathcal{A}(\mathrm{loc})$, so X is thin and, as X is bounded, we know $\mathfrak{S}(X^2) \in \mathcal{A}_{\mathrm{loc}}$. From part (i) we see $\mathfrak{S}((\Pi_p X)^2) \in \mathcal{A}_{\mathrm{loc}}$, and hence $\mathfrak{T}(\Pi_p X) \in \mathcal{A}_{\mathrm{loc}}$. The result follows from the inequality

$$\mathfrak{T}(\Pi_p Y) \leq \sqrt{2}\big(\mathfrak{T}(\Pi_p X) + \mathfrak{T}(\Pi_p Z)\big).$$

□

If X is a semimartingale, we need to consider how to remove problematically large (in particular, non-integrable) jumps, so we can deal with them separately. For any optional set $D \in \Sigma_o$ we define

$$X^{(D)} := X - \mathfrak{S}(I_D \Delta X),$$

so $X^{(D)}$ is X with the jumps occurring in D removed. Consider the family of optional sets $\mathcal{D}(X)$ defined by

$$\mathcal{D}(X) := \{D \in \Sigma_o : \mathfrak{S}(I_D \Delta X) \in \mathcal{V}, X^{(D)} \in \mathcal{S}_{\mathrm{Sp}}\},$$

where $\mathcal{S}_{\mathrm{Sp}}$ denotes the special semimartingales (Definition 11.6.9). From Theorem 11.6.10, we know that $\mathcal{D}(X)$ is nonempty, as it contains the sets $\{|\Delta X| > \epsilon\}$ for any $\epsilon > 0$. Clearly, X is a special semimartingale if and only if $\mathcal{D}(X)$ contains the empty set.

For $X \in \mathcal{S}_{\mathrm{Sp}}$, let $X = M + A$ be the canonical decomposition into a local martingale and predictable finite variation process, and define

$$\hat{X} = A - \mathfrak{S}(\Delta A), \tag{A.2}$$

so that $\hat{X}$ is a continuous finite variation process.

We now define a related set, which we shall use to define $\mathcal{D}(X)$ using only the jumps of X. For any thin process Y, define

$$\begin{aligned}\mathcal{D}'(Y) := \big\{D \in \Sigma_o : \mathfrak{S}(Y I_D) \in \mathcal{V}, \mathfrak{S}(Y^2 I_{D^c})^{1/2} \in \mathcal{A}_{\mathrm{loc}}, \\ \mathfrak{S}(\Pi_p(Y I_{D^c})) \in \mathcal{A}_{\mathrm{loc}}\big\}.\end{aligned}$$

Theorem A.6.5. *(i) Let Y be a thin process. There exists a semimartingale X such that $Y = \Delta X$ if and only if $\mathcal{D}'(Y) \neq \emptyset$, in which case $\mathcal{D}'(Y) = \mathcal{D}(X)$.*

(ii) Let Y be a thin process with $\mathcal{D}'(Y) \neq \emptyset$. Let $D \in \mathcal{D}'(Y)$, N be a continuous local martingale and A be a continuous finite variation process, with $A_0 = N_0 = 0$. Then there exists a unique semimartingale X such that $X^c = N$, $\Delta X = Y$ and $\hat{X}^{(D)} = A$, namely

$$X = \underbrace{N + M}_{\in \mathcal{M}_{\mathrm{loc}}} + \underbrace{\mathfrak{S}(\Pi_p(Y I_{D^c})) + A + \mathfrak{S}(Y I_D)}_{\in \mathcal{V}}$$

where M is the unique purely discontinuous local martingale with $M_{0-} = 0$ and $\Delta M = Y I_{D^c} - \Pi_p(Y I_{D^c})$.

Here $\hat{X}^{(D)}$ denotes the process defined in (A.2) for the special semimartingale $X^{(D)}$.

Proof. We prove the theorem in four steps.

(X exists $\Rightarrow \mathcal{D}'(Y) \supseteq \mathcal{D}(X)$ and $\mathcal{D}'(Y) \neq \emptyset$) Let $X \in \mathcal{S}$, $D \in \mathcal{D}(X)$ and $Y = \Delta X$. By definition, $\mathfrak{S}(Y I_D) \in \mathcal{V}$. Now let $X^{(D)}$ have canonical decomposition $X^{(D)} = M + A$, so that $Y I_{D^c} = \Delta X^{(D)} = \Delta M + \Delta A$. From

Lemma 11.4.6, we know that $\mathfrak{T}(\Delta M) \in A \in \mathcal{A}_{\mathrm{loc}}$ and, as $A \in \mathcal{A}_{\mathrm{loc}}$, we know that $\mathfrak{T}(\Delta A) \in \mathcal{A}_{\mathrm{loc}}$. We have the general inequality

$$\mathfrak{T}(\Delta M + \Delta A) \leq \sqrt{2}\big(\mathfrak{T}(\Delta M) + \mathfrak{T}(\Delta A)\big),$$

hence $\mathfrak{T}(YI_{D^c}) \in \mathcal{A}_{\mathrm{loc}}$. Finally, we note that $\Delta A = \Pi_p(YI_{D^c})$, so $\mathfrak{T}(\Pi_p(YI_{D^c})) \in \mathcal{A}_{\mathrm{loc}}$, which implies $D \in \mathcal{D}'(Y)$.

(X exists $\Rightarrow \mathcal{D}'(Y) \subseteq \mathcal{D}(X)$) Let $X \in \mathcal{S}$, $Y = \Delta X$ and $D \in \mathcal{D}'(Y)$. Then we have $\mathfrak{S}(YI_D) = \mathfrak{S}(\Delta X I_D) \in \mathcal{V}$. Consequently, $X^{(D)} = X - \mathfrak{S}(\Delta X I_D)$ is a semimartingale. It is easy to see that

$$(X^{(D)})^*_t \leq \sup_{s\leq t}\{X^{(D)}_{s-}\} + \mathfrak{T}(YI_{D^c})_t.$$

As $\sup_{s\leq t}\{X^{(D)}_{s-}\}$ is left continuous, it is locally bounded and, as $\mathfrak{T}(YI_{D^c}) \in \mathcal{A}_{\mathrm{loc}}$ by hypothesis, we see that $(X^{(D)})^* \in \mathcal{A}_{\mathrm{loc}}$. Theorem 11.6.10 then implies that $X \in \mathcal{S}_{\mathrm{Sp}}$, and we can conclude that $D \in \mathcal{D}(X)$.

($\mathcal{D}'(Y) \neq \emptyset \Rightarrow X$ is well defined, satisfies the statement of the theorem, and hence exists) Let Y, D, N and A be as stated in (ii). Let $Z = YI_{D^c}$. As $\mathfrak{T}(Z) \in \mathcal{A}_{\mathrm{loc}}$, we know $\mathfrak{T}(Z - \Pi_p Z) \in \mathcal{A}_{\mathrm{loc}}$ by Lemma A.6.4. Hence, by Theorem 11.5.11, there exists a purely discontinuous local martingale M with $\Delta M = Z - \Pi_p Z$. Therefore, we can define X by the formula of the theorem and, as $X \in \mathcal{S}$, $X^c = N$ and $\Delta X = Y$, we see that $D \in \mathcal{D}(X)$. It follows that $X^{(D)} = N + M + \mathfrak{S}(Z) + A$ is in $\mathcal{S}_{\mathrm{Sp}}$ and consequently $\hat{X}^{(D)} = A$.

(X is unique) Suppose $\tilde{X}$ is another semimartingale with $\tilde{X}^c = N$, $\Delta\tilde{X} = Y$ and $\hat{\tilde{X}}^{(D)} = A$. Then we see that $\mathcal{D}(\tilde{X}) = \mathcal{D}'(Y) = \mathcal{D}(X)$; the uniqueness follows from the uniqueness of the canonical decomposition of a special semimartingale (Theorem 11.6.10). □

For a sequence of thin processes $\{Y^i\}_{i\in\mathbb{N}}$, the following lemma shows us there are sets in the intersection of the $\mathcal{D}'(Y^i)$, provided each $\mathcal{D}'(Y^i)$ is not empty. In terms of stochastic integrals, this allows us to decompose a semimartingale simultaneously for a finite collection of integrands, so we can fully exploit the bilinearity of the integral.

Lemma A.6.6. *Let $\{Y^i\}_{i\in\mathbb{N}}$ be a sequence of thin processes with $\mathcal{D}'(Y^i) \neq \emptyset$ for all i. Then, for any $a > 0$ and any $n \in \mathbb{N}$, we have*

$$\bigcup_{i\leq n}\{|Y^i| > a\} \in \bigcap_{i\leq n}\mathcal{D}'(Y^i).$$

Proof. As $\mathcal{D}'(Y^i) \neq \emptyset$, from Theorems 11.6.10 and A.6.5 we know $D^i := \{|Y^i| > a\} \in \mathcal{D}'(Y^i)$ for all $a > 0$. Define $D = \cup_{i\leq n} D^i$, then, for any $t > 0$, the sections of $D \cap [\![0,t]\!]$ contain at most finitely many points (as semimartingales have almost surely finitely many large jumps on the interval $[0,t]$) and therefore $\mathfrak{S}(Y^i I_D) \in \mathcal{V}$. As $D^i \subseteq D$, we see that $D \in \mathcal{D}'(Y^i)$, which yields the result. □

We can now see how this restricts the class of possible integrands.

Theorem A.6.7. *The set $L'(X)$ defined by Theorem A.6.1 is the class of predictable processes H such that $(H^2 \bullet \langle X^c \rangle)^{1/2} \in \mathcal{A}_{\mathrm{loc}}$ and there exists a set $D \in \mathcal{D}(X) \cap \mathcal{D}'(H\Delta X)$ such that*

$$\Big\{ \int_{[0,t]} H_s(\omega) d\hat{X}_s^{(D)}(\omega) \Big\}_{t\geq 0} \in \mathcal{V}.$$

(Here $\hat{X}^{(D)}$ denotes the process defined in (A.2) for the special semimartingale $X^{(D)}$.)

Furthermore, $L'(X)$ is equal to $L(X)$, the class of X-integrable processes, as in Definition 12.3.10.

Proof. As the stochastic integral satisfies the requirements of Theorem A.6.1 whenever it is well defined, it is clear that $L(X) \subseteq L'(X)$. Suppose we have some $H \in L'(X)$. We shall show that H has the stated properties, and that these imply $H \in L(X)$. We first show that $(H^2 \bullet \langle X^c \rangle)^{1/2} \in \mathcal{A}^{\mathrm{loc}}$. As $\mathfrak{I}(H,X)$ is a semimartingale, we know $\langle \mathfrak{I}(H,X)^c \rangle$ exists and is a continuous process, and therefore is locally integrable. Theorem A.6.1(ii) states $\mathfrak{I}(H,X)^c = H \bullet X^c$, and therefore, applying Theorem 12.2.1 locally, we have

$$\langle \mathfrak{I}(H,X)^c \rangle = H \bullet \langle X^c, \mathfrak{I}(H,X)^c \rangle = H^2 \bullet \langle X^c \rangle$$

the integrals on the right being Stieltjes integrals. The stated integrability follows.

We now show the existence of the set D. As both X and $\mathfrak{I}(H,X)$ are semimartingales and $\Delta(\mathfrak{I}(H,X)) = H\Delta X$ by Theorem A.6.1(iii), we know from Theorem A.6.5 that $\mathcal{D}(X) = \mathcal{D}'(\Delta X)$ and $\mathcal{D}(\mathfrak{I}(H,X)) = \mathcal{D}'(H\Delta X)$ are nonempty. By Lemma A.6.6, this implies that, for any $a > 0$,

$$D := \{|\Delta X| > a\} \cup \{|H\Delta X| > a\} \in \mathcal{D}(X) \cap \mathcal{D}'(H\Delta X).$$

As $D \in \mathcal{D}(X)$ we know that $\hat{X}^{(D)}$ is well defined and is a continuous finite variation process. On the other hand, as $D \in \mathcal{D}'(H\Delta X)$, by linearity and the uniqueness of the decomposition in Theorem A.6.5(ii), defining $Y = \mathfrak{I}(H,X)$ we have $\hat{Y}^{(D)} = H \bullet \hat{X}^{(D)} \in \mathcal{V}$, that is H is locally $|d\hat{X}^{(D)}|$-integrable. It follows that D is as required in the theorem, and hence all $H \in L'(X)$ have the stated properties.

Now suppose we are given an H with the stated properties. Using the decomposition in Theorem A.6.5(ii), we can write

$$X = \big(X^c + M\big) + \big(\mathfrak{S}(\Pi_p(YI_{D^c})) + A + \mathfrak{S}(YI_D)\big).$$

As $D \in \mathcal{D}'(\Delta X) \cap \mathcal{D}'(H\Delta X)$, we know that

$$\begin{aligned}\big(H^2 \bullet [X^c + M]\big)^{1/2} &\leq \sqrt{2}\Big(\big(H^2 \bullet \langle X^c \rangle\big)^{1/2} + \mathfrak{T}\big(H(\Delta X I_{D^c} - \Pi_p(\Delta X I_{D^c}))\big)\Big) \\ &\in \mathcal{A}_{\mathrm{loc}}\end{aligned}$$

and

$$H \bullet \big(\mathfrak{S}(\Pi_p(\Delta X I_{D^c})) + \hat{X}^{(D)} + \mathfrak{S}(\Delta X I_D)\big) \in \mathcal{V}$$

(these being Stieltjes integrals). Therefore, we have a decomposition of X satisfying the requirements of Definition 12.3.10, and we see that $H \in L(X)$. □

Recall, from Theorem 12.3.13, that the stochastic integral as defined by Definition 12.3.10 does not depend on the semimartingale decomposition chosen.

Corollary A.6.8. *The stochastic integral defined by Definition 12.3.10 is bilinear and $L(X) = L'(X)$ is a vector space.*

Proof. Suppose $H, H' \in L(X)$. Then $H \bullet X + H' \bullet X$ is a semimartingale. Also, by Lemma A.6.6 there exists a set $D \in \mathcal{D}(X) \cap \mathcal{D}'(H\Delta X) \cap \mathcal{D}'(H'\Delta X)$ such that, using the decomposition of Theorem A.6.5(ii),

$$\big((H+H')^2 \bullet [X^c+M]\big)^{1/2} \leq \sqrt{2}\big((H^2 \bullet [X^c+M])^{1/2} + (H'^2 \bullet [X^c+M])^{1/2}\big) \in \mathcal{A}_{\mathrm{loc}}$$

and

$$(H + H') \bullet \big(\mathfrak{S}(\Pi_p(\Delta X I_{D^c})) + \hat{X}^{(D)} + \mathfrak{S}(\Delta X I_D)\big) \in \mathcal{V}.$$

Therefore, $H + H'$ is also X-integrable, and linearity follows from the linearity of the stochastic integral with respect to a local martingale and of the Stieltjes integral, using this decomposition (cf. Corollary 12.3.21).

Similarly, if X and Y are semimartingales and $H \in L(X) \cap L(Y)$, then by decomposing using a set in $\mathcal{D}(X) \cap \mathcal{D}(Y) \cap \mathcal{D}'(H\Delta X) \cap \mathcal{D}'(H\Delta Y)$ we can see that $H \in L(X + Y)$, and that the desired linearity also holds. □

Corollary A.6.9. *The bilinear map $\mathfrak{I}$ satisfying the conditions of Theorem A.6.1 is unique and so agrees with the stochastic integral defined in Definition 12.3.10.*

Proof. From Theorem A.6.7, we know that we can find a set D in $\mathcal{D}(X) \cap \mathcal{D}'(H\Delta X)$.

Given this set D, take the decomposition of X in Theorem A.6.5. As $(H^2 \bullet [X^c + M])^{1/2} \in \mathcal{A}_{\mathrm{loc}}$ and $H \bullet (X - X^c - M) \in \mathcal{V}$, from the conditions of Theorem A.6.1 and the uniqueness of Exercise 11.7.12 we see that for any bilinear map $\mathfrak{I}$ satisfying the conditions of Theorem A.6.1,

$$\begin{aligned}\mathfrak{I}(H, X) &= \mathfrak{I}(H, X^c) + \mathfrak{I}(H, M) + \mathfrak{I}(H, X - X^c - M)\\ &= H \bullet X^c + H \bullet M + H \bullet (X - X^c - M)\\ &= H \bullet X.\end{aligned}$$

By Theorem 12.3.13, the stochastic integral of Definition 12.3.10 is independent of the decomposition chosen (and hence of the set D we choose in $\mathcal{D}(X) \cap \mathcal{D}'(H\Delta X)$), so $\mathfrak{I}$ is uniquely determined. □

A.6.2 The Bichteler–Dellacherie–Mokobodzki Theorem

In this section, we prove a result, due to Bichteler, Dellacherie and Mokobodzki, which characterizes semimartingales in terms of stochastic integrals. We here present a recent simple proof due to Beiglböck and Siorpaes [7]. We seek to prove the following (Theorem 12.3.26), which we restate for convenience.

Theorem A.6.10 (Bichteler–Dellacherie–Mokobodzki Theorem). *A càdlàg adapted process X is a good integrator (Definition 12.3.24) if and only if it is a semimartingale (Definition 11.6.1).*

We shall prove this in a series of results. The first direction of the theorem is easy.

Lemma A.6.11. *Let X be a semimartingale and H^n a sequence in Λ converging uniformly to a process H. Then $H^n \bullet X$ converges uniformly on compacts in probability (ucp) to $H \bullet X$. In particular, any semimartingale is a good integrator.*

Proof. The semimartingale X has decomposition $X = X_0 + A + M$, where $A \in \mathcal{V}_0$, $M \in \mathcal{M}_{0,\mathrm{loc}}$, and the integrals are given by

$$H^n \bullet X = H^n_0 X_0 + H^n \bullet A + H^n \bullet M.$$

First note that $H^n_0 X_0$ converges to $H_0 X_0$ uniformly. For the martingale part of the integral, fix $\epsilon > 0$ and let T be a stopping time such that M_T is integrable and $P(T < t) < \epsilon$ (such a stopping time exists by the result of Exercise 3.4.16). From Theorem 12.3.8,

$$\|I_{\{T \geq t\}}((H^n - H) \bullet M)^*_t\|_1 \leq C \sup_{s \leq t} \|H^n_s - H_s\|_\infty \|M^*_{T \wedge t}\|_1$$

which can be made arbitrarily small, as H^n converges uniformly to H. By Markov's inequality, for any $\epsilon > 0$ we have $P\big(((H^n - H) \bullet M)^*_t > \epsilon\big) \to 0$ as $n \to \infty$, so the martingale part of the integral converges uniformly on compacts in probability.

As A has paths of finite variation, we know that

$$((H^n - H) \bullet A)^*_t \leq \int_{[0,t]} |H^n_s - H_s||dA|_s \leq \sup_{s \leq t} \|H^n_s - H_s\|_\infty \int_{[0,t]} |dA|_s.$$

Therefore, $H^n \bullet A \to H \bullet A$ uniformly on compacts a.s. (and hence in probability). □

Definition A.6.12. *Let $\pi = \{0 \leq t_0 < t_1 < \ldots < t_n\}$ be an increasing sequence of deterministic times, and X a process with $X_t \in L^1$ for all $t \in [0, \infty[$. The mean variation of X on the partition π is defined by*

$$\mathrm{MV}(X, \pi) = E\Big[|X_0| + \sum_{i=1}^n \big|E[X_{t_i} - X_{t_{i-1}} | \mathcal{F}_{t_i}]\big|\Big].$$

A quasimartingale *is an adapted process X for which there exists a constant C such that the mean variation* $\mathrm{MV}(X) := \sup_\pi \mathrm{MV}(X,\pi) < C$*, the supremum being taken over all finite deterministic partitions π. A local quasimartingale is defined in the usual way.*

Theorem A.6.13 (Rao's Quasimartingale Decomposition). *A càdlàg process X is a local quasimartingale if and only if it has a decomposition $X = Y - Z$, where Y and Z are càdlàg local submartingales. Hence every càdlàg local quasimartingale is a semimartingale.*

Proof. To show that all local quasimartingales have the desired representation, first localize, so we can assume that X is a quasimartingale. Fix $T > 0$ and $n \in \mathbb{N}$ and, for notational convenience, define $s_i = i2^{-n}T$ for $i \in \{0, 1, .., 2^n\}$. Let

$$A_t^n = X_0^+ + \sum_{i=0}^{2^n-1} I_{\{s_{i+1}>t\}} E[X_{s_{i+1}} - X_{s_i}|\mathcal{F}_{s_i}]^+,$$

$$B_t^n = X_0^- + \sum_{i=0}^{2^n-1} I_{\{s_{i+1}>t\}} E[X_{s_{i+1}} - X_{s_i}|\mathcal{F}_{s_i}]^-.$$

As X is a quasimartingale, we know A^n and B^n are integrable. Then let $Y_t^n = E[A_t^n|\mathcal{F}_t]$ and $Z_t^n = E[B_t^n|\mathcal{F}_t]$ for $t \in [0, T]$. Clearly, we have $X_{s_i} = Y_{s_i}^n - Z_{s_i}^n$ for any $i \leq 2^n$. Now note that Y^n and Z^n are increasing in n, by Jensen's inequality. Therefore the limits $Y_t = \lim_{n\to\infty} Y_t^n$ and $Z_t = \lim_{n\to\infty} Z_t^n$ exist in L^1.

It is straightforward to verify, again by Jensen's inequality, that, for any $s \leq t$,

$$Y_s = \sup_n Y_s^n \leq \sup_n E[Y_t^n|\mathcal{F}_s] \leq E\big[\sup_n Y_t^n\big|\mathcal{F}_s\big] = E[Y_t|\mathcal{F}_s].$$

Therefore, Y is a submartingale (but not necessarily càdlàg) and similarly for Z. By Theorem 5.1.8, we define càdlàg processes $\tilde{Y} := \{Y_{t+}\}_{t\geq 0}$ and $\tilde{Z} := \{Z_{t+}\}_{t\geq 0}$, which are also submartingales. As X is càdlàg and our filtration is right-continuous, we see that these will also satisfy $X = \tilde{Y} - \tilde{Z}$. Therefore, X has the required representation on $[0, T]$ and the result follows by pasting as in Lemma 11.6.3.

To show the converse, observe that if Y and Z are càdlàg local submartingales then they have a Doob–Meyer decomposition $Y = M + B$, $Z = N + C$, where B and C are locally integrable increasing processes (Theorem 9.2.7 applied locally to $-Y$ and $-Z$). Then $X = Y - Z = (M - N) + (B - C)$ satisfies, for any sequence $t_0 < t_1 < \ldots, t_n$,

$$\begin{aligned}\sum_{i=1}^{n} \big|E[X_{t_i} - X_{t_{i-1}}|\mathcal{F}_{t_i}]\big| &\leq \sum_{i=1}^{n} \Big(E\big[|B_{t_i} - B_{t_{i-1}}|\,\big|\mathcal{F}_{t_i}\big] + E\big[|C_{t_i} - C_{t_{i-1}}|\,\big|\mathcal{F}_{t_i}\big]\Big)\\ &= E\big[B_{t_n}|\mathcal{F}_{t_i}\big] + E\big[C_{t_n}|\mathcal{F}_{t_i}\big].\end{aligned}$$

As the processes B and C are locally integrable, this guarantees that X is a local quasimartingale. □

We now prove a useful result on convex combinations of stopping times.

Lemma A.6.14. *Fix $t > 0$ and let $\{T_n\}_{n\in\mathbb{N}}$ be a sequence of stopping times such that, for some $\epsilon > 0$, we have $P(T_n \geq t) \geq 1 - \epsilon$ for all n. Then there exists a stopping time T and, for each n, a value N_n and convex weights $w_1^n, \dots, w_{N_n}^n$ (i.e. $w_i^n \geq 0$ for all n and i, and $\sum_{i=1}^{N_n} w_i^n{=}1$) such that $P(T \geq t) \geq 1 - 3\epsilon$ and, for all n sufficiently large,*

$$I_{[0,T]} \leq 2\sum_{k=n}^{N_n} w_k^n I_{[0,T_k]}.$$

Proof. As a consequence of Mazur's lemma (Lemma 1.5.15) and the fact that a set bounded in L^2 is weakly compact (Theorem 1.7.19), for any L^2-bounded sequence $\{x_n\}_{n\in\mathbb{N}}$ there exist vectors $\{y_n\}_{n\in\mathbb{N}}$ and x such that $\|y_n - x\|_2 \to 0$ and $y_n = \sum_{i=n}^{N_n} w_i^n x_n$ for some convex weights $\{w_i^n\}_{i=n}^{N_n}$.

We apply this to the random variables $X_n = I_{\{T_n \geq t\}}$, to obtain weights $\{w_i^n\}_{i=n}^{N_n}$ such that

$$Y_n = \sum_{i=n}^{N_n} w_i^n X_i \to X,$$

the convergence being in L^2. By taking a subsequence (Lemma 1.3.38), we can assume that $Y_n \to X$ a.s.

As $X \leq 1$ and $E[X] \geq 1 - \epsilon$, we deduce that $P(\lim_n Y_n = X \leq 2/3) < 3\epsilon$. Using Egorov's theorem (Theorem 1.3.36), we see that there is a set A with $P(A) \geq 1 - 3\epsilon$ such that $Y_n \geq 1/2$ on A, for all $n \geq N$, for some N. We now define

$$T = \inf_{n\geq N} \inf\Big\{s : \sum_{i=n}^{N_n} \big(w_i^n I_{\{s\in[0,T_n]\}}\big) < 1/2\Big\}.$$

We clearly have $I_{[0,T]} \leq 2\sum_{k=n}^{N_n} w_k^n I_{[0,T_k]}$ and, as $A \subseteq \{T \geq t\}$, we see $P(T \geq t) \geq 1 - 3\epsilon$. ☐

Lemma A.6.15. *Let X be a process bounded uniformly by some $K > 0$. Given a finite deterministic partition π, as in Definition A.6.12, and a stopping time T, define $T_\pi = \inf\{t \in \pi : T \leq t\}$. Then*

$$\mathrm{MV}(X^{T_\pi}, \pi) = E\Big[\sum_{t_i\in\pi} I_{\{t_i<T\}}\big|E[X_{t_{i+1}} - X_{t_i}|\mathcal{F}_{t_i}]\big|\Big]$$

and

$$|\mathrm{MV}(X^{T_\pi}, \pi) - \mathrm{MV}(X^T, \pi)| \leq 2K.$$

Proof. For each $t_i \in \pi$, note that

$$E[X^{T_\pi}_{t_{i+1}} - X^{T_\pi}_{t_i}|\mathcal{F}_{t_i}] = I_{\{t_i<T\}}E[X_{t_{i+1}} - X_{t_i}|\mathcal{F}_{t_i}],$$

which yields the first statement. Furthermore, Jensen's inequality implies

$$|\mathrm{MV}(X^{T_\pi},\pi)-\mathrm{MV}(X^T,\pi)|\le E\Big[\sum_{t_i\in\pi}\big|(X^T_{t_{i+1}}-X^T_{t_i})-(X^{T_\pi}_{t_{i+1}}-X^{T_\pi}_{t_i})\big|\Big].$$

This sum has at most one nonzero term, when $T\in[t_i,t_{i+1}[$, so we have the desired bound. □

Lemma A.6.16. *Any locally bounded good integrator X is a local quasimartingale.*

Proof. By localization, we can assume that X is bounded. Fix $t>0$, and let K be a bound on $|X|$. As X is a good integrator, for any $\epsilon>0$ there exists $C>0$ such that, for any $H\in\Lambda$ with $\sup_s\|H_s\|_\infty\le 1$, we have $P\big((H\bullet X)_t\ge C-2K\big)<\epsilon$. Let $D_n=\{t_i=i2^{-n}t: i\in\mathbb{N}, i\le 2^n\}$. For each n, define

$$H^n:=\sum_{s_i\in D_n} I_{]s_i,s_{i+1}]}\mathrm{sign}(E[X_{s_{i+1}}-X_{s_i}|\mathcal{F}_{s_i}])$$

and

$$T_n=\inf\{s\in D_n:(H^n\bullet X)_s\ge C-2K\}.$$

On the set $\{T_n<t\}$ we have $\big((H^nI_{[0,T_n]})\bullet S\big)_{T_n}=(H^n\bullet S)_{T_n}\ge C-2K$, so $P(T_n\ge t)\ge 1-\epsilon$. Moreover, since $|\Delta X_s|=|X_s-X_{s-}|\le 2K$, we know

$$C\ge E[(H^n\bullet S)^{T_n}_t]=E\Big[\sum_{s_i\in D_n}I_{\{s_i<T_n\}}\big|E[X_{s_{i+1}}-X_{s_i}|\mathcal{F}_{t_i}]\big|\Big]=\mathrm{MV}(X^{T_n},D_n).$$

Now let T be defined by applying Lemma A.6.14 to the sequence $\{T_n\}_{n\in\mathbb{N}}$. For n sufficiently large, we have, in the notation of Lemma A.6.15,

$$\begin{aligned}\mathrm{MV}(X^{T_{D_n}},D_n)&=E\Big[\sum_{s_i\in D_n}I_{\{s_i<T\}}\big|E[X_{s_{i+1}}-X_{s_i}|\mathcal{F}_{t_i}]\big|\Big]\\&\le 2E\Big[\sum_{s_i\in D_n}\sum_{k=n}^{N_n}w^n_kI_{\{s_i<T_k\}}\big|E[X_{s_{i+1}}-X_{s_i}|\mathcal{F}_{t_i}]\big|\Big]\\&=2\sum_{k=n}^{N_n}w^n_k\,\mathrm{MV}(X^{(T_k)_{D_n}},D_n).\end{aligned}$$

By Lemma A.6.15, it follows that

$$\begin{aligned}\mathrm{MV}(X^T,D_n)&\le\mathrm{MV}(X^{T_{D_n}},D_n)+2K\\&\le 2\Big(\sum_{k=n}^{N_n}w^n_k\mathrm{MV}(X^{T_k},D_n)+2K\Big)+2K\\&\le 2C+6K.\end{aligned}$$

Taking $n\to\infty$, we see that the stopped process X^T is a quasimartingale. By Exercise 3.4.16, as t was arbitrary and $P(T\ge t)\ge 1-3\epsilon$, we see that X is a local quasimartingale. □

From this, we can prove the Bichteler–Dellacherie–Mokobodzki theorem.

Proof of Theorem 12.3.26/A.6.10. Lemma A.6.11 proves that any semimartingale is a good integrator, so only the converse remains. Let X be a càdlàg good integrator. From Exercise 3.4.8, we know that X almost surely has at most finitely many discontinuities of size at least 1, that is

$$J_t = \sum_{s\leq t} \Delta X_s I_{\{|\Delta X_s|\geq 1\}}$$

is a finite sum, for almost all ω and all t. Then J has finite variation for each ω, so J is a semimartingale and hence a good integrator (by Lemma A.6.11). By linearity it follows that $X - J$ is a good integrator with bounded jumps. Therefore, $X-J$ is a locally bounded good integrator, and hence a local quasimartingale by Lemma A.6.16. By Rao's decomposition (Theorem A.6.13), it follows that X is a semimartingale. □

A.6.3 Completeness of Stochastic Integrals

In this section, we seek to prove Theorem 12.4.16, which we repeat here for the ease of the reader.

Theorem A.6.17. *For any semimartingale X, the space $\{H\bullet X\}_{H\in L(X)}$ is complete in the semimartingale topology.*

Proof. Unlike Mémin [132], we shall use a prelocalization argument, based on Lemma 16.2.7. Suppose we have a semimartingale X and a sequence $\{H^n\}_{n\in\mathbb{N}} \subset L(X)$ of integrands such that $\{H^n\bullet X\}_{n\in\mathbb{N}}$ converges in the semimartingale topology. We wish to show that there exists $H \in L(X)$ such that $H^n\bullet X \to H\bullet X$.

First suppose that $X \in \mathcal{H}^2_{\mathcal{S}}$ and $H^n\bullet X$ converges in $\mathcal{H}^2_{\mathcal{S}}$. Then, taking the canonical decomposition of the special semimartingale $X = M + A$, we know that $H^n\bullet X$ has canonical decomposition

$$H^n\bullet X = H^n\bullet M + H^n\bullet A.$$

It follows from Lemma 16.2.5 that each of the martingale and finite variation terms must converge, in $\mathcal{H}^2$ and $\mathcal{A}^2$ respectively. Therefore, we must have

$$\sup_{n'>n} E\Big[\int_{[0,\infty]} (H^n - H^{n'})^2 d[M] + \Big(\int_{[0,\infty]} |H^n - H^{n'}|\,|dA|\Big)^2\Big] \to 0$$

as $n \to 0$. From Markov's inequality, we see that the limit H must exist pointwise $(d[M] + |dA|) \times dP$-almost everywhere, at least for a subsequence, and $H^n\bullet X \to H\bullet X$ in $\mathcal{H}^2_{\mathcal{S}}$.

For the original problem with convergence in $\mathcal{S}$, by Lemma 16.2.7, there exists a sequence of stopping times $\{T_m\}_{m\in\mathbb{N}}$ such that, for every m, $\{(H^n\bullet X)^{T_m-}\}_{n\in\mathbb{N}}$ and X^{T_m-} are all in $\mathcal{H}^2_{\mathcal{S}}$, and $(H^n\bullet X)^{T_m-}$ converges in $\mathcal{H}^2_{\mathcal{S}}$. By our earlier argument, we can therefore construct $\{H^{(m)}\}_{m\in\mathbb{N}}$ such that $(H^n \bullet X)^{T_m-} \to (H^{(m)} \bullet X)^{T_m-}$ in $\mathcal{H}^2_{\mathcal{S}}$. By Lemma 16.2.7 and a simple pasting argument, we construct H such that $H^n \bullet X \to H \bullet X$ in $\mathcal{S}$, as desired. $\square$

The case where X is a vector semimartingale and we consider the vector integral (Theorem 12.5.16) can be proven in essentially the same way, and is left as an exercise.

A.7 Novikov's Criterion with Jumps

In this appendix, we give reasonably general conditions, due to Lépingle and Mémin [125] (whose proof we follow), under which a stochastic exponential with jumps is a true martingale.

A.7.1 A Predictable Condition

We seek to prove Theorem 15.4.3, which we reproduce here for the reader's convenience.

Theorem A.7.1. *Suppose that X is a local martingale with $\Delta X \geq -1$ and let $T = \inf\{t : \Delta X_t = -1\} = \inf\{t : \mathcal{E}(X)_t = 0\}$. If the increasing process*

$$\frac{1}{2}\langle X^c\rangle_{t\wedge T} + \sum_{s\leq t\wedge T} \big((1+\Delta X_s)\log(1+\Delta X_s) - \Delta X_s\big)$$

has a predictable compensator B and $E[\exp(B_\infty)] < \infty$, then $\mathcal{E}(X)$ is a uniformly integrable martingale and $\{\mathcal{E}(X)_\infty > 0\} = \{T = \infty\}$ almost surely.

We shall use a series of lemmata to prove this result.

Lemma A.7.2. *Under the assumptions of Theorem A.7.1, X^T is an $\mathcal{H}^1$ martingale, and $\mathcal{E}(X)_{T-} > 0$ a.s.*

Proof. From the assumptions, we see that B_∞ is integrable, which implies that $\sum_{s\leq t\wedge T}\big((1+\Delta X_s)\log(1+\Delta X_s) - \Delta X_s\big)$ is integrable. Using the general inequality

$$(1+x)\log(1+x) - x \geq \begin{cases} x^2/6 & \text{if } |x| \leq 1,\\ x/3 & \text{if } x > 1,\end{cases}$$

this implies that

$$E\Big[\sum_{t\leq T} \Delta X_t^2 I_{\{\Delta X_T \leq 1\}}\Big] < \infty \quad \text{and} \quad E\Big[\sum_{t\leq T} \Delta X_t I_{\{\Delta X_T > 1\}}\Big] < \infty. \qquad \text{(A.3)}$$

Together with the fact that $E[\langle X^c\rangle_T] < \infty$, we see that we can decompose X^T into a martingale in $\mathcal{H}^2$ and a martingale of integrable variation (using the method of Theorem 10.3.4), and so X^T is in $\mathcal{H}^1$. To see that $\mathcal{E}(X)_{T-} > 0$, note that (A.3) implies that, prior to T, X has a.s. at most finitely many jumps with $|\Delta X| \geq 1/2$ and that there is a $k > 0$ such that

$$\sum_{\{t:|\Delta X_t|<1/2\}} \Big(\log(1+\Delta X_t) - \Delta X_t\Big) \geq \sum_{\{t:|\Delta X_t|<1/2\}} -k(\Delta X_t)^2.$$

As $[X]_t$ is almost surely finite, from (A.3), we see that $\prod_{\{t<T\}}(1 + \Delta X_t)e^{-\Delta X_t} > 0$. As X_{T-} is finite and $\langle X\rangle_{T-} < \infty$, we conclude that $\mathcal{E}(X)_{T-} > 0$. □

For notational convenience, we shall write

$$\hat{X}_t = \log(\mathcal{E}(X)_t) - B_t = X_t - \frac{1}{2}\langle X^c\rangle_t + \sum_{s\leq t}\big(\log(1+\Delta X_s) - \Delta X_s\big) - B_t,$$

with the convention $\hat{X} = -\infty$ whenever $\mathcal{E}(X) = 0$.

Lemma A.7.3. *Under the assumptions of Theorem A.7.1, $\mathcal{E}(X)\hat{X}$ is a local martingale.*

Proof. As B is an increasing predictable process and càdlàg, it is locally bounded (Lemma 7.3.20). Let $T_n = \inf\{t : \mathcal{E}(X)_t \leq 1/n\}$. From Lemma A.7.2 we see that $T_n \to T$ a.s. and $P(T_n \neq T) \to 0$. Let N be the local martingale defined by

$$N_t = \frac{1}{2}\langle X^c\rangle_{t\wedge T} + \sum_{s\leq t\wedge T}\big((1+\Delta X_s)\log(1+\Delta X_s) - \Delta X_s\big) - B_t.$$

For each $n \in \mathbb{N}$, let g_n be a C^2 function which is equal to $x\log x$ for all $x \geq 1/n$. Write $Y = \mathcal{E}(X)^{T_n}$. Then, applying Ito's formula,

$$\begin{aligned}
&g_n(Y_t) - Y_t B_{t\wedge T_n} - g_n(Y_0) + Y_0 B_0 \\
&= \int_{]0,t]} (\log(Y_{s-}) + 1) dY_s + \frac{1}{2}\int_{]0,t\wedge T_n]} Y_{s-} d\langle X^c\rangle_s \\
&\quad - \int_{]0,t]} B_s dY_S - \int_{[0,t\wedge T_n]} Y_{s-} dB_s \\
&\quad + \sum_{0<s\leq t\wedge T_n} \big(g_n(Y_s) - Y_{s-}\log(Y_{s-}) - (\log(Y_{s-}) + 1)Y_{s-}\Delta X_s\big).
\end{aligned}$$

We know that

$$\begin{aligned}
Y_t \log(Y_t) &= (1+\Delta X_t)Y_{t-}\log\big((1+\Delta X_t)Y_{t-}\big) \\
&= (1+\Delta X_t)Y_{t-}(\log(Y_{t-}) + \log(1+\Delta X_t))
\end{aligned}$$

so, on the set $\{t < T_n\}$,

$$Y_t \log(Y_t) - Y_t B_t = Y_0 \log(Y_0) - Y_0 B_0 + \int_{]0,t]} (\log(Y_{s-}) + 1 - B_s) dY_s \\ + \int_{]0,t]} Y_{s-} dN_s.$$

A straightforward calculation shows that both sides of this equation will have the same jump at time T_n, so the equality holds on the set $\{t \leq T_n\}$. Taking $n \to \infty$, we find that, on the set $\{t \leq T\}$,

$$\begin{aligned}\mathcal{E}(X)_t \hat{X}_t &= Y_t \log(Y_t) - Y_t B_t \\ &= Y_0 \log(Y_0) - Y_0 B_0 + \int_{]0,t]} (\log(Y_{s-}) + 1 - B_s) dY_s + \int_{]0,t]} Y_{s-} dN_s.\end{aligned}$$

The terms on the right are local martingales, as the integrands are locally bounded. On the set $\{t \geq T\}$, by definition $\mathcal{E}(X)_t \hat{X}_t$ is constant, and hence a local martingale, so the result is proven. □

Lemma A.7.4. *Let* $\lambda \in]0,1]$. *Under the conditions of Theorem A.7.1, there is a positive local martingale,* $Z^{(\lambda)}$ *stopped at* T*, and a decreasing positive process* $D^{(\lambda)}$ *stopped at* T*, with* $D_{0-} = 1$*, such that*

$$(\mathcal{E}(X))^\lambda = Z^{(\lambda)} D^{(\lambda)}.$$

These processes satisfy

$$Z^{(\lambda)} \leq (\mathcal{E}(X))^\lambda \exp((1-\lambda)B)$$

and, for any stopping time R, $Z_R^{(\lambda)} \to \mathcal{E}(M)_R$ *a.s. as* $\lambda \to 1$.

Proof. The processes can be explicitly given by

$$\begin{aligned} Z_t^{(\lambda)} &= \exp\Big(\lambda X_{t\wedge T} - \frac{1}{2}\langle \lambda X^c \rangle_{t\wedge T}\Big) \prod_{0 \leq u \leq t\wedge T} (1 + \lambda \Delta X_u) e^{-\lambda \Delta X_u}, \\ D_t^{(\lambda)} &= \exp\Big(-\frac{\lambda - \lambda^2}{2} \langle X^c \rangle_{t\wedge T}\Big) \prod_{0 \leq u \leq t\wedge T} \frac{(1+\Delta X_u)^\lambda}{1 + \lambda \Delta X_u}. \end{aligned}$$

We observe the convergence $Z_R^{(\lambda)} \to \mathcal{E}(M)_R$ a.s. as $\lambda \to 1$ immediately.

Recalling that B is locally bounded, take $\{T_n\}_{n\in\mathbb{N}}$ to be a series of stopping times such that $B^{T_n} \leq n$ and $\mathcal{E}(X)^{T_n}$ and $(\mathcal{E}(X)\hat{X})^{T_n}$ are martingales for each n. For n fixed, write Q for the probability measure defined by $dQ/dP = \mathcal{E}(X)^{T_n}$. From Lemmata A.7.3 and 15.2.1, we observe that $\hat{X}^{T_n}$ is a Q-martingale. Thus $(\lambda - 1)\hat{X}^{T_n}$ is also a Q-martingale. By construction,

$$\begin{aligned}\mathcal{E}(X^{T_n}) \exp\big((\lambda - 1)\hat{X}^{T_n}\big) &= \big(\mathcal{E}(X^{T_n})\big)^\lambda \exp\big((1-\lambda)B^{T_n}\big) \\ &\leq \big(\mathcal{E}(X^{T_n})\big)^\lambda \exp\big((1-\lambda)n\big).\end{aligned}$$

From this, as $\big(\mathcal{E}(X^{T_n})\big)^{\lambda}$ is a positive local P-supermartingale with $\mathcal{E}(X)_{0-} = 1$, we have

$$E_Q[\exp((\lambda-1)\hat{X}_{T_n})] \leq \exp((1-\lambda)n).$$

Therefore, $\exp((\lambda-1)\hat{X}^{T_n})$ is a Q-submartingale. Changing measure again, we see that $S^{T_n} := \mathcal{E}(X^{T_n})\exp((\lambda-1)\hat{X}^{T_n})$ is a P-submartingale bounded in $L^1(P)$.

Using the decomposition $\big(\mathcal{E}(X)\big)^{\lambda} = Z^{(\lambda)}D^{(\lambda)}$ stated above, we can write $S = Z^{(\lambda)}H$, where H is the predictable process $H = D^{(\lambda)}\exp((1-\lambda)B)$. It follows that

$$dS_t = Z^{(\lambda)}_{t-}dH_t + H_t dZ^{(\lambda)}_t,$$

from which we can deduce the Doob–Meyer decomposition of S. As $\{Z^{(\lambda)}_{t-}\}_{t\geq 0}$ is positive on $[\![0,T]\!]$ and S is a submartingale, H must be an increasing process, and so, on $[\![0,T_n]\!]$, we have

$$1 \leq D^{(\lambda)}\exp((1-\lambda)B) \qquad \text{and} \qquad Z^{(\lambda)} \leq \mathcal{E}(X)\exp((1-\lambda)B).$$

We now let $n\to\infty$, which yields the desired bound on $[\![0,T[\![$. As $\Delta X_T = -1$, we know that $Z^{(\lambda)}_t = 0 = \mathcal{E}(X)_t$ for all $t \geq T$. Therefore, the bound is established for all times t. □

Lemma A.7.5. *Under the conditions of Theorem A.7.1, for any $\lambda \in]0,1[$, the process $Z^{(\lambda)}$ constructed in Lemma A.7.4 is uniformly integrable.*

Proof. For each $n\in\mathbb{N}$, let $S_n = \inf\{t : B_t \geq n\}$. By the first Doob inequality (Theorem 5.1.2(i)), we know that $kP\big(\sup_t(\mathcal{E}(X)_t)^{\lambda} > k\big) \leq E[\mathcal{E}(X)_0] = 1$ for all $k\geq 0$, and so, by integrating with respect to $d(k^{\lambda})$, we obtain

$$E\big[\sup_t \big(\mathcal{E}(X)_t\big)^{\lambda}\big] \leq (1-\lambda)^{-1}.$$

From Lemma A.7.4, we have

$$E\big[\sup_{t<S_n} Z^{(\lambda)}_t\big] \leq \frac{\exp((1-\lambda)n)}{1-\lambda}.$$

Let $\{T_m\}_{m\in\mathbb{N}}$ be a sequence of finite stopping times reducing $Z^{(\lambda)}$. Then

$$1 = E[Z^{(\lambda)}_{S_n\wedge T_m}] = E[Z^{(\lambda)}_{T_m}I_{\{T_m\leq S_n\}}] + E[Z^{(\lambda)}_{S_n}I_{\{S_n<T_m\}}].$$

By dominated convergence, as $m\to\infty$, the first term on the right converges to $E[Z^{(\lambda)}_{S_n}I_{\{S_n=\infty\}}]$ and the second to $E[Z^{(\lambda)}_{S_m}I_{\{S_n<\infty\}}]$. Therefore, we can see that $E[Z^{(\lambda)}_{S_n}] = 1$.

To finish, we write

$$E[Z^{(\lambda)}_{\infty}] = 1 + E[Z^{(\lambda)}_{\infty}I_{\{S_n<\infty\}}] - E[Z^{(\lambda)}_{S_n}I_{\{S_n<\infty\}}]$$

and, using the bound on $Z^{(\lambda)}$ from Lemma A.7.4 and Hölder's inequality,

$$\begin{aligned} E[Z^{(\lambda)}_{S_n} I_{\{S_n<\infty\}}] &\leq E[(\mathcal{E}(X))^{\lambda}_{S_n} \exp((1-\lambda)B_{S_n}) I_{\{S_n<\infty\}}] \\ &\leq (E[\mathcal{E}(X)_{S_n}])^{\lambda} (E[\exp(B_{S_n}) I_{\{S_n<\infty\}}])^{1-\lambda} \\ &\leq (E[\exp(B_{S_n}) I_{\{S_n<\infty\}}])^{1-\lambda}. \end{aligned}$$

As we assumed that $E[\exp(B_\infty)] < \infty$, and we know $P(S_n < \infty) \to 0$ as $n \to \infty$, we see that $E[\exp(B_{S_n}) I_{\{S_n<\infty\}}] \to 0$ as $n \to \infty$. It follows that $E[Z^{(\lambda)}_\infty] \geq 1$. As $Z^{(\lambda)}$ is a positive local martingale, we know $E[Z^{(\lambda)}_\infty] \leq Z^{(\lambda)}_0 = 1$, so $E[Z^{(\lambda)}_\infty] = 1$ and we conclude that Z is a uniformly integrable martingale. □

We can now prove Theorem 15.4.3/A.7.1.

Proof. If $\lambda \in]0,1]$, we note that, from the definition of $\hat{X}$ and the bound on $Z^{(\lambda)}$ in Lemma A.7.4,

$$Z^{(\lambda)} \leq \exp(B) \exp(\lambda \hat{X}) \qquad \text{and} \qquad Z^{(\lambda)} \leq \mathcal{E}(X) \exp((\lambda-1)\hat{X}).$$

Let $T_k = \inf\{t : \hat{X}_t \leq -k\}$ for $k \in \mathbb{N}$. Then we can immediately deduce

$$\begin{aligned} Z^{(\lambda)}_{T_k} &\leq I_{\{T_k<\infty\}} \exp(B_{T_k}) \exp(-\lambda k) + I_{\{T_k=\infty\}} \mathcal{E}(X)_{T_k} \exp((1-\lambda)k) \\ &\leq I_{\{T_k<\infty\}} \exp(B_\infty) + I_{\{T_k=\infty\}} \exp(k) \mathcal{E}(X)_{T_k}. \end{aligned} \tag{A.4}$$

The last line of this inequality is integrable and independent of λ. Therefore, the family $Z^{(\lambda)}$ is uniformly integrable, and by the convergence stated in Lemma A.7.4,

$$E[\mathcal{E}(X)_{T_k}] = \lim_{\lambda \to 1} E[Z_{T_k}] = 1.$$

We now write

$$E[\mathcal{E}(X)_\infty] = 1 + E[\mathcal{E}(X)_\infty I_{\{T_k<\infty\}}] - E[\mathcal{E}(X)_{T_k} I_{\{T_k<\infty\}}]$$

and, using (A.4), in the limit $\lambda \to 1$ we have

$$E[\mathcal{E}(X)_t I_{\{T_k<\infty\}}] \leq E[\exp(B_\infty)] \exp(-k),$$

which tends to zero as $k \to \infty$. Therefore, we observe $E[\mathcal{E}(X)_\infty] = 1$, as desired. □

A.7.2 An Optional Condition

We now prove Theorem 15.4.5, which we again reproduce for the reader's convenience.

Theorem A.7.6. *Suppose that X is a local martingale with $\Delta X > -1$ and*

$$E\Big[\exp\Big(\frac{1}{2}\langle X^c\rangle_\infty\Big)\prod_t (1+\Delta X_t)\exp\Big(-\frac{\Delta X_t}{1+\Delta X_t}\Big)\Big] < \infty.$$

Then $\mathcal{E}(X)$ is a uniformly integrable martingale and $\mathcal{E}(X)_\infty > 0$ a.s.

Lemma A.7.7. *Under the conditions of Theorem A.7.6, X is in $\mathcal{H}^1$ and $\mathcal{E}(X)_\infty > 0$ a.s.*

Proof. It is clear that $E[\langle X^c\rangle_\infty] < \infty$. The inequality

$$(1+x)\exp\Big(-\frac{x}{1+x}\Big) \geq 1 + \frac{3-e}{e}x \qquad \text{for } x > 1$$

implies that

$$E\Big[\sum_t \Delta X_t I_{\{\Delta X \geq 1\}}\Big] < \infty.$$

For $|x| < 1$, we can find a constant $k > 0$ such that

$$\log(1+x) - \frac{x}{1+x} \geq ku^2,$$

which implies

$$E\Big[\sum_t (\Delta X_t)^2 I_{\{|\Delta X| < 1\}}\Big] < \infty.$$

Combining these, we have

$$\begin{aligned} E\big[[X]_\infty^{1/2}\big] &\leq \Big(E\Big[\langle X^c\rangle + \sum_t (\Delta X)^2 I_{\{|\Delta X|<1\}}\Big]^{1/2}\Big) + E\Big[\sum_t \Delta X_t I_{\{\Delta X \geq 1\}}\Big] \\ &< \infty. \end{aligned}$$

By the BDG inequalities, this implies that $X \in \mathcal{H}^1$ and, as in Lemma A.7.2, $\mathcal{E}(X)_\infty > 0$ a.s. □

The following lemma allows us to use complex analytic methods to address the integrability of our processes.

Lemma A.7.8. *There exists an $a > 1$ and $b > e$ such that if z is a complex number with $\Re(z) \leq a$ and $b|\Im(z)| \leq |\Re(z)|$ (where $\Re$ and $\Im$ denote the real and imaginary parts respectively), then $|(1-z)e^z| \leq 1$.*

Proof. Let $x = \Re(z)$. If z satisfies the stated conditions for some $b > e$, then

$$|(1-z)e^z| \leq \Big[(1-x)^2 + \frac{x^2}{b^2}\Big]^{1/2} e^x.$$

The real function $f(x) = [(1-x)^2 + x^2/b^2]^{1/2}e^x$ is increasing for $x \leq 0$, decreasing for $x \in [0, (b^2-1)/(b^2+1)]$, and then increasing thereafter. Therefore, to ensure that $f(x) \leq 1$ for all $x \leq a$, it is sufficient to choose some $a > 1$ such that $(a-1)e^a < 1$ and some $b > e$ such that

$$\Big[(1-a)^2 + \frac{a^2}{b^2}\Big]e^{2a} \leq 1.$$

□

Lemma A.7.9. *If $\Delta X > -1$, define the processes A and $\tilde{X}$ by*

$$\begin{aligned} A_t &= \frac{1}{2}\langle X^c\rangle + \sum_{0\leq s\leq t}\Big(\log(1+\Delta X_s) + \frac{\Delta X_s}{1+\Delta X_s}\Big), \\ \tilde{X}_t &= X_t - \langle X^c\rangle_t - \sum_{0\leq s\leq t}\frac{\Delta X_s}{1+\Delta X_s} \\ &= \log(\mathcal{E}(X)_t) - A_t. \end{aligned}$$

If a and b are as in Lemma A.7.8 and λ is a complex number satisfying

$$1 - a \leq \Re(\lambda) \leq 1 - b|\Im(\lambda)|$$

then

$$|\mathcal{E}(\lambda X)_t| \leq \exp(A_t)\exp(\Re(\lambda)\tilde{X}_t) \leq \mathcal{E}(X_t)\exp((\Re(\lambda)-1)\tilde{X}_t).$$

Proof. Following Lemma A.7.8, for all $x > -1$ we can write $z = (1-\lambda)\frac{x}{1+x}$, and we observe

$$|1+\lambda z|\exp\Big(-\Re(\lambda)\frac{x}{1+x}\Big) \leq (1+x)\exp\Big(-\frac{x}{1+x}\Big).$$

At the same time, as $\Re(\lambda) \leq 1 - |\Im(\lambda)|$, we have $\Re(\lambda^2/2 - \lambda + 1/2) \geq 0$. Therefore, we have the inequality

$$\begin{aligned} |\mathcal{E}(\lambda X)_t| &\leq \exp\Big(\Re(\lambda)X_t - \Big(\Re(\lambda) - \frac{1}{2}\Big)\langle X^c\rangle_T\Big)\prod_{0\leq s\leq t}|1+\lambda\Delta X_s|\exp(-\Re(\lambda)\Delta X_s) \\ &\leq \exp(A_t)\exp(\Re(\lambda)\tilde{X}_t). \end{aligned}$$

The second inequality follows by replacing A_t by $\log(\mathcal{E}(X)_t) - \tilde{X}_t$. □

We can now prove Theorem 15.4.5/A.7.6.

Proof. First note that, under the conditions of Theorem A.7.6, using the notation and result of Lemma A.7.7, the random variables A_∞ and $\tilde{X}_\infty$ exist and are almost surely finite. Therefore the inequalities of Lemma A.7.9 remain valid for $t = \infty$, and $\mathcal{E}(\lambda X)_\infty$ is almost surely an analytic function of λ.

Let $T_k = \inf\{t : \tilde{X}_t \leq -k\}$. If $1 - a \leq \Re(\lambda) \leq 1 - b|\Im(\lambda)|$, $\Re(\lambda) < 0$ and $t < T_k$, then from Lemma A.7.9 we have

$$|\mathcal{E}(\lambda M)_t| \leq \exp(A_\infty)\exp(-k\Re(\lambda)).$$

As $\Delta\tilde{X} = \Delta X/(1+\Delta X)$ and $-1 < \Delta X_{T_k} \le 0$, we have

$$|\mathcal{E}(\lambda X)_{T_k}| = |1+\lambda\Delta X_{T_k}|\,|\mathcal{E}(\lambda X)_{T_k-}| \le (1+|\lambda|)|\mathcal{E}(\lambda X)_{T_k-}|.$$

Hence, if $1-a \le \Re(\lambda) \le 1-b|\Im(\lambda)|$ and $\Re(\lambda) < 0$, then

$$|\mathcal{E}(\lambda X)_t| \le \exp(A_\infty)\exp(-k\Re(\lambda))(1+|\lambda|). \tag{A.5}$$

The real and imaginary parts of $\mathcal{E}(\lambda X)^{T_k}$ are, therefore, in $\mathcal{H}^1$, which implies that

$$E[\mathcal{E}(\lambda X)_{T_k}] = 1. \tag{A.6}$$

Again applying Lemma A.7.9, we see that if $\Re(\lambda) \ge 0$, then

$$|\mathcal{E}(\lambda X)_{T_k}| I_{\{T_k<\infty\}} \le \exp(A_\infty)\exp\big(-k\Re(\lambda)\big) I_{\{T_k<\infty\}} \tag{A.7}$$

and so, from Lemma A.7.8,

$$|\mathcal{E}(\lambda X)_{T_k}| I_{\{T_k=\infty\}} \le \mathcal{E}(M)_{T_k}\exp\big(-k(1-\Re(\lambda)\big) I_{\{T_k=\infty\}}. \tag{A.8}$$

From these inequalities (A.5, A.7, A.8), we see that $\mathcal{E}(\lambda X)_{T_k}$ is bounded by an integrable random variable independently of λ, for all λ in the set

$$\Lambda = \{\lambda : 1-a \le \Re(\lambda) \le 1-b|\Im(\lambda)|\}.$$

As $\mathcal{E}(\lambda X)_{T_k}$ is almost surely an analytic function of λ on the interior of Λ and is continuous on the boundary, we conclude that $E[\mathcal{E}(\lambda X)_{T_k}]$ is also analytic on the interior of Λ and continuous on the boundary. Together with (A.6), this shows that $E[\mathcal{E}(X)_{T_k}] = 1$ or, equivalently,

$$E[\mathcal{E}(X)_{T_k} I_{\{T_k<\infty\}}] + E[\mathcal{E}(X)_{T_k} I_{\{T_k=\infty\}}] = 1.$$

From (A.7), the first term of this sum converges to zero as $k \to \infty$, and it follows that $E[\mathcal{E}(X)_\infty] = 1$. $\square$

Remark A.7.10. Okada [141] gives an alternative proof of a modified version of this result, which is closer to including Kazamaki's criterion for processes with jumps. His assumption is that, for some $\alpha \in [0,1[$ and a nonnegative constant C, the set

$$\Bigg\{\exp\Big(\alpha X_S + \Big(\frac{1}{2}-\alpha\Big)\langle M^c\rangle_S - (1-\alpha)C\langle M^c\rangle_S^{1/2}\Big) \\ \times \prod_{0\le t\le S}(1+\Delta X_t)\exp\Big(-\Delta X_t + (1-\alpha)\frac{\Delta X_t^2}{1+\Delta X_t}\Big)\Bigg\}_{S\in\mathcal{T}_b}$$

is uniformly integrable, where S ranges over the bounded stopping times $\mathcal{T}_b$. At first glance, this seems considerably more general than the results of Lépingle and Mémin. However, the uniform integrability assumption is very strong and can only be easily verified in the case $\alpha = 0, C = 0$, which corresponds to the assumption of Theorem A.7.6.

Note that verifying this condition is particularly difficult as there is no guarantee that $\exp(X)$ is a true submartingale when X is a local martingale (as local submartingales are not necessarily submartingales).

Sokol [169] gives a version of Lépingle and Mémin's result where the jumps of X are assumed to be bounded strictly away from -1, and gives an optimal condition in this context. Protter and Shimbo [153] also study related questions and demonstrate various counterexamples and extensions.

A.8 BMO Spaces

In Remark 10.1.12 we saw that the spaces $\mathcal{H}^p$ and $\mathcal{H}^q$, where $q^{-1} + p^{-1} = 1$ are dual spaces (up to isomorphism). The case $p = 1$ was left unstudied, and in particular the Kunita–Watanabe inequality (Corollary 11.4.2) was not considered in this case. In this appendix, we shall give some basic results regarding the BMO space, which fills this gap in our understanding.

In the subsequent appendices, we also see how this theory naturally arises in the study of BSDEs with bounded terminal conditions, and hence appears in the theory of stochastic control.

The term BMO arises from classical analysis, where it stands for 'bounded mean oscillation', and is the dual of the Hardy space H^1. This theory was first modified for the stochastic setting by Garsia (in discrete time, see [85]), and then extended by Burkholder, Gundy, Herz, Stein, Doléans-Dade, Meyer, Kazamaki and others to the continuous time martingale setting. Our presentation is roughly guided by that of Meyer [133] (see also He, Wang and Yan [94]), with the addition of the more general semimartingale theory discussed in Meyer [135].

Before proceeding to the definition, we observe the following general equality. For τ any stopping time, M any local martingale,

$$E\big[(M_\infty - M_{\tau-})^2\big|\mathcal{F}_\tau\big] = E\big[[M]_\infty - [M]_{\tau-}\big|\mathcal{F}_\tau\big] \qquad \text{a.s.}$$

Definition A.8.1. *Let M be a local martingale. We say that M is a BMO martingale, and write $M \in \mathcal{H}^{\mathrm{BMO}}$, if there exists a constant k such that, for any stopping time $\tau \geq 0$,*

$$E\big[(M_\infty - M_{\tau-})^2\big|\mathcal{F}_\tau\big] = E\big[[M]_\infty - [M]_{\tau-}\big|\mathcal{F}_\tau\big] \leq k \qquad \text{a.s.}$$

We write, with $\mathcal{T}$ the set of all stopping times,

$$\|M\|^2_{\mathrm{BMO}} = \sup_{\tau\in\mathcal{T}} \big\|E\big[(M_\infty - M_{\tau-})^2\big|\mathcal{F}_\tau\big]\big\|_\infty.$$

If A is a finite variation process, then we say A is $\mathcal{A}$-BMO, and write $A \in \mathcal{A}^{\mathrm{BMO}}$, if $\|A\|_{\mathcal{A}^{\mathrm{BMO}}}$ is finite, where

$$\|A\|_{\mathcal{A}^{\mathrm{BMO}}} = \sup_{\tau\in\mathcal{T}} \Big\|E\Big[\int_{[\tau,\infty]} |dA_s|\Big|\mathcal{F}_\tau\Big]\Big\|_\infty.$$

If X is a semimartingale, then we say X is BMO, and write $X \in \mathcal{H}^{\mathrm{BMO}}_{\mathcal{S}}$, if $X = M + A$ for some $M \in \mathcal{H}^{\mathrm{BMO}}$, $A \in \mathcal{A}^{\mathrm{BMO}}$.

In the same way as for the $\mathcal{H}^p$ and $\mathcal{H}^p_{\mathcal{S}}$ norms, in particular from Remark 16.2.4, we can see that the $\mathcal{H}^{\mathrm{BMO}}$ and $\mathcal{H}^{\mathrm{BMO}}_{\mathcal{S}}$ norms are consistent for the martingales, and $\mathcal{H}^{\mathrm{BMO}}_{\mathcal{S}}$ and $\mathcal{A}^{\mathrm{BMO}}$ norms are consistent for predictable finite variation processes. Our attention will mainly be on the BMO martingales.

Remark A.8.2. By replacing a stopping time τ by τ_A, for $A \in \mathcal{F}_\tau$, we see that

$$E\big[(M_\infty - M_{\tau-})^2\big|\mathcal{F}_\tau\big] \leq k$$

is equivalent to

$$E\big[(M_\infty - M_{\tau-})^2\big] \leq kP(\tau < \infty),$$

and so the martingale BMO norm can be written

$$\|M\|^2_{\mathrm{BMO}} = \sup_{\tau\in\mathcal{T}}\left\{\frac{E\big[(M_\infty - M_{\tau-})^2\big]}{P(\tau<\infty)}\right\},$$

which is clearly a seminorm (and is a norm after considering indistinguishable processes to be equal). We shall see that it is also complete.

Remark A.8.3. A martingale M is BMO if and only if the jumps of M (including any jump at zero) are uniformly bounded, and if the process

$$E\big[(M_\infty - M_t)^2\big|\mathcal{F}_t\big] = E[M^2_\infty|\mathcal{F}_t] - M_t^2$$

is uniformly bounded. For example, a Brownian motion stopped at a deterministic time is BMO (even though it is not a bounded martingale).

Remark A.8.4. By taking $\tau = 0$, we observe immediately that $\mathcal{H}^{\mathrm{BMO}} \subset \mathcal{H}^2$ and $\mathcal{H}^{\mathrm{BMO}}_{\mathcal{S}} \subset \mathcal{H}^1_{\mathcal{S}}$. Applying Garsia's lemma (Theorem 8.2.21), we see that

$$E\big[[M]^p_\infty\big] \leq p^p\|M\|_{\mathrm{BMO}} \quad \text{and} \quad E\Big[\Big(\int_{[0,\infty]}|dA|_s\Big)^p\Big] \leq p^p\|A\|_{\mathcal{A}^{\mathrm{BMO}}},$$

so $\mathcal{H}^{\mathrm{BMO}} \subset \mathcal{H}^p$, $\mathcal{A}^{\mathrm{BMO}} \subset \mathcal{A}^p$ and $\mathcal{H}^{\mathrm{BMO}}_{\mathcal{S}} \subset \mathcal{H}^p_{\mathcal{S}}$ for all $p < \infty$. It follows that all processes in $\mathcal{H}^{\mathrm{BMO}}_{\mathcal{S}}$ are special semimartingales, and we define a norm on $\mathcal{H}^{\mathrm{BMO}}_{\mathcal{S}}$ by taking the *canonical* decomposition $X = M + A$, and defining $\|X\|_{\mathrm{BMO}} = \|M\|_{\mathrm{BMO}} + \|A\|_{\mathcal{A}^{\mathrm{BMO}}}$.

Lemma A.8.5. *We know $\mathcal{H}^\infty \subset \mathcal{H}^{\mathrm{BMO}}$ and $\mathcal{H}^\infty_{\mathcal{S}} \subset \mathcal{H}^{\mathrm{BMO}}_{\mathcal{S}}$. Furthermore, any bounded martingale is in $\mathcal{H}^{\mathrm{BMO}}$.*

Proof. It is enough to consider the semimartingale case. From the definition of $\mathcal{H}^\infty_{\mathcal{S}}$, we have

$$\Big\|[M]_\infty\Big\|_\infty + \Big\|\int_{[0,\infty]}|dA_s|\Big\|_\infty < \infty$$

for the canonical decomposition $X = M + A$. As the processes involved are nondecreasing, this quantity is larger than the $\mathcal{H}^{\mathrm{BMO}}_{\mathcal{S}}$ norm defined in Remark A.8.4, which yields the result.

That any bounded martingale is in $\mathcal{H}^{\mathrm{BMO}}$ is immediate from the definition. □

Remark A.8.6. It is easy to check, using Doob's inequalities, that for any bounded martingale, $\|M\|_{\mathrm{BMO}} \leq \sqrt{5}\|M^*_\infty\|_\infty$.

Lemma A.8.7. *If $X \in \mathcal{H}^{\mathrm{BMO}}_{\mathcal{S}}$, then $|\Delta X| < \|X\|_{\mathrm{BMO}}$. In particular, X is locally bounded.*

Proof. Suppose X is a local martingale. For any stopping time T,

$$(\Delta M)^2_T = [M]_T - [M]_{T-} \leq E\big[[M]_\infty - [M]_{T-}|\mathcal{F}_T] \leq \|M\|^2_{\mathrm{BMO}} \quad \text{a.s.}$$

As the jumps of M are on a thin set, the result follows. A similar argument holds when X is in $\mathcal{A}^{\mathrm{BMO}}$ and, hence, whenever $X \in \mathcal{H}^{\mathrm{BMO}}_{\mathcal{S}}$. It follows that, if $T_n = \inf\{t : |X_t| \geq n\}$, then $|X_{T_n}| \leq n + \|X\|_{\mathrm{BMO}}$, so X is locally bounded. □

Remark A.8.8. Let $M \in \mathcal{H}^{\mathrm{BMO}}$ have decomposition $M = M^c + M^d$, where M^c is continuous and M^d is totally discontinuous. Then, as $[M] = \langle M^c\rangle + [M^d]$, we have $\|M^c\|_{\mathrm{BMO}} \leq \|M\|_{\mathrm{BMO}}$ and $\|M^d\|_{\mathrm{BMO}} \leq \|M\|_{\mathrm{BMO}}$. Therefore, $M^c \in \mathcal{H}^{\mathrm{BMO}}$ and $M^d \in \mathcal{H}^{\mathrm{BMO}}$.

The following lemma describes a situation where one often encounters BMO martingales.

Lemma A.8.9. *Let A be an adapted, integrable, increasing, càdlàg process (i.e. $A \in \mathcal{A}^+$) and M the càdlàg version of the martingale $\{E[A_\infty|\mathcal{F}_t]\}_{t\geq 0}$.*

(i) If there exists a constant k such that

$$0 \leq M_t - A_{t-} \leq k \qquad \textit{for all } t,$$

then $M \in \mathcal{H}^{\mathrm{BMO}}$ and $\|M\|_{\mathrm{BMO}} \leq \sqrt{3}k$.

(ii) If A is predictable, $A_0 = 0$ and there exists a constant k such that

$$0 \leq M_t - A_t \leq k \qquad \textit{for all } t,$$

then $M \in \mathcal{H}^{\mathrm{BMO}}$ and $\|M\|_{\mathrm{BMO}} \leq 2\sqrt{3}k$.

In both cases, it also follows that $M - A \in \mathcal{H}^{\mathrm{BMO}}_{\mathcal{S}}$ and $\|M - A\|_{\mathrm{BMO}} \leq E[A_\infty] + \|M\|_{\mathrm{BMO}}$.

Proof. The predictable case (ii) can be reduced to the adapted case (i) as follows. Let $Y = M - A$, so $0 \leq Y \leq k$. As A is predictable, so are its jumps, and it follows that, for any predictable stopping time T,

$$E[\Delta Y_T|\mathcal{F}_{T-}] = E[\Delta M_T + \Delta A_T|\mathcal{F}_{T-}] = \Delta A_T,$$

so $0 \leq \Delta A \leq k$. Therefore, $0 \leq M - A_- \leq 2k$, and we are in case (i), (where $A_- = \{A_{t-}\}_{t\geq 0}$).

To prove case (i), let $X = M - A_-$, so $0 \le X \le k$. For any stopping time T, using the product rule for Stieltjes integrals, one can verify that

$$\begin{aligned}(A_\infty - A_{T-})^2 &= 2\int_{[T,\infty[} (A_\infty - A_{s-})dA_s - \sum_{s\ge T} (\Delta A_s)^2 \\ &\le 2\int_{[T,\infty[} (A_\infty - A_{s-})dA_s.\end{aligned}$$

Taking an expectation, as $\Pi_o(A_\infty - A_-) = X$, we have

$$E[(A_\infty - A_{T-})^2] \le 2E\Big[\int_{[T,\infty[} X_s dA_s\Big],$$

so

$$E[(A_\infty - A_{T-})^2] \le 2kE[A_\infty - A_{T-}] \le 2kE[X_T] \le 2k^2 P(T < \infty).$$

As in Remark A.8.2, by replacing T with T_B for some $B \in \mathcal{F}_T$, we conclude that $E[(A_\infty - A_{T-})^2|\mathcal{F}_T] \le 2k^2$. We then observe that

$$\begin{aligned}E[(M_\infty - M_{T-})^2|\mathcal{F}_T] &\le E[(A_\infty - A_{T-})^2 + X_{T-}^2|\mathcal{F}_T] \\ &\le E[(A_\infty - A_{T-})^2|\mathcal{F}_T] + k^2 \le 3k^2,\end{aligned}$$

as desired. □

The following examples are due to Garsia (see [133] for a presentation), we state them without proof.

Lemma A.8.10. *(i) Let X be a local martingale, B an adapted increasing process. If the process XB is bounded by 1, then the local martingale $M = B_- \bullet X$ is in BMO, and $\|M\|_{\mathrm{BMO}} \le \sqrt{6}$.*
(ii) Let U be a predictable process such that there exists an increasing process B with $|U| \le B$, and X and B satisfy (i). Then $U \bullet X$ is also in BMO.

A.8.1 Fefferman's Inequality

With these preliminaries on the BMO space, we can move towards the elegant Fefferman inequality, which is a nontrivial extension of the Kunita–Watanabe inequality to BMO processes (cf. Corollary 11.4.2). We follow the argument of Meyer [133]. The earliest version of this result in the continuous case is due to Getoor and Sharpe [87].

Theorem A.8.11 (Fefferman's Inequality for Martingales). *There exists a constant c such that, for M and N two local martingales,*

$$E\big[|[M,N]_\infty|\big] \le c\|M\|_{\mathcal{H}^1}\|N\|_{\mathrm{BMO}}.$$

Remark A.8.12. Using essentially the same proof, one can also show the slightly more general result

$$E\Big[\int_{[0,\infty[}|X_s||d[M,N]_s|\Big] \leq \sqrt{2}E\Big[\Big(\int_{[0,\infty[}X_s^2 d[M]_s\Big)^{1/2}\Big]\|N\|_{\mathrm{BMO}}$$

for any *optional* process X. (Taking $X = 1$ we observe the case of the theorem, by the BDG inequality.) We prove the more general result.

Proof. Let $C_t = \int_{[0,t]} X_s^2 d[M]_s$ and define the optional positive processes H and K by

$$H_t^2 = \frac{X_t^2}{\sqrt{C_t} + \sqrt{C_{t-}}}, \qquad K_t^2 = \sqrt{C_t}.$$

Using the product rule for Stieltjes integrals, we know

$$H_t^2 d[M]_t = \frac{dC_t}{\sqrt{C_t} + \sqrt{C_{t-}}} = d\sqrt{C_t}$$

and

$$H_t^2 K_t^2 \geq X_t^2/2, \qquad d[M]\text{-a.e.}$$

As $|d[M,N]|$ is absolutely continuous with respect to $d[M]$, using the Kunita–Watanabe inequality (Theorem 11.4.1), we see

$$\frac{1}{\sqrt{2}}E\Big[\int_{[0,\infty[}|X_s||d[M,N]_s|\Big] \leq E\Big[\int_{[0,\infty[}H_s K_s|d[M,N]_s|\Big] \leq \sqrt{E_1}\sqrt{E_2}.$$

where

$$E_1 = E\Big[\int_{[0,\infty[}H_s^2 d[M]_s\Big] = E\Big[\sqrt{C_\infty}\Big] = E\Big[\Big(\int_{[0,\infty[}X_s^2 d[M]_s\Big)^{1/2}\Big],$$
$$E_2 = E\Big[\int_{[0,\infty[}K_s^2 d[N]_s\Big] = E\Big[\int_{[0,\infty[}([N]_\infty - [N]_{s-})dK_s^2\Big]$$

(using integration by parts to calculate E_2). The optional projection of $\{[N]_\infty - [N]_{s-}\}_{s\geq 0}$ is $\big\{E\big[[N]_\infty\big|\mathcal{F}_s\big] - [N]_{s-}\big\}_{s\geq 0}$, and this is bounded by $\|N\|_{\mathrm{BMO}}^2$. As K^2 is an optional process,

$$\begin{aligned} E\Big[\int_{[0,\infty[}([N]_\infty - [N]_{s-})dK_s^2\Big] &= E\Big[\int_{[0,\infty[}(E\big[[N]_\infty\big|\mathcal{F}_s\big] - [N]_{s-})dK_s^2\Big] \\ &\leq \|N\|_{\mathrm{BMO}}^2 E\Big[\sqrt{C_\infty}\Big]. \end{aligned}$$

By substitution, we have

$$\frac{1}{\sqrt{2}}E\Big[\int_{[0,\infty[}|X_s||d[M,N]_s|\Big] \leq E\Big[\Big(\int_{[0,\infty[}X_s^2 d[M]_s\Big)^{1/2}\Big]\|N\|_{\mathrm{BMO}}$$

and the result follows. □

Corollary A.8.13. *For T a stopping time and $A \in \mathcal{F}_T$, replacing X with $XI_{[\![T,\infty[\![}I_A$, we obtain*

$$E\Big[\int_{[T,\infty[}|X_s||[M,N]_\infty|\Big|\mathcal{F}_T\Big] \leq \sqrt{2}E\Big[\Big(\int_{[T,\infty[}X_s^2 d[M]_s\Big)^{1/2}\Big|\mathcal{F}_T\Big]\|N\|_{\mathrm{BMO}}.$$

Using Fefferman's inequality, we can study BMO as the dual space of $\mathcal{H}^1$ (cf. Remark 10.1.12). This result is sometimes called Fefferman's theorem.

Theorem A.8.14. *The dual of $\mathcal{H}^1$ is isomporphic to $\mathcal{H}^{\mathrm{BMO}}$, that is,*

(i) any continuous linear function $\phi : \mathcal{H}^1 \to \mathbb{R}$ can be written $\phi(M) = E[M_\infty N_\infty] = E\big[[M,N]_\infty\big]$ for some $N \in \mathcal{H}^{\mathrm{BMO}}$ and
(ii) for any $N \in \mathcal{H}^{\mathrm{BMO}}$, the map $M \mapsto E\big[[M,N]_\infty\big]$ is a continuous linear function.

Proof. Statement (ii) is a direct consequence of Fefferman's inequality and the linearity of the quadratic covariation and expectation. To prove statement (i), we suppose without loss of generality that $\|\phi\|_{\mathrm{op}} = 1$, that is, $\sup\{\phi(M)\} = 1$, where the supremum is taken over M in the unit ball in $\mathcal{H}^1$. (The case when $\|\phi\|_{\mathrm{op}} = 0$ is trivial, and all others can be rescaled accordingly.) As $\mathcal{H}^2 \subset \mathcal{H}^1$ and the norm on $\mathcal{H}^2$ is stronger than that on $\mathcal{H}^1$, we see that ϕ induces a linear function on $\mathcal{H}^2$ with norm ≤ 1. Therefore, as $\mathcal{H}^2$ is reflexive, there exists a square integrable martingale N such that

$$\phi(M) = E[M_\infty N_\infty] = E\big[[M,N]_\infty\big]$$

for all $M \in \mathcal{H}^2$. As $\mathcal{H}^2$ is dense in $\mathcal{H}^1$ (Theorem 10.1.7), this extends uniquely to a continuous linear function on $\mathcal{H}^1$, so it remains only to show that N is in $\mathcal{H}^{\mathrm{BMO}}$.

We begin by considering the jumps of N. Let T be a stopping time, and we suppose T is either predictable or totally inaccessible. Let Z be a bounded, $\mathcal{F}_T$-measurable random variable such that $\|Z\|_{L^1} \leq 1$. Let $M = ZI_{[\![T,\infty[\![} - \Pi_p^*(ZI_{[\![T,\infty[\![})$, so M is a martingale of integrable variation. We can calculate

$$E\Big[\int_{[0,\infty[}|dM_s|\Big] \leq 2\|Z\|_{L^1} \leq 2,$$

so we know $\|M\|_{\mathcal{H}^1} = \|M\|_{\mathcal{H}^1_S} \leq 2$. On the other hand, as Z is bounded, M is square integrable, so

$$\phi(M) = E[M_\infty N_\infty] = E\big[[M,N]_\infty\big] = E[\Delta M_T \Delta N_T].$$

In the case where T is totally inaccessible, we have $\Delta M_T = Z$, so $|E[Z\Delta N_T]| \leq |\phi(M)| \leq 2$. Taking a supremum over the possible choices for Z, we have $\|\Delta N_T\|_{L^\infty} \leq 2$.

In the case where T is predictable, we have $\Delta M_T = Z - E[Z|\mathcal{F}_{T-}]$. As N is a martingale and T is predictable, $E[\Delta N_T|\mathcal{F}_{T-}] = 0$, so

$$E[\Delta M_T \Delta N_T] = E[Z\Delta N_T] - E\big[\Delta N_T E[Z|\mathcal{F}_{T-}]\big] = E[Z\Delta N_T].$$

As above, it follows that $\|\Delta N_T\|_{L^\infty} \leq 2$. By Theorem 6.2.9, we see that $\|\Delta N_T\|_{L^\infty} \leq 2$ for any stopping time T.

For T a general stopping time, we now consider the martingale $M := I_{]\!]T,\infty[\![} \bullet N$. We know $[M]_\infty = [M,N]_\infty = [N]_\infty - [N]_T$ and hence, for some $c > 0$, by the BDG inequality we have

$$\|M\|_{\mathcal{H}^1} \leq cE\big[([N]_\infty - [N]_T)^{1/2}\big].$$

As the term inside the expectation is zero whenever $T = \infty$, by the Cauchy–Schwarz inequality,

$$E\big[([N]_\infty - [N]_T)^{1/2}\big] \leq E\big[[N]_\infty - [N]_T\big]^{1/2} P(T < \infty)^{1/2}.$$

Therefore, we know that, for some constant $c > 0$, which may vary from line to line,

$$\begin{aligned} E\big[[N]_\infty - [N]_T\big] &= E[M_\infty N_\infty] = \phi(M) \\ &\leq c\|M\|_{\mathcal{H}^1} \\ &= cE\big[([N]_\infty - [N]_T)^{1/2}\big] \\ &\leq cE\big[[N]_\infty - [N]_T\big]^{1/2} P(T < \infty)^{1/2}. \end{aligned}$$

It follows that, for some $c > 0$,

$$E\big[[N]_\infty - [N]_T\big] \leq cP(T < \infty).$$

Replacing T by T_A for $A \in \mathcal{F}_T$, we observe that

$$E\big[[N]_\infty - [N]_T\big|\mathcal{F}_T\big] \leq c$$

and so, as we have seen that N has bounded jumps, $N \in \mathcal{H}^{\mathrm{BMO}}$ (Remark A.8.3). □

A.8.2 Semimartingale BMO

We now present two results for semimartingales. The first is a version of Émery's inequality (Lemma 16.2.9) for BMO processes (due to Meyer [135]), while the second is a version of Fefferman's inequality.

Theorem A.8.15 (Émery's Inequality for BMO). *For $p \in [1,\infty[$, there exists a constant c_p such that, for any semimartingale X and predictable process H,*

$$\|H \bullet X\|_{\mathcal{H}^p_S} \leq c_p \|H\|_{S^p} \|X\|_{\mathrm{BMO}}.$$

Note that, in this result, the p is the same on both sides, which was earlier only the case when BMO was replaced by $\mathcal{H}^\infty_{\mathcal{S}}$ on the right-hand side.

To prove the above result, we first assume $X \in \mathcal{H}^{\mathrm{BMO}}_{\mathcal{S}}$ (or the result is trivial), then take the canonical decomposition $X = M + A$, where $M \in \mathcal{H}^{\mathrm{BMO}}$ and $A \in \mathcal{A}^{\mathrm{BMO}}$ is predictable and of finite variation, and prove the bound for each component separately.

Lemma A.8.16. *For any stopping time T, any measurable process H and any $A \in \mathcal{A}^{\mathrm{BMO}}$,*

$$E\Big[\int_{[T,\infty[} |H_s||dA_s|\Big|\mathcal{F}_T\Big] \le \|A\|_{\mathcal{A}^{\mathrm{BMO}}} E[H^*_\infty|\mathcal{F}_T].$$

Consequently, for any $p \in [1,\infty[$, there exists $c_p > 0$ such that

$$\|H \bullet A\|_{\mathcal{H}^p_{\mathcal{S}}} \le c_p \|H\|_{S^p} \|A\|_{\mathcal{A}^{\mathrm{BMO}}}.$$

Proof. Let $B_t = \int_{[0,t]} |dA_s|$. Recall that $H^*_t := \lim_{s\downarrow t} \sup_{u\le s} |H_s|$. Then we have

$$\begin{aligned}
&E\Big[\int_{[T,\infty]} |H_{s-}|\,|dA_s|\Big|\mathcal{F}_T\Big]\\
&\le E\Big[\int_{[T,\infty]} H^*_s dB_s\Big|\mathcal{F}_T\Big]\\
&= E\Big[\int_{[T,\infty]} (B_\infty - B_{s-})dH^*_s + H^*_{T-}(B_\infty - B_{T-})\Big|\mathcal{F}_T\Big]\\
&\le \|A\|_{\mathcal{A}^{\mathrm{BMO}}}\, E[(H^*_\infty - H^*_{T-}) + H^*_{T-}|\mathcal{F}_T]\\
&= \|A\|_{\mathcal{A}^{\mathrm{BMO}}}\, E[H^*_\infty|\mathcal{F}_T]
\end{aligned}$$

To pass to the penultimate line, we use the fact that H^* is optional, and hence we can replace $B_\infty - B_{s-}$ by its optional projection, which is bounded by $\|A\|_{\mathcal{A}^{\mathrm{BMO}}}$. The first statement is then proven.

For $p = 1$, the second statement follows from the first by taking an expectation. For $p > 1$, the statement follows from Garsia's lemma (Theorem 8.2.21). □

Lemma A.8.17. *Let $M \in \mathcal{H}^{\mathrm{BMO}}$ and H be a predictable process. Then for any $p \in [1,\infty[$ there exists a constant c_p such that*

$$\|H \bullet M\|_{\mathcal{H}^p} \le c_p \|H\|_{S^p} \|M\|_{\mathrm{BMO}}.$$

Proof. If $p > 1$, then let q be the Hölder conjugate of p (so $p^{-1} + q^{-1} = 1$), and let $q =$ "BMO" if $p = 1$, for notational convenience.

By truncation and monotone convergence, it is enough to consider the case when H is bounded. Then the martingale $H \bullet M$ is in $\mathcal{H}^{\mathrm{BMO}} \subset \mathcal{H}^p$, and by Remark 10.1.12, Theorem A.8.14 and the definition of the dual space, we have

$$\|H \bullet M\|_{\mathcal{H}^p} \leq c_p \sup_N E\big[[H \bullet M, N]_\infty\big]$$

where N takes values in the unit ball in $\mathcal{H}^q$. Therefore, it is enough to prove that, if $\|M\|_{\mathrm{BMO}} \leq 1$ and $\|N\|_{\mathcal{H}^q} \leq 1$, then $E\big[[H \bullet M, N]_\infty\big] \leq c_p \|H\|_{S^p}$ for all bounded processes H.

To proceed, we consider two cases. First suppose $p > 1$. Then we can write $E\big[[H \bullet M, N]_\infty\big] = E\big[[M, H \bullet N]_\infty\big]$. From Émery's inequality for $\mathcal{H}^p$ (Lemma 16.2.9), we have

$$\|H \bullet N\|_{\mathcal{H}^1} \leq c_p \|H\|_{S^p} \|N\|_{\mathcal{H}^q} \leq c_p \|H\|_{S^p}.$$

The result follows from Fefferman's inequality (Theorem A.8.11).

Now suppose $p = 1$. Then we write $E\big[[H \bullet M, N]_\infty\big] = E\big[(H \bullet [M, N])_\infty\big]$, and we note that $|d[M, N]| \leq (d[M] + d[N])/2$, which is of finite variation. The result follows from Lemma A.8.16. □

Theorem A.8.18 (Fefferman's Inequality for Semimartingales). *There exists a constant c such that, for X and Y any two semimartingales,*

$$E\Big[\int_{[0,\infty[} |d[X, Y]|_s\Big] \leq c \|Y\|_{\mathrm{BMO}} \|X\|_{\mathcal{H}^1_S}.$$

Proof. We take the canonical decompositions $X = M + A$, $Y = N + B$, where $M \in \mathcal{H}^1$, $A \in \mathcal{A}$, $N \in \mathcal{H}^{\mathrm{BMO}}$ and $B \in \mathcal{A}^{\mathrm{BMO}}$. Then

$$[X, Y] = [X, N + B] = [X, B] + [A, N] + [M, N]$$

and we prove the inequality for each term on the right. The result then follows by the triangle inequality.

For $[X, B]$, as B is of finite variation, we know that, for any $p \geq 1$,

$$\begin{aligned} \Big\| \int_{[0,\infty[} |d[X, B]|_s \Big\|_{L^p} &= \Big\| \sum_s |\Delta X_s||\Delta B_s| \Big\|_{L^p} \leq \Big\| \int_{[0,\infty[} |\Delta X_s||dB_s| \Big\|_{L^p} \\ &\leq c_p \|\Delta X\|_{S^p} \|B\|_{\mathrm{BMO}} \leq c_p \|X\|_{\mathcal{H}^p_S} \|B\|_{\mathrm{BMO}}, \end{aligned}$$

where the penultimate inequality is due to Lemma A.8.16.

For $[A, N]$, as N has jumps bounded by $\|N\|_{\mathrm{BMO}}$ (Lemma A.8.7),

$$\begin{aligned} \Big\| \int_{[0,\infty[} |d[A, N]|_s \Big\|_{L^p} &= \Big\| \sum_s |\Delta A_s||\Delta N_s| \Big\|_{L^p} \leq \Big\| \sum_s |\Delta A_s| \Big\|_{L^p} \|N\|_{\mathrm{BMO}} \\ &\leq \Big\| \int_{[0,\infty[} |dA_s| \Big\|_{L^p} \|N\|_{\mathrm{BMO}} \leq \|A\|_{\mathcal{H}^p_S} \|N\|_{\mathrm{BMO}}. \end{aligned}$$

For $[M, N]$, we have the usual Fefferman inequality for martingales (Theorem A.8.11). Taking $p = 1$ and combining these estimates yields the result. □

Remark A.8.19. Meyer [135] proves the more general result that, for any $p \in [1, \infty[$ or $p =$ "BMO", there exists a constant c_p such that

$$\Big\| \int_{[0,\infty[} |d[Y,X]|_s \Big\|_{L^p} \leq c_p \|Y\|_{\mathrm{BMO}} \|X\|_{\mathcal{H}^p_{\mathcal{S}}}.$$

A.8.3 Relationship with Stochastic Exponentials

The real strength of the theory of BMO martingales is felt when we come to considering their stochastic exponentials. In particular, they form a class of martingales whose stochastic exponentials are also martingales, even though the Novikov and Kazamaki criteria may fail.

The following inequality is due to Strook [172].

Lemma A.8.20. *Let X be an adapted càdlàg process. Assume that $\lim_t X_t = X_\infty$ exists a.s. and is finite. If there exists a nonnegative integrable random variable ξ such that, for any stopping time T,*

$$E\big[|X_\infty - X_{T-}|\big|\mathcal{F}_T\big] \leq E[\xi|\mathcal{F}_T] \qquad \text{a.s.}$$

then, for all $\lambda \geq 0$ and $\mu > 0$, we have

$$\mu P(X^*_\infty \geq \lambda + \mu) \leq 2E[\xi I_{\{X^*_\infty \geq \lambda\}}].$$

Proof. Let $0 < \mu' < \mu$. Put $T = \inf\{t : |X_t| \geq \lambda\}$ and $S = \inf\{t : |X_t| \geq \lambda + \mu'\}$. Then $T \leq S$, $X_{T-} \leq \lambda$ and

$$\{X^*_\infty > \lambda + \mu'\} \subseteq \{|X_S| \geq \lambda + \mu'\} \subseteq \{|X_T| > \lambda\} \cap \{|X_S - X_{T-}| \geq \mu'\}.$$

As $|X_S - X_{T-}| \leq |X_\infty - X_{T-}| + |X_\infty - X_S|$, we see that

$$\begin{aligned} &P(X^*_\infty > \lambda + \mu') \\ &\leq \frac{1}{\mu'} E\big[|X_S - X_{T-}| I_{\{|X_T| \geq \lambda\}}\big] \\ &\leq \frac{1}{\mu'} \Big(E\big[|X_\infty - X_{T-}| I_{\{|X_T| \geq \lambda\}}\big] + E\big[|X_\infty - X_S| I_{\{|X_T| \geq \lambda\}}\big] \Big). \end{aligned}$$

Since $\{|X_T| \geq \lambda\} \in \mathcal{F}_T$, by assumption we have

$$E\big[|X_\infty - X_{T-}| I_{\{|X_T| \geq \lambda\}}\big] \leq E\big[\xi I_{\{|X_T| \geq \lambda\}}\big] \leq E\big[\xi I_{\{X^*_\infty \geq \lambda\}}\big].$$

As X is càdlàg, $\lim_{n\to\infty} X_{(S+1/n)-} = X_S$ and so, by Fatou's lemma,

$$\begin{aligned} E\big[|X_\infty - X_S| I_{\{|X_T| \geq \lambda\}}\big] &\leq \lim_{n\to\infty} E\big[|X_\infty - X_{(S+1/n)-}| I_{\{|X_T| \geq \lambda\}}\big] \\ &\leq E[\xi I_{\{|X_T| \geq \lambda\}}]. \end{aligned}$$

Combining these inequalities, we have

$$\mu' P(X^*_\infty > \lambda + \mu') \leq 2E\big[\xi I_{\{X^*_\infty \geq \lambda\}}\big].$$

The result follows by taking the limit $\mu' \uparrow \mu$. $\square$

This leads to a probabilistic version of the classical John–Nirenberg inequality.

Theorem A.8.21 (Probabilistic John–Nirenberg Inequality). *Let X be an adapted càdlàg process such that $\lim_{t\to\infty} X_t = X_\infty$ exists a.s. and is finite. If there is a constant $c > 0$ such that for any stopping time T,*

$$E\big[|X_\infty - X_{T-}|\big|\mathcal{F}_T\big] \leq c \qquad \text{a.s.}$$

then for any $0 \leq \lambda < (8c)^{-1}$,

$$E[e^{\lambda X^*_\infty}] < \frac{2e}{1-8c\lambda},$$

and, for any stopping time T,

$$E[e^{\lambda|X_\infty - X_{T-}|}|\mathcal{F}_T] < \frac{2e}{1-8c\lambda} \qquad \text{a.s.}$$

Proof. By Lemma A.8.20, for $n \geq 1$ we have

$$4cP(X^*_\infty \geq 4nc) \leq 2cP(X^*_\infty \geq 4(n-1)c).$$

Hence $P(X^*_\infty \geq 4nc) \leq 2^{-n} \leq \exp(-n/2)$. It follows that

$$\begin{aligned} E[e^{\lambda X^*_\infty}] &\leq \sum_{n=0}^{\infty} e^{4c\lambda(n+1)} P(4cn \leq X^*_\infty < 4c(n+1)) \\ &\leq e^{4c\lambda} \sum_{n=0}^{\infty} e^{-(1/2-4c\lambda)n} = \frac{e^{4c\lambda}}{1-e^{-(1/2-4c\lambda)}}. \end{aligned}$$

As $e^{-x} \leq 1 - x/\sqrt{e}$ for $x \in [0, 1/2]$, we obtain

$$E[e^{\lambda X^*_\infty}] \leq \frac{e^{1/2+4c\lambda}}{1/2-4c\lambda} < \frac{2e}{1-8c\lambda}.$$

To obtain the conditional inequality, we note that the above argument can be applied to give $E[e^{\lambda X^*_\infty} I_A] < \frac{2e}{1-8c\lambda} P(A)$ for any $A \in \mathcal{F}_0$, that is, $E[e^{\lambda X^*_\infty}|\mathcal{F}_0] < \frac{2e}{1-8c\lambda}$ a.s. We can then apply the argument to the process $Y_t := X_{T+t} - X_{T-}$ in the filtration given by $\mathcal{G}_t = \mathcal{F}_{T+t}$, to obtain

$$E[e^{\lambda(\sup_{t\geq T}|X_t - X_{T-}|)}|\mathcal{F}_T] = E[e^{\lambda Y^*_\infty}|\mathcal{G}_0] < \frac{2e}{1-8c\lambda} \qquad \text{a.s.}$$

□

Applying this inequality to BMO martingales, we obtain the following results.

Theorem A.8.22. *Let $M \in \mathcal{H}^{\mathrm{BMO}}$ and $\|M\|_{\mathrm{BMO}} = m$.*

(i) For $0 \leq \lambda < (8m)^{-1}$, we have

$$E[e^{\lambda M^*_\infty}] < \frac{2e}{(1 - 8m\lambda)}.$$

(ii) For any $0 \leq \lambda < m^{-2}$ and any stopping time T we have

$$E[e^{\lambda([M]_\infty - [M]_{T-})}|\mathcal{F}_T] < \frac{1}{1 - \lambda m^2}.$$

Proof. To prove (i), let T be a stopping time. By Jensen's inequality, we know

$$E\big[|X_\infty - X_{T-}|\big|\mathcal{F}_T\big] \leq E\big[(X_\infty - X_{T-})^2\big|\mathcal{F}_T\big]^{1/2} \leq m,$$

and the result follows by Theorem A.8.21.

To prove (ii), first note that a weaker inequality can be obtained by applying Theorem A.8.21 directly to $[M]$. To obtain the stated result, consider the increasing process $A = [M]$. As $M \in \mathcal{H}^{\mathrm{BMO}}$, for any stopping time T we have

$$E[A_\infty - A_{T-}|\mathcal{F}_T] \leq m^2 \qquad \text{a.s.}$$

By Garsia's lemma (Theorem 8.2.21) and Lemma 8.2.18, we know $E[A^n_\infty] \leq n!m^{2n}$, and so $E[\exp(\lambda A_\infty)] \leq \frac{1}{1-\lambda m^2}$. The conditional expectation version of the result can then be proven in the same way as the second statement in Theorem A.8.21. □

Corollary A.8.23. *Let $1 < p < \infty$. There are positive constants c_p and C_p such that, for any uniformly integrable martingale M,*

$$c_p\|M\|_{\mathrm{BMO}} \leq \sup_{T\in\mathcal{T}} \big\|E\big[|M_\infty - M_T|^p\big|\mathcal{F}_T\big]\big\|_\infty^{1/p} \leq C_p\|M\|_{\mathrm{BMO}}.$$

Proof. If $p \geq 2$, then the first inequality follows directly from Jensen's inequality (with $c_p = 1$). For $p > 1$, the BDG and L^p inequalities can be used.

For the second inequality, assume $\|M\|_{\mathrm{BMO}} > 0$ (otherwise the result is trivial). Then, from Theorem A.8.21,

$$E\Big[\exp\Big(\frac{|M_\infty - M_T|}{e + 8\|M\|_{\mathrm{BMO}}}\Big)\Big|\mathcal{F}_T\Big] \leq 2.$$

From this the conclusion is straightforward, via Taylor expansion of the exponential. □

The following results are adapted from Doléans-Dade and Meyer [60] and Izumisawa, Sekiguchi and Shiota [104]. Condition (A_p) is sometimes referred to as a *Muckenhoupt* condition, see [60]. In the continuous case, the importance of this condition for martingales is explored by Izumisawa and Kazamaki [103].

Theorem A.8.24. *The following are equivalent, for M a local martingale.*

(i) M is a BMO martingale, and there exists a positive constant $\epsilon > 0$ such that $\epsilon < 1 + \Delta M < \epsilon^{-1}$.

(ii) $Z := \mathcal{E}(M)$ has the following three properties:

- *(Condition (S)) For some constant $\epsilon > 0$ we have $\epsilon < Z/Z_- < \epsilon^{-1}$.*
- *(Positivity) $Z_\infty > 0$ a.s. (Note that Condition (S) implies Z is a nonnegative supermartingale, so Z_∞ is well defined)*
- *(Condition (A_p), for some p) There exist constants $p > 1$ and $K > 0$ such that*

$$E\big[(Z_T/Z_\infty)^{\frac{1}{p-1}}\big|\mathcal{F}_T\big] \le K \qquad \text{a.s.}$$

for any stopping time T.

Furthermore, the constants p and K in (ii) can be taken to depend only on $\|M\|_{\mathrm{BMO}}$ and ϵ.

Proof. (i $\Rightarrow$ ii) It is easy to check that Condition (S) and positivity are satisfied, using the facts $\epsilon < 1 + \Delta M < \epsilon^{-1}$ and $\|M\|_{\mathrm{BMO}} < \infty$. We assume without loss of generality that ϵ is small.

We need to show that Condition (A_p) is satisfied for some p. The inequality

$$e^{x-(x^2)/(2\epsilon^2)} \le 1 + x \le e^x$$

for $0 < a < (1-\epsilon)^{-1}$ and $-1+\epsilon \le x \le (1-\epsilon)/(2a)$ implies the inequalities

$$\begin{aligned} (1+x)^{-a} &\le (1-2ax)^{1/2} \exp\left(\frac{ax^2}{2\epsilon^2}\right), \\ 1 + ax &\le (1+x)^a \exp\left(\frac{ax^2}{2\epsilon^2}\right). \end{aligned} \tag{A.9}$$

We can choose $a > 0$ such that $k_a := (4a^2 + a)/\epsilon^2 < 1/\|M\|^2_{\mathrm{BMO}}$ and $(1 - \epsilon)/(2a) > \epsilon^{-1}$. For any stopping time T, by the Cauchy–Schwarz inequality,

$$\begin{aligned} &E[(Z_T/Z_\infty)^a|\mathcal{F}_T] \\ &= E\Big[e^{a(M_T-M_\infty)+\frac{a}{2}(\langle M^c\rangle_\infty-\langle M^c\rangle_T)} \prod_{u>T} \frac{e^{a\Delta M_u}}{(1+\Delta M_u)^a}\Big|\mathcal{F}_T\Big] \\ &\le E\Big[e^{a(M_T-M_\infty)+\frac{a}{2}(\langle M^c\rangle_\infty-\langle M^c\rangle_T)} \\ &\qquad\quad \times \prod_{u>T} (1-2a\Delta M_u)^{1/2} e^{a\Delta M_u + a(\Delta M_u)^2/(2\epsilon^2)}\Big|\mathcal{F}_T\Big] \\ &= E\Big[\Big(e^{a(M_T-M_\infty)-a^2(\langle M^c\rangle_\infty-\langle M^c\rangle_T)} \prod_{u>T} (1-2a\Delta M_u)^{1/2} e^{a\Delta M_u}\Big) \\ &\qquad\quad \times \Big(e^{\frac{1}{2}(2a^2+a)(\langle M^c\rangle_\infty-\langle M^c\rangle_T)+\frac{a}{2\epsilon^2}\sum_{u>T}(\Delta M_u)^2}\Big)\Big|\mathcal{F}_T\Big] \\ &\le E[\mathcal{E}(2aM)_\infty/\mathcal{E}(2aM)_T|\mathcal{F}_T]^{1/2} E[e^{k_a([M]_\infty-[M]_T)}|\mathcal{F}_T]^{1/2}. \end{aligned}$$

The first term in the last line is less than 1, while Theorem A.8.22(ii) implies the second is bounded by $(1-k_a\|M\|^2_{\mathrm{BMO}})^{-1/2}$. Condition (A_p) follows, and we note that values of $p=1+a^{-1}$ and k_a can be explicitly given in terms of $\|M\|_{\mathrm{BMO}}$ and ϵ.

(ii $\Rightarrow$ i) As Z is positive, we know that $1+\Delta M>0$. Then Condition (S) guarantees that $1+\Delta M>\epsilon$ for some $\epsilon>0$. Condition (S) also implies that the jumps of M are bounded above by a constant k, and we can assume $k\geq 1$. From Condition (A_p), writing $a=1/(p-1)$ for convenience, we know that, for any stopping time T,

$$\begin{aligned} K &\geq E[(Z_T/Z_\infty)^a|\mathcal{F}_T] \\ &= E\Big[e^{a(M_T-M_\infty)+\frac{a}{2}(\langle M^c\rangle_\infty-\langle M^c\rangle_T)}\prod_{u>T}\frac{e^{a\Delta M_u}}{(1+\Delta M_u)^a}\Big|\mathcal{F}_T\Big]. \end{aligned}$$

The inequality $e^x-1-x\geq x^2/e$ (for $x\geq -1$) implies that, for $x\in\,]-1,k]$,

$$\frac{e^x}{1+x}\geq 1+\frac{x^2}{e(1+x)}\geq 1+\frac{x^2}{e(1+k)}\geq e^{bx^2},$$

where $0<b\leq(\log(1+\frac{k^2}{e(1+k)})/k^2)\wedge\frac{1}{2}$. Then

$$\begin{aligned} K &\geq E[(Z_T/Z_\infty)^a|\mathcal{F}_T] \\ &\geq E\big[e^{aM_\infty-M_T+ab(\langle M^c\rangle_\infty-\langle M^c\rangle_T)+ba\sum_{u>T}\Delta M_u^2}\big|\mathcal{F}_T\big] \\ &= E\big[e^{a(M_T-M_\infty)+ab([M]_\infty-[M]_T)}\big|\mathcal{F}_T\big]. \end{aligned} \tag{A.10}$$

If we suppose for a moment that M is a uniformly integrable martingale, then Jensen's inequality yields

$$\begin{aligned} K &\geq \exp\big(E[a(M_T-M_\infty)+ab([M]_\infty-[M]_T)|\mathcal{F}_T]\big) \\ &= \exp(E[ab([M]_\infty-[M]_T)|\mathcal{F}_T]), \end{aligned}$$

and so $E\big[[M]_\infty-[M]_T\big|\mathcal{F}_T\big]$ is bounded. As M has bounded jumps, it follows that M is a BMO martingale. On the other hand, if we do not assume that M is a true martingale, then by localization and rearranging (A.10) we see that there exists a sequence of stopping times $T_n\to\infty$ such that M^{T_n} is a BMO martingale for each n, with the uniform bound $\|M^{T_n}\|_{\mathrm{BMO}}\leq\frac{1}{ab}\log(K)+k^2$. Taking $n\to\infty$, we have a sequence which is uniformly bounded in the BMO norm, so the limit M is in $\mathcal{H}^{\mathrm{BMO}}$ by monotone convergence. □

Corollary A.8.25. *Under the conditions of Theorem A.8.24, there exist constants $k\geq 1$, $\delta>0$ and $\eta>0$, which depend only on ϵ, that is, the bound for the jumps of M, and on $\|M\|_{\mathrm{BMO}}$, such that, for any stopping time T, we have*

$$Z_T\leq kZ_{T-}\qquad P(Z_\infty<\delta Z_{T-}|\mathcal{F}_T)\leq 1-\eta.$$

Proof. The first inequality is Condition (S), while the second follows from Conditions (S) and (A_p) of the theorem, together with Markov's inequality. $\square$

This implies that BMO martingales of this type yield processes $Z = \mathcal{E}(M)$ which satisfy the conditions of the following theorem, due to Émery [72], based on work of Coifman and Fefferman [35]. In particular, if M is a BMO martingale with jumps in $[-1+\epsilon, \epsilon^{-1}]$, then we see that $\mathcal{E}(M)$ is an $\mathcal{H}^p$ martingale, for some $p > 1$ (and hence uniformly integrable).

Theorem A.8.26. *Let Z be a positive local martingale such that $Z_0 = 1$. Suppose there are three constants $k \geq 1$, $\delta > 0$ and $\eta > 0$ such that for any stopping time T, we have*

$$Z_T \leq k Z_{T-}, \qquad P(Z_\infty < \delta Z_{T-} | \mathcal{F}_T) \leq 1 - \eta.$$

Then Z is a martingale in $\mathcal{H}^p$ for some $p > 1$. More precisely, there exist maps $p = p(k, \eta, \delta) > 1$ and $c = c(k, \eta, \delta) < \infty$ such that $\|Z\|_{\mathcal{H}^p} \leq c$.

Proof. Without loss of generality, we can assume that $\delta \leq 1$. We shall begin by reducing to the case where Z is bounded.

For $n \geq 2$, let $R = \inf\{t : Z_t > n\}$, so the martingale Z^R is bounded by kn. On the set $\{Z^R_\infty < \delta Z^R_{T-}\}$, we have $Z^R_\infty < Z^R_{T-}$, and so $R \geq T$. Therefore, $Z^R_{T-} = Z_{T-} \leq n$, hence $Z_R = Z^R_\infty \leq \delta n \leq n$. This implies that, on $\{Z^R_\infty < \delta Z^R_{T-}\}$, we know $R = \infty$ and $Z_\infty < \delta Z_{T-}$. Consequently,

$$P(Z^R_\infty < \delta Z^R_{T-} | \mathcal{F}_T) \leq P(Z_\infty < \delta Z_{T-} | \mathcal{F}_T) \leq 1 - \eta$$

and the stopped martingale Z^R also satisfies the hypotheses of the theorem. If we can show that the result holds for Z^R, as p and c depend only on k, η and δ, a monotone convergence argument with $n \to \infty$ implies that it also holds for Z, as desired.

Now suppose Z is bounded. By Doob's L^p inequality (Theorem 4.5.6), our problem is to give a bound on $\|Z_\infty\|_{L^p}$ for some $p > 1$. By assumption, for any $T \in \mathcal{T}$, we know that $P(Z_\infty \geq \delta Z_{T-} | \mathcal{F}_T) \geq \eta$, so

$$P(Z_\infty \geq \delta Z_{T-}, T < \infty) \geq \eta P(T < \infty).$$

For $\lambda > 1$, we apply this to the stopping time $T = \inf\{t : Z_t > \lambda\}$. On $\{T < \infty\}$, we have $Z_{T-} \leq \lambda \leq Z_T$, and so $\lambda/k \leq Z_{T-}$ and $Z_T \leq k\lambda$. It follows that

$$\eta P(T < \infty) \leq P(Z_\infty \geq \delta\lambda/k),$$

and we deduce (always for $\lambda > 1$),

$$\begin{aligned} E[Z_\infty I_{\{Z_\infty > \lambda\}}] &\leq E[Z_\infty I_{\{T < \infty\}}] = E[Z_T I_{\{Z_\infty > \lambda\}}] \\ &\leq k\lambda P(T < \infty) \leq \frac{k\lambda}{\eta} P(Z_\infty \geq \delta\lambda/k). \end{aligned}$$

For any $p > 1$, we integrate this inequality with respect to the measure $I_{\{\lambda>1\}}(p-1)\lambda^{p-2}\,d\lambda$ and obtain

$$E\left[Z_\infty(Z_\infty^{p-1}-1)I_{\{Z_\infty>1\}}\right] \leq \frac{k(p-1)}{\eta p}E\Big[\Big(\Big(\frac{kZ_\infty}{\delta}\Big)^p - 1\Big)I_{\{Z_\infty>\delta/k\}}\Big].$$

Setting $u = \frac{k(p-1)}{\eta p}\left(\frac{k}{\delta}\right)^p$, as

$$E[Z_\infty I_{\{Z_\infty>1\}}] \leq E[Z_\infty] = 1,$$

we can write

$$E[Z_\infty^p I_{\{Z_\infty>1\}}] - 1 \leq uE[Z_\infty^p] \leq uE[Z_\infty^p I_{\{Z_\infty>1\}}] + u.$$

Leaving k, η and δ fixed, we send $p \to 1$, so $u \to 0$. For p sufficiently close to 1, we have $u < 1$. Fix such a p. It follows that

$$E[Z_\infty^p] \leq E[Z_\infty^p I_{\{Z_\infty<1\}}] + 1 \leq \frac{1+u}{1-u} + 1 = \frac{2}{1-u} < \infty.$$

As p and u depend only on k, η and δ, the result follows. □

Remark A.8.27. A consequence of this estimate is that the reverse Hölder inequality (Corollary 15.4.8) holds for $\mathcal{E}(M)$, where M is a BMO martingale with jumps in $[-1+\epsilon, \epsilon]$. That is, for $p > 1$ sufficiently small, there exists a constant c_p (depending only on p, $\|M\|_{\mathrm{BMO}}$ and ϵ) such that, for every stopping time T,

$$E[Z_\infty^p|\mathcal{F}_T] \leq c_p Z_T^p.$$

A.9 Non-Lipschitz BSDEs

In this appendix, we consider the solution of BSDEs with weaker conditions on the driver than those in Chapter 19. We restrict our attention to the one-dimensional case and assume the terminal value ξ and the values of the driver evaluated at zero $f(\omega, t, 0, 0, 0)$ are bounded (uniformly in (ω, t)), and the driver is balanced. On the other hand, we allow either

(i) f to be a function of quadratic growth with respect to (z, θ), or
(ii) the Lipschitz constant of f with respect to (z, θ) to depend on (ω, t) in a predictable way (with some integrability assumptions).

The key difficulty in doing this is that, unlike the case of SDEs, we cannot exploit localization techniques in our proofs – we need to show that the equations have solutions with the fixed horizon T, as this is where the terminal condition is prescribed. The result we give is not the most general possible, but is sufficient for many problems, particularly in stochastic control and mathematical finance.

BSDEs with drivers of quadratic growth in z (in the continuous setting) were first considered by Kobylanski [119]. Tevzadze [176] gives a construction based on slicing the terminal value into small pieces, for each of which a contraction mapping approach can be used. Briand and Elie [24] give a construction based on Malliavin calculus techniques. Extending away from bounded terminal values, results have been obtained by Briand and Hu [25] and Barrieu and El Karoui [4], amongst others. See also counterexamples in Delbaen, Hu and Bao [52] and approaches with jumps in Delong [55]. For the case of stochastic Lipschitz bounds, Hamadène and Lepeltier [93] give an approach in the continuous case, which we generalize.

We assume, as in Chapter 19, that we are in a filtered probability space with an $N \leq \infty$ dimensional Brownian motion W, and a random measure μ with deterministic compensator $\mu_p(d\zeta, dt) = \nu(d\zeta)dt$, such that the pair $(W, \tilde{\mu})$ has the predictable representation property. Recall that a (scalar) BSDE is an equation of the form

$$dY_t = -f(\omega, t, Y_t, Z_t, \Theta_t)dt + Z_t dW_t + \int_{\mathcal{Z}} \Theta_t(\zeta)\tilde{\mu}(d\zeta, dt)$$

with terminal value $Y_T = \xi$ for some $\mathcal{F}_T$-measurable random variable ξ, where f is a real-valued $\Sigma_p \otimes \mathcal{B}(\mathbb{R} \times \mathbb{R}^N \times L^2(\nu))$-measurable function.

A.9.1 Quadratic BSDEs

In all this section, we fix a deterministic final time $T < \infty$.

By restricting our attention to bounded terminal conditions (and assuming $f(\omega, t, 0, 0, 0)$ is bounded), we can make use of the following connection between BSDEs and the theory of BMO martingales, which extends a result in Touzi [177] (see also Hu, Imkeller and Müller [96]).

Theorem A.9.1. *Let f be the driver for a BSDE, with solution (Y, Z, Θ). Assume that Y is uniformly bounded and, for some constant $c > 0$,*

$$|f(\omega, t, y, z, \theta)| \leq c(1 + |y| + \|z\|^2 + \|\theta\|_\nu^2) \qquad dt \times dP\text{-a.e.}$$

Then the martingale part $Z \bullet W + \Theta * \tilde{\mu}$ is in $\mathcal{H}^{\mathrm{BMO}}$ and its BMO norm can be bounded in terms of $\|Y_T^*\|_\infty$ and c.

Proof. Let $\{\tau_n\}_{n\in\mathbb{N}}$ be a localizing sequence for the local martingale $Z \bullet W + \Theta * \tilde{\mu}$, and write $T_n = T \wedge \tau_n$. As Y is bounded and $\Delta Y_t = \int_{\mathcal{Z}} \Theta_t(\zeta)\mu(d\zeta, t)$, we see that Θ is pointwise bounded (for at least one representative in $L^2(\nu)$). In particular, $\Theta^2 * \tilde{\mu}$ is a well-defined local martingale, and by monotone convergence,

$$E\Big[\int_{]\tau,T]} \Theta^2(\zeta)\mu(d\zeta, dt)\Big] = E\Big[\int_{]\tau,T]} \Theta^2(\zeta)\nu(d\zeta)dt\Big] \leq \|\Theta * \tilde{\mu}\|_{\mathrm{BMO}}$$

for any stopping time τ. For notational clarity, we omit the (ω, t, Y_t) arguments of f. By Itô's formula and the boundedness of Y, for any stopping time τ and any $\beta \geq 0$ we have

$$
\begin{aligned}
e^{\beta\|Y_T^*\|_\infty} &\geq e^{\beta Y_{T_n}} - e^{\beta Y_\tau} \\
&= \int_{]\tau,T_n]} \beta e^{\beta Y_{t-}}\Big(- f(Z_t,\Theta_t)dt + Z_t dW_t + \int_{\mathcal{Z}} \Theta_t(\zeta)\tilde{\mu}(d\zeta,dt)\Big) \\
&\quad + \frac{\beta^2}{2}\int_{]\tau,T_n]} e^{\beta Y_{t-}}\|Z_t\|^2 dt + \sum_{t\in]\tau,T_n]} e^{\beta Y_{t-}}\Big(e^{\beta\Delta Y_t} - 1 - \beta\Delta Y_t\Big) \\
&\geq \int_{]\tau,T_n]} \beta e^{\beta Y_{t-}}\Big(- f(Z_t,\Theta_t) + \frac{\beta}{2}e^{\beta Y_{t-}}\|Z_t\|^2\Big)dt \\
&\quad + \int_{]\tau,T_n]} \beta e^{\beta Y_{t-}}\Big(Z_t dW_t + \int_{\mathcal{Z}} \Theta_t(\zeta)\tilde{\mu}(d\zeta,dt)\Big) \\
&\quad + \frac{\beta^2}{2}\sum_{t\in]\tau,T_n]} e^{\beta Y_{t-}}(\Delta Y_t)^2.
\end{aligned}
$$

Taking an expectation (and using the optional stopping theorem) and rearranging, we obtain

$$
\begin{aligned}
&\frac{\beta^2}{2}E\Big[\int_{]\tau,T_n]} e^{-\beta Y_{t-}}\|Z_t\|^2 dt + \int_{]\tau,T_n]} e^{-\beta Y_{t-}}(\Theta_t(\zeta))^2\mu(d\zeta,dt)\Big|\mathcal{F}_\tau\Big] \\
&= \frac{\beta^2}{2}E\Big[\int_{]\tau,T_n]} e^{-\beta Y_{t-}}\|Z_t\|^2 dt + \sum_{u\in]\tau,T_n]} e^{-\beta Y_{t-}}(\Delta Y_t)^2\Big|\mathcal{F}_\tau\Big] \\
&\leq e^{\beta\|Y_T^*\|_\infty} + \beta E\Big[\int_{]\tau,T_n]} e^{\beta Y_{t-}} f(Z_t,\Theta_t)dt\Big|\mathcal{F}_\tau\Big].
\end{aligned}
$$

Letting $n \to \infty$, by monotone convergence, we see this inequality holds with T_n replaced by T.

As we know $|f(y,z,\theta)| \leq c(1 + |y| + \|z\|^2 + \|\theta\|_\nu^2)$, we see that

$$
\begin{aligned}
&\frac{\beta^2}{2}e^{-\beta\|Y_T^*\|_\infty}E\Big[\int_{]\tau,T]}\|Z_t\|^2 dt + \int_{\mathcal{Z}\times]\tau,T]}(\Theta_t(\zeta))^2\mu(d\zeta,dt)\Big|\mathcal{F}_\tau\Big] \\
&\leq e^{\beta\|Y_T^*\|_\infty} + c\beta E\Big[\int_{]\tau,T]} e^{\beta Y_{t-}}(1 + |Y_{t-}| + \|Z_t\|^2 + \|\Theta_t\|_\nu^2)dt\Big|\mathcal{F}_\tau\Big].
\end{aligned}
$$

Setting $\beta = 4c$, we conclude

$$
\begin{aligned}
&8c^2 e^{-4c\|Y_T^*\|_\infty}E\big[[Z\bullet W + \Theta * \tilde{\mu}]_T - [Z\bullet W + \Theta * \tilde{\mu}]_\tau\big|\mathcal{F}_\tau\big] \\
&= 8c^2 e^{-4c\|Y_T^*\|_\infty}E\Big[\int_{]\tau,T]}\|Z_t\|^2 dt + \int_{]\tau,T_n]}(\Theta_t(\zeta))^2\mu(d\zeta,dt)\Big|\mathcal{F}_\tau\Big] \\
&\leq e^{4c\|Y_T^*\|_\infty} + 4c^2 E\Big[\int_{]\tau,T]} e^{4cY_{t-}}(1 + |Y_{t-}|)dt\Big|\mathcal{F}_\tau\Big] < \infty.
\end{aligned}
$$

As τ was arbitrary, $Z \bullet W + \Theta * \mu \in \mathcal{H}^{\mathrm{BMO}}$. We note that this also gives an explicit bound on the BMO norm in terms of $\|Y_T^*\|_\infty$ and c. $\square$

We now consider BSDEs with drivers of the following type.

Definition A.9.2. *We say a map $f : \Omega \times [0,T] \times \mathbb{R}^n \times L^2(\nu) \to \mathbb{R}$ (predictable in (ω,t), Borel measurable in its other arguments) is an ϵ-balanced quadratic driver if there exists $K > 0$ such that*

$$|f(\omega,t,z,\theta) - f(\omega,t,z',\theta)| \leq K(1 + \|z\| + \|z'\|)\|z - z'\|$$

and there exists a function β, which is Borel measurable in (y,z,θ,θ') and predictable in (ω,t), satisfies $\epsilon < 1 + \beta(\zeta) < \epsilon^{-1}$ ν-a.e., and is such that

$$f(\omega,t,z,\theta) - f(\omega,t,z,\theta') = \int_{\mathcal{Z}} (\theta(\zeta) - \theta'(\zeta))\beta(\zeta;\omega,t,z,\theta,\theta')\nu(d\zeta)$$

and

$$\|\beta(\cdot;\omega,t,z,\theta,\theta')\|_\nu \leq \epsilon^{-1}(\|\theta\|_\nu + \|\theta'\|_\nu)$$

for all t,z,θ,θ', P-almost surely. Without loss of generality, we may assume $K = \epsilon^{-1}$, to save notation.

It is easy to verify that an ϵ-balanced quadratic driver satisfies the growth condition of Theorem A.9.1.

Remark A.9.3. We have assumed that our driver does not depend directly on y. This is not strictly necessary, and we can introduce Lipschitz dependence on y without much difficulty. However, the notation becomes more involved, and the case where f does not depend on y is of primary interest for stochastic control problems.

We now connect solutions of BSDEs, assuming they exist, with conditional expectations under changed measures.

Lemma A.9.4. *Suppose (Y,Z,Θ) and (Y',Z',Θ') are solutions to the BSDEs with respective ϵ-balanced quadratic drivers f,f' and terminal values ξ,ξ', where, for some $K > 0$,*

$$|\xi| + \int_{[0,T]} |f(\omega,s,0,0)|ds \leq K \qquad \text{a.s.}$$

and similarly for ξ', f'. Assume Y, Y' are bounded. Write

$$\begin{aligned}
\phi_t &:= f(\omega,t,Z'_t,\Theta'_t) - f'(\omega,t,Z'_t,\Theta'_t),\\
\alpha_t &:= \frac{f(\omega,t,Y_t,Z_t,\Theta'_t) - f(\omega,t,Y_t,Z'_t,\Theta'_t)}{\|Z_t - Z'_t\|^2}(Z_t - Z'_t)^\top
\end{aligned}$$

and $\beta_t(\zeta) = \beta(\zeta;\omega, t, Y_t, Z_t, \Theta_t, \Theta'_t)$ *from Definition A.9.2. If* $\int_{[0,t]} \phi_s ds$ *is uniformly bounded, then*

$$Y_t - Y'_t = E^Q\Big[\xi - \xi' + \int_{]t,T]} \phi_s ds \Big| \mathcal{F}_t\Big],$$

where $dQ/dP = \mathcal{E}(\alpha^\top \bullet W + \beta * \tilde{\mu})_T$.

Under these assumptions, $\mathcal{E}(\alpha^\top \bullet W + \beta * \tilde{\mu}) \in \mathcal{H}^p$ *for some* $p > 1$ *depending only on* K *and* ϵ.

Proof. First note that $\|\alpha_t\| \leq K(\|Z_t\| + \|Z'_t\|)$ and $\|\beta\|_\nu \leq K(\|\Theta_t\|_\nu + \|\Theta'_t\|_\nu)$, from the bounds on f. From Theorem A.9.1, we know that $Z \bullet W + \Theta * \tilde{\mu}$ and $Z' \bullet W + \Theta' * \tilde{\mu}$ are in $\mathcal{H}^{\mathrm{BMO}}$, and their BMO norms can be bounded in terms of ϵ, K and the bounds on Y and Y'. It follows from the definition of BMO that $\alpha^\top \bullet W + \beta * \tilde{\mu}$ is in $\mathcal{H}^{\mathrm{BMO}}$, with a bound on its BMO norm depending on ϵ, K and the bounds on Y and Y'. We also know

$$\epsilon \leq 1 + \Delta(\alpha^\top \bullet W + \beta * \tilde{\mu}) = 1 + \Delta(\beta * \tilde{\mu}) = 1 + \int_{\mathcal{Z}} \beta(\zeta)\mu(d\zeta, t) \leq \epsilon^{-1},$$

and so by Theorem A.8.26, $\mathcal{E}(\alpha^\top \bullet W + \beta * \tilde{\mu}) \in \mathcal{H}^p$ for some p depending on ϵ, K and the bounds on Y and Y'.

Write $\Lambda_t = \mathcal{E}(\alpha \bullet W + \beta * \tilde{\mu})_t$. A simple application of the Itô product rule shows that $\{\Lambda_t(Y_t - Y'_t + \int_{[0,t]} \phi_s ds)\}_{t \geq 0}$ is a P-local martingale. As there is a constant c such that $\big|Y_\tau - Y'_\tau + \int_{[0,\tau]} \phi_s ds\big| < c$, and $\Lambda \in \mathcal{H}^p$ for some $p > 1$, for any stopping time τ we know

$$E\Big[\Big|\Lambda_\tau\Big(Y_\tau - Y'_\tau + \int_{[0,\tau]} \phi_s ds\Big)\Big|^p\Big] \leq c^p E\big[|\Lambda_\tau|^p\big] \leq c^p \|\Lambda\|_{\mathcal{H}^p}$$

and so we see that $\Lambda(Y - Y' + \int_{[0,\cdot]} \phi_s ds)$ is of class (D), in particular, it is a uniformly integrable martingale. It follows that

$$Y_t - Y'_t = E\Big[\frac{\Lambda_T}{\Lambda_t}\Big(\xi - \xi' + \int_{]t,T]} \phi_s ds\Big)\Big|\mathcal{F}_t\Big] = E^Q\Big[\xi - \xi' + \int_{]t,T]} \phi_s ds \Big| \mathcal{F}_t\Big]$$

as desired.

Taking $\xi' = 0, f' = 0$, we see that

$$Y_t = E^Q\Big[\xi + \int_{]t,T]} f(\omega, t, 0, 0) ds \Big| \mathcal{F}_t\Big].$$

where $dQ/dP = \mathcal{E}(\alpha^\top \bullet W + \beta * \tilde{\mu})$, and ρ, α and β are chosen accordingly. Consequently, any bounded solution is bounded by the bounds on ξ and $f(\omega, t, 0, 0, 0)$, that is, $2K$. Therefore, the BMO norm of $Z \bullet W + \Theta * \tilde{\mu}$ depends only on ϵ and K, as does p. □

Theorem A.9.5. *Suppose we are in the setting described at the beginning of Chapter 19. Consider a BSDE with bounded terminal value $\xi \in L^\infty(\mathcal{F}_T)$ and an ϵ-balanced quadratic driver f, such that $\int_{[0,T]} |f(\omega, t, 0, 0, 0)|dt$ is a.s. bounded, and f can be uniformly approximated by a sequence of twice Fréchet differentiable functions with bounded second derivative. Then there exists a solution to the BSDE (Y, Z, Θ), where Y is a bounded semimartingale and $Z \bullet W + \Theta * \tilde{\mu} \in \mathcal{H}^{\mathrm{BMO}}$.*

The requirement that f can be approximated by twice Fréchet differentiable functions (Definition A.9.10) is known to be unnecessary in most cases, however this assumption greatly simplifies the proof. We prove the result in stages. Using Lemma A.9.4, it is easy to show uniqueness of the solution.

Theorem A.9.6. *For BSDEs of the type considered in Theorem A.9.5, the comparison theorem is satisfied. In particular, the solution is unique (in the class of solutions with Y bounded).*

Proof. This is almost immediate from Lemma A.9.4. If $\xi \geq \xi'$ and $f(\omega, t, z, \theta) \geq f'(\omega, t, z, \theta)$, then ϕ is positive and

$$Y_t - Y_t' = E^Q\Big[\xi - \xi' + \int_{]t,T]} \phi_s ds \Big| \mathcal{F}_t\Big] \geq 0.$$

Uniqueness of the solution follows from taking $\xi = \xi'$ and $f = f'$. □

Theorem A.9.7. *For BSDEs of the type considered in Theorem A.9.5, consider solutions Y, Y' to two BSDEs with the same ϵ-balanced driver f and respective bounded terminal values ξ, ξ'. Then there exists a constant $q > 1$, depending on the bounds on ξ, ξ', $\int_{[0,T]} |f(\omega, s, 0, 0)|ds$ and ϵ, such that*

$$E\big[\sup_{s\in[t,T]} |Y_s - Y_s'|^q \big| \mathcal{F}_t\big] \leq C_q E\Big[\Big(|\xi - \xi'| + \int_{]t,T]} |\phi_s| ds\Big)^{q^2/(q-1)} \Big| \mathcal{F}_t\Big]^{1-1/q}$$

and

$$\begin{aligned} &E\Big[\int_{]t,T]} (\|Z_s - Z_s'\|^2 + \|\Theta_s - \Theta_s'\|_\nu^2) ds \Big| \mathcal{F}_t\Big] \\ &\leq E\big[(\xi - \xi')^2 \big| \mathcal{F}_t\big] \\ &\qquad + C_q' E\big[\sup_{s\in[t,T]} |Y_s - Y_s'|^q \big| \mathcal{F}_t\big] + 2E\Big[\int_{]t,T]} |Y_s - Y_s'||\phi_s| ds \Big| \mathcal{F}_t\Big] \end{aligned}$$

where ϕ is as in Lemma A.9.4 and C_q and C_q' are constants depending only on q, ϵ and the bounds on ξ, ξ' and $\int_{[0,T]} |f(\omega, s, 0, 0)|ds$.

Proof. We know $Y_t - Y_t' = E^Q[\xi - \xi' + \int_{]t,T]} \phi_s ds | \mathcal{F}_t]$ for some Q with a density Λ in $\mathcal{H}^p$, where $p > 1$ depends on the quantities mentioned in the theorem.

Let $q = \sqrt{p}$ and write $\tilde{\xi} = |\xi - \xi'| + \int_{]t,T]} |\phi_s| ds$. Then we have, by Doob's inequality,

$$\begin{aligned} E[\sup_{s\in[t,T]} |Y_s - Y_s'|^q|\mathcal{F}_t] &\leq E\big[\sup_{s\in[t,T]} |\Lambda_T/\Lambda_s|^q|\tilde{\xi}|^q\big|\mathcal{F}_t\big] \\ &\leq E\big[\sup_{s\in[t,T]} |\Lambda_T/\Lambda_s|^p\big|\mathcal{F}_t\big]^{1/q} E[|\tilde{\xi}|^{q^2/(q-1)}\big|\mathcal{F}_t\big]^{1-1/q}. \end{aligned} \tag{A.11}$$

Using the Reverse Hölder inequality (Corollary 15.4.8), we see that

$$C_q := \Big\|E\big[\sup_{s\in[t,T]} |\Lambda_T/\Lambda_s|^p\big|\mathcal{F}_t\big]^{1/q}\Big\|_\infty < \infty$$

gives the first bound.

To prove the second bound, we expand $(Y_t - Y_t')^2$ using Itô's formula, to obtain

$$\begin{aligned} d(Y_t - Y_t')^2 = \|Z_t - Z_t'\|^2 dt &+ \|\Theta - \Theta\|_\nu^2 dt + (Y_t - Y_t')(Z_t - Z_t')dW \\ &+ (Y_t - Y_t')\int_{\mathcal{Z}} (\Theta_t(\zeta) - \Theta_t'(\zeta))\tilde{\mu}(d\zeta, dt) \\ &+ -2(Y_t - Y_t')(f(Z_t, \Theta_t) - f'(Z_t', \Theta_t'))dt \end{aligned}$$

where we omit the other arguments of f and f' for clarity.

Integrating and taking a conditional expectation (with localization and taking a limit) we have

$$\begin{aligned} E[(\xi - \xi')^2|\mathcal{F}_t] \geq (Y_t - Y_t')^2 &+ E\Big[\int_{]t,T]} (\|Z_s - Z_s'\|^2 + \|\Theta_s - \Theta_s\|_\nu^2) ds\Big|\mathcal{F}_t\Big] \\ &- 2E\Big[\int_{]t,T]} |Y_s - Y_s'|\,|f(Z_s, \Theta_s) - f(Z_s', \Theta_s')| ds\Big|\mathcal{F}_t\Big] \\ &- 2E\Big[\int_{]t,T]} |Y_s - Y_s'||\phi_s| dt\Big|\mathcal{F}_t\Big] \end{aligned} \tag{A.12}$$

and, for some constant C, which may vary from line to line,

$$\begin{aligned} &E\Big[\int_{]t,T]} |Y_t - Y_t'|\,|f(Z_t, \Theta_t) - f'(Z_t', \Theta_t')| dt\Big|\mathcal{F}_t\Big] \\ &\leq CE\Big[\sup_{t\leq s<T}\{|Y_s - Y_s'|^*\}\int_{]t,T]} (1 + \|Z_s\|^2 + \|Z_s'\|^2 + \|\Theta_s\|_\nu^2 + \|\Theta_s'\|_\nu^2) ds\Big|\mathcal{F}_s\Big] \\ &\leq CE\big[\big(\sup_{t\leq s<T}\{|Y_s - Y_s'|^*\}\big)^q\big|\mathcal{F}_t\big]^{1/q} \\ &\qquad \times E\Big[\Big(\int_{]t,T]} (1 + \|Z_s\|^2 + \|Z_s'\|^2 + \|\Theta_s\|_\nu^2 + \|\Theta_s'\|_\nu^2)\Big)^{q/(q-1)} ds\Big|\mathcal{F}_t\Big]^{1-1/q} \\ &\leq CE\big[\big(\sup_{t\leq s<T}\{|Y_s - Y_s'|^*\}\big)^q\big|\mathcal{F}_t\big]^{1/q}. \end{aligned} \tag{A.13}$$

The final line follows from the bound on $\|Z\bullet W + \Theta * \tilde{\mu}\|_{\mathcal{H}^{q/(q-1)}}$ implied by Theorem A.9.1. Substituting (A.13) and (A.11) into (A.12) gives the desired result. $\square$

Following the ideas of Tevzadze [176], we now prove the existence result of Theorem A.9.5 in a sequence of Lemmata. First, we assume that f is purely quadratic (has no linear term), and that the terminal value is small. Under these conditions, we can derive a contraction mapping result, and so obtain existence via Picard iteration. Second, we show how, assuming f is twice differentiable, we can piece together equations with small terminal values, and so solve the equation for any bounded terminal condition. Finally, we use an approximation argument to see that the requirement that f is twice differentiable can be partially relaxed. This presentation is based on Touzi [177], who considers only the continuous case.

Definition A.9.8. *We write $L^{\mathrm{BMO}}(\langle W\rangle)\times L^{\mathrm{BMO}}(\langle\tilde{\mu}\rangle)$ for the set of processes (Z,Θ) such that $Z\bullet W$ and $\Theta * \tilde{\mu}$ are both in $\mathcal{H}^{\mathrm{BMO}}$. Note that $L^{\mathrm{BMO}}(\langle W\rangle)\times L^{\mathrm{BMO}}(\langle\tilde{\mu}\rangle)\subset L^2(\langle W\rangle)\times L^2(\langle\tilde{\mu}\rangle)$. The norm on this space is given by $\|Z\bullet W+\Theta*\tilde{\mu}\|_{\mathrm{BMO}}$.*

Lemma A.9.9. *Consider a BSDE satisfying the conditions of Theorem A.9.5, and, in addition, suppose $f(\omega,t,0,0)=0$,*

$$|f(\omega,t,z,\theta)-f(\omega,t,z',\theta')|\leq k\big(\|z\|+\|z'\|+\|\theta\|_\nu+\|\theta'\|_\nu\big)\big(\|z-z'\|+\|\theta-\theta'\|\big)$$

and

$$\|\xi\|_{L^\infty}\leq 1/(256k).$$

Then the BSDE has a (unique) solution $(Y,Z,\Theta)\in S^\infty\times L^{\mathrm{BMO}}(\langle W\rangle)\times L^{\mathrm{BMO}}(\langle\tilde{\mu}\rangle)$.

Proof. Consider the map $\Phi:(y,z,\theta)\mapsto(Y,Z,\Theta), S^\infty\times L^{\mathrm{BMO}}(\langle W\rangle)\times L^{\mathrm{BMO}}(\langle\tilde{\mu}\rangle)\mapsto S^\infty\times L^{\mathrm{BMO}}(\langle W\rangle)\times L^{\mathrm{BMO}}(\langle\tilde{\mu}\rangle)$ defined by

$$Y_t=\xi+\int_{]t,T]}f(\omega,s,z_s,\theta_s)ds-\int_{]t,T]}Z_s dW_s-\int_{\mathcal{Z}\times]t,T]}\Theta_s(\zeta)\tilde{\mu}(d\zeta,ds).$$

This map is well defined, given that $(W,\tilde{\mu})$ have the predictable representation property and (z,θ) are in $L^2(\langle W\rangle)\times L^2(\langle\tilde{\mu}\rangle)$. We seek to show that Φ is at least locally a contraction. Before doing this, we show that Φ maps a certain bounded subset of $S^\infty\times L^{\mathrm{BMO}}(\langle W\rangle)\times L^{\mathrm{BMO}}(\langle\tilde{\mu}\rangle)$ to itself.

Applying Ito's formula to $|Y|^2$, for any stopping time τ we have (in the same way as (A.12)),

$$\begin{aligned}&|Y_\tau|^2+E\Big[\int_{]\tau,T]}\big(\|Z_s\|^2+\|\Theta_s\|_\nu^2\big)ds\Big|\mathcal{F}_\tau\Big]\\&\leq E\Big[\xi^2+\int_{]\tau,T]}2Y_s f(\omega,s,z_s,\theta_s)ds\Big|\mathcal{F}_\tau\Big]\\&\leq\|\xi\|_{L^\infty}^2+2\|Y\|_{S^\infty}E\Big[\int_{]\tau,T]}|f(\omega,s,z_s,\theta_s)|ds\Big|\mathcal{F}_\tau\Big].\end{aligned}$$

Using $2ab \leq \frac{1}{4}a^2 + 4b^2$ and the norm on f, we have

$$\begin{aligned}
&|Y_\tau|^2 + E\Big[\int_{]\tau,T]} \big(\|Z_s\|^2 + \|\Theta_s\|_\nu^2\big)ds\Big|\mathcal{F}_\tau\Big] \\
&\leq \|\xi\|_{L^\infty}^2 + \frac{1}{4}\|Y\|_{S^\infty}^2 + 16k^2 E\Big[\int_{]\tau,T]} \big(\|z\|^2 + \|\theta\|_\nu^2\big)ds\Big|\mathcal{F}_\tau\Big]^2.
\end{aligned}$$

Taking a supremum over all stopping times, we find

$$\|Y\|_{S^\infty}^2 + \|Z{\bullet}W + \Theta * \tilde{\mu}\|_{\mathrm{BMO}}^2 \leq 2\|\xi\|_{L^\infty}^2 + \frac{1}{2}\|Y\|_{S^\infty} + 32k^2\|z{\bullet}W + \theta * \tilde{\mu}\|_{\mathrm{BMO}}^4$$

and, consequently,

$$\|Y\|_{S^\infty}^2 + \|Z{\bullet}W + \Theta * \tilde{\mu}\|_{\mathrm{BMO}}^2 \leq 4\|\xi\|_{L^\infty}^2 + 64k^2\|z{\bullet}W + \theta * \tilde{\mu}\|_{\mathrm{BMO}}^4.$$

The key problem is the fourth power on the right-hand side. For this reason, we assume that ξ is small. If we assume that

$$\|y\|_{S^\infty}^2 + \|z{\bullet}W + \theta * \tilde{\mu}\|_{\mathrm{BMO}}^2 \leq R^2 := \frac{1}{2048k^2},$$

then

$$\|Y\|_{S^\infty}^2 + \|Z{\bullet}W + \Theta * \tilde{\mu}\|_{\mathrm{BMO}}^2 \leq 4\|\xi\|_{L^\infty}^2 + 64k^2R^4 \leq R^2.$$

Therefore, we see that Φ maps the ball of radius R in $S^\infty \times L^{\mathrm{BMO}}(\langle W\rangle) \times L^{\mathrm{BMO}}(\langle\tilde{\mu}\rangle)$ to a subset of itself.

Now let $(Y, Z, \Theta) = \Phi(y, z, \theta)$ and $(Y', Z', \Theta') = \Phi(y', z', \theta')$, where (y, z, θ) and (y', z', θ') are in the ball of radius R. Using (A.12), and similar arguments to above, we deduce

$$\begin{aligned}
&\|Y - Y'\|_{S^\infty}^2 + \|(Z - Z'){\bullet}W + (\Theta - \Theta') * \tilde{\mu}\|_{\mathrm{BMO}}^2 \\
&\leq 16 \sup_\tau E\Big[\int_{]\tau,T]} |f(z_t, \theta_t) - f(z_t', \theta_t')|dt\Big|\mathcal{F}_\tau\Big]^2 \\
&\leq 32k^2 \sup_\tau E\Big[\int_{]\tau,T]}(\|z\| + \|z'\| + \|\theta\|_\nu + \|\theta'\|_\nu)(\|z - z'\| + \|\theta - \theta'\|)dt\Big|\mathcal{F}_\tau\Big]^2 \\
&\leq 32k^2 \sup_\tau E\Big[\int_{]\tau,T]}(\|z\| + \|z'\| + \|\theta\|_\nu + \|\theta'\|_\nu)^2 dt\Big|\mathcal{F}_\tau\Big] \\
&\qquad \times \sup_\tau E\Big[\int_{]\tau,T]}(\|z - z'\| + \|\theta - \theta'\|_\nu)^2 dt\Big|\mathcal{F}_\tau\Big] \\
&\leq 256k^2R^2 \sup_\tau E\Big[\int_{]\tau,T]}(\|z - z'\| + \|\theta - \theta'\|_\nu)^2 dt\Big|\mathcal{F}_\tau\Big] \\
&\leq 1024k^2R^2\|(z - z'){\bullet}W + (\theta - \theta') * \tilde{\mu}\|_{\mathrm{BMO}}^2 \\
&\leq \frac{1}{2}\|(z - z'){\bullet}W + (\theta - \theta') * \tilde{\mu}\|_{\mathrm{BMO}}^2.
\end{aligned}$$

Therefore, Φ is a contraction on the ball of radius R. It follows that there exists a unique fixed point in this ball. This guarantees existence of solutions to our BSDE. Global uniqueness follows from the comparison theorem (Theorem A.9.6). □

We now see that we can perturb an existing BSDE solution by a small change in the terminal value and the driver.

Definition A.9.10. *Recall that, for H a Hilbert space, a function $f : \mathbb{R}^N \times L^2(\nu) \to H$ is called* Fréchet differentiable *if, for every (z, θ) there exists a bounded linear operator $Df(z, \theta) : H \to H$ such that*

$$\lim_{h_z, h_\theta \to 0} \frac{\|f(z + h_z, \theta + h_\theta) - f(z, \theta) - (Df(z, \theta))(h_z, h_\theta)\|_H}{\|h_z\| + \|h_\theta\|_\nu} \to 0.$$

When $H = \mathbb{R}$, we can write

$$(Df(z, \theta))(h_z, h_\theta) = D_z f^\top h_z + \int_{\mathcal{Z}} D_\theta f(\zeta) h_\theta(\zeta) d\nu$$

for some $D_z f \in \mathbb{R}^N$ and $D_\theta f \in L^2(\nu)$. We define the norm $\|D_{(z,\theta)} f\| = \|D_z f\| + \|D_\theta f\|_\nu$.

Remark A.9.11. We can similarly define a twice differentiable function, and the derivative will take values in the tensor space $(\mathbb{R}^N \times L^2(\nu))^{\otimes 2}$ with the corresponding norm. A general fact we shall use, which follows from the mean value theorem, is that, if f is twice differentiable and $\|D^2 f\| \leq k$, then $\|D_{(z,\theta)} f\| \leq \|Df(0, 0)\| + k(\|z\| + \|\theta\|_\nu)$.

Lemma A.9.12. *Suppose that, for a given terminal value ξ and driver f satisfying the conditions of Theorem A.9.5, the BSDE with data (ξ, f) admits a unique solution (Y, Z, Θ). If (ξ', f') satisfy the conditions of Theorem A.9.5 and are such that*

- $\|\xi - \xi'\|_{S^\infty} < 1/(256k)$
- *$f(\omega, t, 0, 0) - f'(\omega, t, 0, 0) = 0$ $dt \times dP$-a.e. and*
- *$f - f'$ is twice Fréchet differentiable in (z, θ), $\|Df(0, 0)\| \leq k$ and $\|D^2 f(z, \theta)\| \leq k$ for some k*

then the BSDE with data (ξ', f') admits a unique solution.

Proof. Define (omitting the ω, t arguments for simplicity),

$$f_\delta(z, \theta) := f'(z + Z_t, \theta + \Theta_t) - f(Z_t, \Theta_t).$$

As f_δ is Fréchet differentiable at $(0,0)$, we can define the processes $\phi_z = D_z f_\delta(0,0) \in \mathbb{R}^N$ and $\phi_\theta = D_\theta f_\delta(0,0) \in L^2(\nu)$. As f_δ is ϵ-balanced, we observe that $\phi_\theta > -1+\epsilon$. Define

$$g(z,\theta) := f_\delta(z,\theta) - z\phi_z - \int_{\mathcal{Z}} \phi_\theta(\zeta)\theta(\zeta)\nu(d\zeta).$$

By the mean value theorem and the fact $g(z,\theta) = g(z,\theta) - g(0,0)$, there exists a stochastic λ such that

$$\begin{aligned}
&\big|g(z,\theta) - g(z',\theta')\big| \\
&\le \big\|Df_\delta(\lambda z + (1-\lambda)z', \lambda\theta + (1-\lambda)\theta') - Df_\delta(0,0)\big\|(\|z-z'\| + \|\theta-\theta'\|_\nu) \\
&\le k\big(\|\lambda z + (1-\lambda)z'\| + \|\lambda\theta + (1-\lambda)\theta'\|_\nu\big)(\|z-z'\| + \|\theta-\theta'\|_\nu) \\
&\le k\big(\|z\| + \|z'\| + \|\theta\|_\nu + \|\theta'\|_\nu\big)(\|z-z'\| + \|\theta-\theta'\|_\nu).
\end{aligned}$$

We seek to decompose our solution as $Y' = Y + Y^\delta$, and similarly for Z^δ and Θ^δ. Existence of Y' is equivalent to that of Y^δ. Therefore, we now consider the BSDE with data $(f_\delta, \xi - \xi')$, and rewrite it as

$$\begin{aligned}
dY_t^\delta &= -f_\delta(Z^\delta, \Theta^\delta)dt + Z_t^\delta dW_t + \int_{\mathcal{Z}} \Theta_t^\delta(\zeta)\tilde{\mu}(d\zeta, dt) \\
&= -g(Z,\Theta)dt + Z_t dW_t^Q + \int_{\mathcal{Z}} \Theta_t(\zeta)\tilde{\mu}^Q(d\zeta, dt),
\end{aligned}$$

where $dW_t^Q = dW_t - \phi_z dt$ and $\tilde{\mu}^Q(d\zeta,dt) = \tilde{\mu}(d\zeta,dt) - \phi_\theta\nu(d\zeta)dt$. We know that $\|\phi_z\| \le k(1+\|Z\|)$, $|\phi_\theta(\zeta)| \le k(1+|\theta(\zeta)|)$ and $\phi_\theta > -1+\epsilon$. Therefore, $\phi_z^\top \bullet W + \phi_\theta * \tilde{\mu}$ is a BMO martingale, and $dQ/dP = \mathcal{E}(\phi_z^\top \bullet W + \phi_\theta * \tilde{\mu})_T$ defines an equivalent measure Q. By Theorem 15.2.8, W^Q and $\tilde{\mu}^Q$ have the predictable representation property under Q. Therefore, by Lemma A.9.9, there exists[1] a solution $(Y^\delta, Z^\delta, \Theta^\delta)$, and hence a solution (Y', Z', Θ'). Uniqueness again follows from the comparison theorem. □

Lemma A.9.13. *Consider a BSDE satisfying the conditions of Theorem A.9.5 and suppose f is twice continuously Fréchet differentiable with bounded second derivative. Then the BSDE admits a (unique) solution in $S^\infty \times L^{\mathrm{BMO}}(\langle W\rangle) \times L^{\mathrm{BMO}}(\langle\tilde{\mu}\rangle)$.*

[1]This argument is slightly problematic, as so far we have only considered BSDEs in the case when the compensator of the random measure is deterministic. However, all our previous results can be adapted to allow for the compensator to be stochastic, with a resultant increase in notational complexity. By Theorem 15.2.8, the predictable representation property holds under the new measure, and the key is to reweight the norm $\|\cdot\|_\nu$ by the Radon–Nikodym density of the compensator with respect to a deterministic reference measure $\nu(d\zeta)dt$. We leave the details to the interested reader.

Proof. Let k be a bound on $\|Df(0,0)\| + \sup_{z,\theta}\|D^2 f(z,\theta)\|$. We begin by setting (Y^0, Z^0, Θ^0) as the solution to the BSDE with terminal condition $Y^0_T = 0$ and driver $f(0,0)$. Now fix $n > 256k\|\xi\|_{L^\infty}$. As our desired terminal condition ξ is bounded, we know that $\xi^i := \frac{1}{n}\xi$ satisfies the bound $\|\xi^i\|_{L^\infty} < 1/256k$, for each $i \leq n$. We then recursively define the BSDE drivers

$$f^i(z,\theta) = f(z + Z^{i-1}_t, \theta + \Theta^{i-1}_t) - f(Z^{i-1}_t, \Theta^{i-1}_t),$$

where (Y^i, Z^i, Θ^i) is the solution to the BSDE with data $(f^i, \frac{i}{n}\xi)$. The solutions (Y^i, Z^i, Θ^i) exist by induction, using Lemma A.9.12. Finally, as

$$\sum_{i\leq n} f^i(Z^i, \Theta^i) = f\Big(\sum_{i\leq n} Z^i, \sum_{i\leq n}\Theta^i\Big) \quad \text{and} \quad \sum_{i\leq n}\xi^i = \xi,$$

we see that

$$(Y, Z, \Theta) := \Big(\sum_{i\leq n} Y^i, \sum_{i\leq n} Z^i, \sum_{i\leq n}\Theta^i\Big)$$

is a solution to the BSDE with data (f, ξ). Uniqueness follows from Theorem A.9.6. □

Proof of Theorem A.9.5. By assumption, it is possible to approximate f uniformly by a sequence of twice Fréchet differentiable ϵ-balanced quadratic drivers with bounded second derivatives. Each of these approximations gives a solution to the BSDE, by the previous lemma. The continuity of Theorem A.9.7 then implies that the solutions to these approximations converge in $S^p \times L^{\mathrm{BMO}}(\langle W\rangle) \times L^{\mathrm{BMO}}(\langle\tilde{\mu}\rangle)$ for some $p > 1$. Boundedness of Y and uniqueness of the solution follows from Theorem A.9.6. □

A.9.2 Stochastic Lipschitz BSDEs

Another class of BSDEs which arises in applications, particularly in control problems, is when the Lipschitz constant of the driver f is itself a (potentially unbounded) stochastic process.

Example A.9.14. At least in the linear case, Theorem 15.2.8 gives us a simple approach. For simplicity of exposition, consider the case where a single Brownian motion W has the predictable representation property. For a predictable process α, consider the BSDE given by

$$dY_t = -(\alpha Z_t)dt + Z_t dW_t, \qquad Y_T = \xi$$

for $\xi \in L^\infty(\mathcal{F}_T)$. Then we can rewrite the BSDE in the form

$$dY_t = Z_t(dW_t - \alpha dt) = Z_t d\tilde{W}, \qquad Y_T = \xi.$$

Assuming that $\mathcal{E}(\alpha \bullet W)$ is a true martingale (which does not require that α is bounded), we know from Girsanov's theorem that $\tilde{W}$ is a Q-Brownian motion, where $dQ/dP = \mathcal{E}(\alpha \bullet W)_T$. Theorem 15.2.8 guarantees that $\tilde{W}$ has the predictable representation property, so there exists a Z such that $Y_t = E^Q[\xi|\mathcal{F}_t]$ and $dY_t = Z_t d\tilde{W}_t$. Therefore, for bounded terminal values, we know that this linear BSDE admits a solution. As Q is the same for all ξ, it is easy to establish that the solution is unique and the comparison theorem must hold.

Our aim is to extend this example to the case of a nonlinear driver.

Definition A.9.15. *We say that a driver f for a BSDE is stochastic Lipschitz if there is a predictable process K such that, for all $z, z' \in \mathbb{R}^N$ and $\theta, \theta' \in L^2(\nu)$,*

$$|f(\omega,t,z,\theta) - f(\omega,t,z',\theta')| \leq K_t(\omega)(\|z-z'\| + \|\theta-\theta'\|_\nu) \qquad dt \times dP - \text{a.e.}$$

We again say that it is balanced if it satisfies Definition 19.3.1.

Note that we exclude dependence of f on y, purely for the sake of simplicity.

Theorem A.9.16. *Let (Y, Z, Θ) be a solution to a BSDE with a balanced, stochastic Lipschitz driver. Suppose one of the following integrability conditions holds.*

(i) $E\big[\exp\big(2\int_{[0,T]} K_t^2 dt\big)\big] < \infty$.

(ii) For X a martingale (or nonnegative submartingale) with $E[e^{a\|X_T\|^2}] < \infty$ for some $a > 0$, for some $k > 0$, we have $K_t^2 \leq k(1 + X_t^*)$.

(iii) For X a martingale (or nonnegative submartingale) with $E[e^{a\|X_T\|}] < \infty$ for some $a > 0$, for some $k > 0$, we have $K_t^2 \leq k\big(1 + (X_t^*)^{1/2}\big)$.

Then $Y_t = E^Q[\xi|\mathcal{F}_t]$ for some probability measure Q with density process $\Lambda \in \mathcal{M}$.

Proof. Define $\Lambda = \mathcal{E}(\alpha^\top \bullet W + \beta * \tilde{\mu})$ and $dQ/dP = \Lambda_T$, where, for notational simplicity,

$$\alpha_t := \frac{f(\omega,t,Z_t,0) - f(\omega,t,0,0)}{\|Z_t\|^2}(Z_t)^\top,$$
$$\beta_t(\zeta) = \beta(\zeta;\omega,t,Z_t,\Theta_t,0).$$

From the stochastic Lipschitz conditions on f, we know

$$\frac{d}{dt}\langle \alpha^\top \bullet W + \beta * \tilde{\mu}\rangle = \|\alpha_t\|^2 + \|\beta_t\|_\nu^2 \leq 2(K_t(\omega))^2.$$

Then, we see that case (i) corresponds to Novikov's condition (Corollary 15.4.4), case (ii) corresponds to the conditions of Example 15.5.6, while case (iii) corresponds to the conditions of Example 15.5.9. In any of these cases, it follows that Λ is a uniformly integrable martingale. Bayes' rule (Exercise 5.7.1) then implies that

$$Y_t = E^Q[\xi|\mathcal{F}_t].$$

□

Remark A.9.17. The constant 2 in condition (i) of Theorem A.9.16 is not particularly tight. In the continuous case, where $\nu \equiv 0$, using Theorem 15.4.6(ii) we can weaken condition (i) to

$$|f(\omega, t, z) - f(\omega, t, z')| \leq K_t(\omega)\|z - z'\|$$

and $E\big[\exp\big(\frac{1}{2}\int_{[0,T]} K_t^2 dt\big)\big] < \infty$. Further weakening is also possible using the argument of Example 15.5.4.

Remark A.9.18. Under the conditions of Theorem A.9.16, we see from Jensen's inequality that $E\big[\int_{[0,T]} |K_t|^p dt\big] < \infty$ for every $p < \infty$.

Corollary A.9.19. *Given the conditions of Theorem A.9.16, suppose (Y, Z, Θ) solves a BSDE with data (f, ξ), and (Y', Z', Θ') a BSDE with data (f', ξ'). If ξ is bounded, then we also know $Z \bullet W + \Theta * \tilde{\mu} \in \mathcal{H}^2$, with a bound depending only on K and $\|\xi\|_\infty$. If ξ and ξ' are both bounded, we have the estimate*

$$\begin{aligned}
&E\Big[\int_{]0,T]} (\|\delta Z_s\|^2 + \|\delta\Theta_s\|_\nu^2) ds\Big] \\
&\leq E\big[(\xi - \xi')^2\big] + CE\Big[\int_{]0,T]} |Y_s - Y_s'|^3 ds\Big]^{1/3} + 2E\Big[\int_{]0,T]} |Y_s - Y_s'||\phi_s| ds\Big],
\end{aligned}$$

for C a constant depending only on K, $\|\xi\|_\infty$, $\|\xi'\|_\infty$ and T, where $\phi_t = f(\omega, t, Z', \Theta') - f'(\omega, t, Z', \Theta')$.

Proof. If ξ is bounded, then so is $Y_t = E^Q[\xi|\mathcal{F}_t]$, and $\|Y_T^*\|_\infty = \|\xi\|_\infty$. Using Young's inequality,

$$f(\omega, t, y, z, \theta) \leq \frac{(K_t(\omega))^2}{2} + \frac{3}{2}(1 + \|z\|^2 + \|\theta\|_\nu^2).$$

The same arguments as in Theorem A.9.1, with $\tau = 0$, yield

$$\begin{aligned}
&9e^{-6\|Y_T^*\|_\infty} E\Big[\int_{]0,T]} \|Z_t\|^2 dt + \int_{[0,T_n]} (\Theta_t(\zeta))^2 \mu(d\zeta, dt)\Big] \\
&\leq e^{6\|Y_T^*\|_\infty} + 9E\Big[\int_{[0,T]} e^{6Y_{t-}}\Big(1 + \frac{(K_t(\omega))^2}{2} + |Y_{t-}|\Big) dt\Big] < \infty,
\end{aligned}$$

and so, using the BDG inequalities, we see that $Z \bullet W + \Theta * \tilde{\mu} \in \mathcal{H}^2$, with a bound depending only on K and $\|Y_T^*\|_\infty = \|\xi\|_\infty$. Similarly for (Y', Z', Θ').

Writing $\delta Y = Y - Y'$, and similarly for δZ and $\delta\Theta$, taking (A.12) with $t = 0$, we obtain the estimate

$$\begin{aligned}
&E\Big[\int_{]0,T]} (\|\delta Z_s\|^2 + \|\delta\Theta_s\|_\nu^2) ds\Big] - E[(\xi - \xi')^2] \\
&\leq 2E\Big[\int_{]0,T]} |\delta Y_s|\,|f(Z_s, \Theta_s) - f(Z_s', \Theta_s')| ds\Big] + 2E\Big[\int_{]0,T]} |\delta Y_s||\phi_s| ds\Big] \\
&\leq 2E\Big[\int_{]0,T]} |\delta Y_s| K_s(\|\delta Z_s\| + \|\delta\Theta_s\|_\nu) ds\Big] + 2E\Big[\int_{]0,T]} |\delta Y_s||\phi_s| ds\Big].
\end{aligned}$$

Expanding using Hölder's, Young's and Jensen's inequalities,

$$
\begin{aligned}
&E\Big[\int_{]0,T]} |\delta Y_s| K_s\,(\|\delta Z_s\| + \|\delta\Theta_s\|_\nu) ds\Big] \\
&\leq E\Big[\int_{]0,T]} |\delta Y_s|^3 ds\Big]^{1/3} E\Big[\int_{]0,T]} K_t^{3/2}(\|\delta Z_s\| + \|\delta\Theta_s\|_\nu)^{3/2} ds\Big]^{2/3} \\
&\leq E\Big[\int_{]0,T]} |\delta Y_s|^3 ds\Big]^{1/3} E\Big[\int_{]0,T]} \Big(\frac{K_t^6}{4} + \frac{(\|\delta Z_s\| + \|\delta\Theta_s\|_\nu)^2}{4/3}\Big) ds\Big]^{3/2} \\
&\leq CE\Big[\int_{]0,T]} |\delta Y_s|^3 ds\Big]^{1/3},
\end{aligned}
$$

where C is a constant depending only on K, $\|\xi\|_\infty$, $\|\xi'\|_\infty$ and T. Combining these inequalities gives the desired estimate. □

We can now prove that these BSDEs have unique solutions, using an argument based on that of Hamadène and Lepeltier [93].

Theorem A.9.20. *Consider a BSDE with bounded terminal condition $\xi \in L^\infty(\mathcal{F}_T)$, and balanced stochastic Lipschitz driver f such that $\int_{[0,T]} |f(\omega, t, 0, 0)| dt \in L^\infty(\mathcal{F}_T)$ and the conditions of Theorem A.9.16 are satisfied. Then there exists a unique bounded solution to the BSDE and the comparison theorem holds.*

Proof. For each $n, m \in \mathbb{N}$, We define the approximation to the driver

$$
\begin{aligned}
f^{n,m} &= I_{\{K_t<n\}} f^+ - I_{\{K_t<m\}} f^- \\
&= (I_{\{K_t<n\}} f) \wedge f \vee (I_{\{K_t<m\}} f).
\end{aligned}
$$

From the first representation of $f^{n,m}$ it is clear that $f^{n,m}$ is Lipschitz with constant $n \vee m$, is increasing in n and decreasing in m. From Lemma 19.3.8 and the second representation, we see that $f^{n,m}$ is balanced, for every n, m, so the comparison theorem holds for these approximations.

As $f^{n,m}$ is uniformly Lipschitz, the BSDE with data $(\xi, f^{n,m})$ has a solution, denoted $(Y^{n,m}, Z^{n,m}, \Theta^{n,m})$. As $f^{n,m}$ is balanced, we see from the comparison theorem that $Y^{n,m}$ is increasing in n and decreasing in m, therefore a diagonal argument yields a sequence Y^{n_k,m_k} which converges pointwise, for all t and almost all ω. As $Y^{n,m}$ is uniformly bounded, dominated convergence implies that Y^{n_k,m_k} is a Cauchy sequence in $L^p([0,T] \times \Omega)$ for every $p \in [1, \infty[$.

We also know that, for any n, n', m, m', if $k = n \wedge n' \wedge m \wedge m'$ then

$$
\begin{aligned}
&E\Big[\Big(\int_{[0,T]} |Y_s^{n,m} - Y_s^{n',m'}||f^{n,m}(Z_s^{n,m}, \Theta_s^{n,m}) - f^{n',m'}(Z_s^{n,m}, \Theta_s^{n,m})|ds\Big)\Big] \\
&\leq E\Big[\int_{[0,T]} |Y_s^{n,m} - Y_s^{n',m'}|^3 ds\Big]^{1/3} \\
&\qquad \times E\Big[\int_{[0,T]} I_{\{K_t>k\}} K_t^{3/2}(1 + \|Z_s^{n,m}\| + \|\Theta_s^{n,m}\|)^{3/2} ds\Big]^{3/2} \\
&\leq E\Big[\int_{[0,T]} |Y_s^{n,m} - Y_s^{n',m'}|^3 ds\Big]^{1/3} \\
&\qquad \times E\Big[\int_{[0,T]} I_{\{K_t>k\}} \Big(\frac{K_t^4}{4} + \frac{(1 + \|Z_s^{n,m}\| + \|\Theta_s^{n,m}\|)^2}{4/3} ds\Big)\Big]^{3/2}.
\end{aligned}
$$

Therefore, the estimates of Corollary A.9.19 imply that $(Z^{m_k,n_k}, \Theta^{m_k,n_k})$ also converges in $L^2(\langle W\rangle) \times L^2(\langle\tilde{\mu}\rangle)$ as $k \to \infty$, and hence pointwise almost everywhere for a subsequence. Writing (Z, Θ) for the limit, as $f^{m,n} \to f$ locally uniformly, we see that, at least for a subsequence,

$$
f^{m_k,n_k}(Z_s^{m_k,n_k}, \Theta_s^{m_k,n_k}) \to f(Z_s, \Theta_s) \qquad dt \times dP - \text{a.e.}
$$

We therefore have the existence of a triple (Y, Z, Θ) which solves the BSDE for almost all (t, ω). It follows that Y has a right continuous modification, which solves the BSDE for all t, for almost all ω. Uniqueness follows from the comparison theorem, which is a direct consequence of Theorem A.9.16. □

Remark A.9.21. It is also clear that the argument of Theorem A.9.7 will hold, which gives a useful continuity estimate on the solutions of these BDSEs.

Finally, in the Markovian case, for either quadratic or stochastic Lipschitz drivers, we have the following version of Theorem 19.5.1.

Theorem A.9.22. *Suppose ξ and f depend on ω only through the value of a 'forward' Markov process X with infinitesimal generator $\mathcal{L}_t$ of the form in Definition 17.4.1. Let v be a* bounded *$C_\nu^{1,2}$ function and suppose that v is a solution to the following semilinear parabolic PIDE*

$$
\begin{cases}
\quad 0 = \dfrac{\partial v}{\partial s}(s,x) + \mathcal{L}_t v(s,x) + f(s, x, v(s,x), \partial_x v(s,x)\sigma(s,x), \tilde{v}(s,x)), \\
v(T,x) = \psi(x),
\end{cases}
\tag{A.14}
$$

where $\tilde{v}(s,x)$ denotes the element of $L^2(\nu)$ given by the map

$$
\zeta \mapsto v\big(s, x + g(\zeta, s, x)\big) - v(s,x).
$$

Then

$$\begin{aligned} Y_s^{(t,x)} &= v(s, X_s^{(t,x)}), \\ Z_s^{(t,x)} &= \partial_x v(s, X_s^{(t,x)})\sigma(s, X_s^{(t,x)}), \\ \Theta_s^{(t,x)}(\zeta) &= \tilde{v}(\zeta; t, X_s^{(t,x)}) = v\big(s, X_s^{(t,x)} + g(\zeta, s, X_s^{(t,x)})\big) - v(s, X_s^{(t,x)}), \end{aligned}$$

where $(Y^{(t,x)}, Z^{(t,x)}, \Theta^{(t,x)})$ *is the unique solution of the BSDE (19.7) and the equalities are in* S^∞, $L^2(\langle W \rangle)$ *and* $L^2(\langle \tilde{\mu} \rangle)$ *respectively. In particular,*

$$Y_t^{(t,x)} = v(t, x).$$

Proof. As in Theorem 19.5.1, we simply apply Ito's rule to v, and see that $Y_t = v(t, X_t)$ solves the BSDE. As v is bounded, the solution lies in the relevant space (by Theorem A.9.1 or Corollary A.9.19). Uniqueness of solutions then yields the result. □

Remark A.9.23. To show that the solution of the BSDE is a viscosity solution of the PDE, as least in the continuous setting, we can proceed in much the same way as in Theorem 19.5.3.

For the stochastic Lipschitz case with $K_t \leq k(1 + \|X_t\|)$ or $K_t \leq k(1 + \|X_t\|^{1/2})$, for X the forward process, we simply begin by localizing in x at the first step. We then have a uniformly Lipschitz BSDE, and the result follows as before.

In the quadratic setting, a detailed proof is given by Kobylanski [119].

A.10 Filippov's Implicit Function Lemma

When considering optimal control, we typically obtain the optimizer in an implicit form, that is, for some function H we know that the optimal u^* is the minimizer of $H(\omega, t, u)$. This raises the question of whether this minimum can be achieved with a *measurable* control $u : \Omega \times [0, \infty] \to U$. This is a significant point, as otherwise the equations considered often do not have meaningful solutions.

Example A.10.1. To see that a non-measurable optimizer is indeed possible, consider the simple function $H(t, u) = 1 - u^2$ defined for u in $[-1, 1]$. Then for every t, the values $u = \pm 1$ will minimize H. Let A be a non-measurable set in time, then the non-measurable process $u(t) = 2I_{\{t \in A\}} - 1$ is a non-measurable minimizer of H.

We will now show that, given fairly weak conditions on the terms involved, it is possible to select a minimizer which is measurable in a useful way. The key step is the following theorem, which is a variation due to Beneš [10] of a result of McShane and Warfield [130], which extended the fundamental result of Filippov [78]. (Our proof is from [10] and [130].)

Theorem A.10.2. *Let $(M, \mathcal{M})$ be a measurable space, A a separable metric space and U a topological space which is the union of countably many compact metrizable subsets of itself. Consider a function $k : M \times U \to A$ such that $k(\cdot, u)$ is $\mathcal{M}$-measurable for every $u \in U$ and $k(x, \cdot)$ is continuous for every $x \in M$. Let $y : M \to A$ be an $\mathcal{M}$-measurable map with*

$$y(x) \in k(x, U) \quad \text{for all } x \in M.$$

Then there exists an $\mathcal{M}$-measurable map $u : M \to U$ such that

$$y(x) = k(x, u(x)).$$

Proof. We present the proof as a series of cases. As we shall see, Case 1 is the key step, as the general case can be reduced to this.

Case 1: Suppose that $U = L$, for L a closed subset of $[0, \infty[$. As one might expect, our strategy is to partition L, define an approximation of u on this partition, then let the partition converge.

Our first task, to allow us to define the partition, is to show that for any compact C, we have

$$\zeta(C) := \{x : y(x) \in k(x, L \cap C)\} \in \mathcal{M}.$$

Let π_m be a countable cover of A by open sets of diameter $\leq 2^{-m}$. Then we claim that

$$\zeta(C) = \bigcap_m \bigcup_{S \in \pi_m} \{x : y(x) \in S \text{ and } k(x, L \cap C) \cap S \neq \emptyset\}. \tag{A.15}$$

To show this, we see that if $x \in \zeta(C)$ then $y(x) \in k(x, L \cap C)$. For each m, there is an open set $S \in \pi_m$ with $y(x) \in S$, so that $y(x) \in k(x, L \cap C) \cap S \neq \emptyset$. Conversely, if x is in the set on the right of (A.15), then for every m there is a set S of diameter $\leq 2^{-m}$ such that $y(x) \in S$ and $k(x, L \cap C) \cap S \neq \emptyset$. Therefore, $y(x)$ is a distance at most 2^{-m} away from the set $k(x, L \cap C)$. Since this is true for every m, we know $y(x)$ is in the closure of $k(x, L \cap C)$. As $k(x, \cdot)$ is continuous and $L \cap C$ is compact, we know $k(x, L \cap C)$ is closed, so we see $y(x) \in k(x, L \cap C)$, that is, $x \in \zeta(C)$.

It is apparent that, for S open,

$$\{x : k(x, L \cap C) \cap S \neq \emptyset\} = \bigcup_{u \in L \cap C} \{x : k(x, u) \in S\}.$$

Each set in the possibly uncountable union on the right belongs to $\mathcal{M}$. We show that a countable union can be used, which ensures the term on the right is in $\mathcal{M}$. Let D be a countable dense set in $L \cap C$, and $\{u_n\}_{n \in \mathbb{N}} \subseteq D$ satisfy $u_n \to u$, where u is chosen such that $k(x, u) \in S$. Then $k(x, u_n) \to k(x, u)$, by continuity of $k(x, \cdot)$. Therefore, $k(s, u_n) \in S$ for all n sufficiently large, as S is open.

Therefore, for some n depending on S, u, and x, we know that

$$\begin{aligned} k(x,u) \in S &\Rightarrow x \in \{w : k(w,u_n) \in S\} \\ &\Rightarrow x \in \bigcup_m \{w : k(w,u_m) \in S\} \subseteq \bigcup_{u \in D} \{x : k(x,u) \in S\}. \end{aligned}$$

It follows that $\bigcup_{u \in L \cap C} \{x : k(x,u) \in S\} = \bigcup_{u \in D} \{x : k(x,u) \in S\}$, that is, the union can be taken over a countable set. Therefore, for S open,

$$\{x : k(x, L \cap C) \cap S \neq \emptyset\} \in \mathcal{M}.$$

As y is $\mathcal{M}$-measurable, $\{x : y(x) \in S\} \in \mathcal{M}$, so by (A.15) we conclude that $\zeta(C) \in \mathcal{M}$.

Given this measurability result, we now define a partition of L. For $q, j \in \mathbb{N}$, we consider the set

$$B_j^q := \big\{x : y(x) \in \big(k(x, L \cap [0, j2^{-q}]) \setminus k(x, L \cap [0, (j-1)2^{-q}])\big)\big\} \in \mathcal{M}.$$

On B_j^q, we define the approximation

$$u_q(x) := \inf \big\{L \cap](j-1)2^{-q}, j2^{-q}]\big\}$$

and, for notational convenience, set $j_q(x) = j$, noting that

$$L \cap](j-1)2^{-q}, j2^{-q}] = \emptyset \text{ implies } B_j^q = \emptyset.$$

We know u_q and j_q are $\mathcal{M}$-measurable functions, as $\{x : j_q(x) = j\} = B_j^q \in \mathcal{M}$. Note that $j_{q+1}(x) \in \{2j_q(x), 2j_q(x) - 1\}$. Thus

$$\begin{aligned} &u_{q+1}(x) \\ &= \inf \big\{u : u \in L \cap](j_{q+1}(x) - 1)2^{-(q+1)}, j_{q+1}(x)2^{-(q+1)}]\big\} \\ &= \begin{cases} \inf \big\{L \cap](j_q(x) - 1/2)2^{-q}, j_q(x)2^{-q}]\big\} & \text{if } j_{q+1}(x) = 2j_q(x), \\ \inf \big\{L \cap](j_q(x) - 1)2^{-q}, (j_q(x) - 1/2)2^{-q}]\big\} & \text{if } j_{q+1}(x) = 2j_q(x) - 1, \end{cases} \\ &\geq u_q(x). \end{aligned}$$

Therefore, $u_q(x)$ is increasing in q. We show that $u_q(x)$ is bounded. Suppose $x \in B_j^0$, so that $j_0(x) = j$, then

$$u_0(x) = \inf \big\{L \cap](j-1), j]\big\} \leq j.$$

Then $j_1(x) = 2j$ or $2j - 1$, and

$$\begin{aligned} u_1(x) &= \begin{cases} \inf \big\{L \cap](2j - 1/2)2^{-1}, (2j)2^{-1}]\big\} & \text{if } j_1(x) = 2j, \\ \inf \big\{L \cap](2j - 1)2^{-1}, (2j - 1/2)2^{-1}]\big\} & \text{if } j_1(x) = 2j - 1, \end{cases} \\ &\leq j. \end{aligned}$$

In general, the same argument yields

$$u_{q+1}(x) \leq (2j_q)2^{-(q+1)} \leq j,$$

so $u_q \uparrow u$ for some ($\mathcal{M}$-measurable) map u. For all x, we have $u_q(x) \in L$, and so $\{u(x)\}_{x \in M} \subseteq L$, as L is closed.

Given that we have defined the limit u, we now show that $y(x) = k(x, u(x))$, which will establish the result in the case $U = L$. To establish a contradiction, suppose for the moment that

$$y(x) \neq k(x, u(x)). \tag{A.16}$$

Then there exists an $x \in M$ and a neighbourhood V of $k(x, u(x))$ such that $y(x) \notin V$ (as A is a separable metric space, and therefore Hausdorff). We know $k(x, \cdot)$ is continuous, so the preimage $k(x, \cdot)^{-1}V$ includes a neighbourhood of $u(x)$. Hence there exist j and q such that

$$u(x) \in L \cap](j-1)2^{-q}, j2^{-q}] \subseteq L \cap [(j-1)2^{-q}, j2^{-q}] \subseteq k(x, \cdot)^{-1}V. \tag{A.17}$$

It follows that $k(x, L \cap [(j-1)2^{-q}, j2^{-q}]) \subseteq V$, and so

$$\big(k(x, L \cap [0, j2^{-q}]) \setminus k(x, L \cap [0, (j-1)2^{-q}])\big) \subseteq V.$$

For this j and q, we have $u_q(x) \in L$ and

$$u_q(x) \leq u(x) \leq j2^{-q},$$

so $j_q(x) \leq j$. Suppose $j_q(x) \leq j-1$. Then

$$j_{q+1}(x) \leq 2j_q(x) \leq 2(j-1), \qquad \text{and } j_{q+n} \leq 2^n(j-1).$$

But $u_{q+n} \leq j_{n+q}(x)2^{-(q+n)}$, so

$$u_{q+n}(x) \leq 2^n(j-1)2^{-(q+n)} = (j-1)2^{-q}.$$

Therefore, $u(x) \leq (j-1)2^{-q}$. This contradicts the assumption of (A.17) that $u(x) \in](j-1)2^{-q}, j2^{-q}] \cap L$. Therefore, $j_q(x) = j$, that is, $x \in B^q_j$, or

$$y(x) \in \big(k(x, L \cap [0, j2^{-q}]) \setminus k(x, L \cap [0, (j-1)2^{-q}])\big) \subseteq V,$$

which contradicts $y(x) \notin V$. Therefore, it must be the case that $y(x) = k(x, u(x))$, that is, (A.16) cannot hold.

Case 2: We now let U be any space such that there is a closed subset L of $[0, \infty[$ and a continuous surjective map $\phi : L \to U$. By Case 1, there is a measurable function $T : M \to L$ such that $y(x) = k(y, \phi(T(x)))$. We set $u = \phi \circ T$, and need only to verify that u is $\mathcal{M}$-measurable. If F is a closed subset of U, then $\phi^{-1}(F)$ is closed, and so $\{x : T(x) \in \phi^{-1}(F)\} \in \mathcal{M}$. Thus u is $\mathcal{M}$-measurable.

Case 3: We now prove the theorem as stated. Let $\{K_n\}_{n\in\mathbb{N}}$ be a sequence of compact metrizable sets with $\bigcup_n K_n = U$. Since every compact metric space is the continuous image of the Cantor set (see Hocking and Young [95, Theorem 3.28]), for each positive integer n there is a closed subset L_n of $[2n-1, 2n]$ (which contains a translate of the Cantor set) and a continuous function $\phi_n : L_n \to K_n$ whose range is K_n. Define $L = \cup_n L_n$, and define ϕ to be the function on L which coincides with ϕ_n on L_n. Then we are in the setting of Case 2, and the proof is complete. □

We can now prove the following useful result (Theorem 21.3.4 in the main text).

Theorem A.10.3. *Let U be a topological space which is the union of countably many compact metrizable subsets of itself and X be a separable metric space. Let $G : \Omega \times [0,\infty[\times X \times U \to \mathbb{R}$ be such that*

(i) $G(\cdot,\cdot,x,u)$ is Σ_p-measurable (i.e. predictable), for every $u \in U$, $x \in X$,
(ii) $G(\omega,t,\cdot,u)$ is continuous, uniformly in $u \in U$, for $dP \times dt$-almost all (ω,t),
(iii) $G(\omega,t,x,\cdot)$ is continuous, for $dP \times dt$-almost all (ω,t) and all $x \in X$,
(iv) $\operatorname{ess\,inf}_{u\in U} G(\omega,t,x,u) > -\infty$ for $dP \times dt$-almost all (ω,t) and all $x \in X$,

where in (iv), the essential infimum is taken in the predictable processes, and defined $dP\times dt$-a.e. Then, for every $\epsilon > 0$, there exists a $\Sigma_p \otimes \mathcal{B}(X)$-measurable function $u^\epsilon(\omega,t,x)$ taking values in U such that, for every x,

$$G(\omega,t,x,u^\epsilon(\omega,t,x)) < \operatorname{ess\,inf}_{u\in U} G(\omega,t,x,u) + \epsilon \qquad dt \times dP - a.e.$$

If we also know that

(v) for $dP \times dt$-almost all (ω,t) and all $x \in X$, there exists $v \in U$ such that $G(\omega,t,x,v) = \operatorname{ess\,inf}_{u\in U} G(\omega,t,x,u)$,

then there exists a predictable, $\mathcal{B}(X)$-measurable function $u^* : \Omega\times\mathbb{R}^+\times X \to U$ such that

$$G(\omega,t,x,u^*(\omega,t,x)) = \operatorname{ess\,inf}_{u\in U} G(\omega,t,x,u) \qquad dt \times dP - a.e.$$

The functions $G(\omega,t,x,u^\epsilon(\omega,t,x))$ (and $G(\omega,t,x,u^*(\omega,t,x))$, when defined) have the same modulus of continuity with respect to x as G.

Proof. For x in a countable dense subset of X, define

$$y(\omega,t,x) = \operatorname{ess\,inf}_{u\in U} G(\omega,t,x,u),$$

the infimum being taken in the predictable processes and defined $dP \times dt$-a.e. For each x in our dense set, we see that the essential infimum equals a pointwise infimum over a countable subset of $\mathcal{U}$, and hence is measurable. By the continuity of G with respect to x, we can find a version of y which is

$dP \times dt$-a.e. continuous with respect to x, and hence uniquely extends to a continuous function on all of X, and this function is a version of the essential infimum.

Suppose (i - v) hold. By (v), we know $y(\omega, t, x) \in G(\omega, t, x, U)$, after possibly modifying on a $dP \times dt$-null predictable set. For each $x \in X$, applying Theorem A.10.2 with $(M, \mathcal{M}) = (\Omega \times [0, \infty[\times X, \Sigma_p \otimes \mathcal{B}(X))$, we have the existence of a $\Sigma_p \otimes \mathcal{B}(X)$-measurable map u^* with $G(\omega, t, x, u^*(\omega, t, x)) = y(\omega, t, x)$, $dt \times dP$-a.e. As y is continuous with respect to x, we see that $G(\omega, t, x, u^*(\omega, t, x))$ is also continuous with respect to x (with the same modulus of continuity).

Now suppose only (i - iv) hold. Then let y be defined as before, and define the function

$$G^\epsilon(\omega, t, x, u) = \begin{cases} G(\omega, t, x, u) & \text{if } G(\omega, t, x, u) - y(\omega, t) > 2\epsilon, \\ \lambda_\epsilon(G(\omega, t, x, u), y(\omega, t)) & \text{if } G(\omega, t, x, u) - y(\omega, t, x) \in [\epsilon, 2\epsilon], \\ y(\omega, t, x) & \text{if } G(\omega, t, x, u) - y(\omega, t, x) < \epsilon. \end{cases}$$

where $\lambda_\epsilon(a, b) = (a - b - \epsilon)(a - b)/\epsilon + b$. As this is a continuous interpolation between G and y, it is clearly predictable in (ω, t) and $dP \times dt$-a.e. continuous in x and u. We know that

- $G - G^\epsilon < \epsilon$ for $dP \times dt$-almost all (ω, t) and all (x, u),
- $\operatorname{ess\,inf}_u G^\epsilon = \operatorname{ess\,inf}_u G$ for $dP \times dt$-almost all (ω, t) and all $x \in X$ and
- G^ϵ attains its minimum with respect to u, for all x, $dt \times dP$-a.e.

Therefore, using the previous result, there exists a predictable process u^ϵ such that $G^\epsilon(\omega, t, u^\epsilon_t) = \operatorname{ess\,inf}_u G^\epsilon(\omega, t, u)$, and hence

$$G(t, \omega, u^\epsilon_t) < G^\epsilon(t, \omega, u^\epsilon_t) + \epsilon = \operatorname{ess\,inf}_{u \in U} G(\omega, t, u) + \epsilon.$$

$\square$

Remark A.10.4. If we know that U is a *compact* metrizable set, then the continuity assumption (iii) guarantees immediately that the infimum is attained as stated in (v).

Theorem A.10.5. *(i) For each $n \in \mathbb{N}$, let $f^n : X \to A$ be a measurable function, where $(X, \mathcal{X})$ is a measurable space and A is a compact metrizable set. Then there exists a measurable function $g : X \to A$ such that $g(x)$ is a limiting value of $\{f^n(x)\}_{n \in \mathbb{N}}$ for every $x \in X$ (i.e. for each value of x, a subsequence of $f^n(x)$ converges to $g(x)$).*

(ii) The above result also holds if A is a closed, convex, bounded subset of a separable Hilbert space, endowed with the weak topology.

Proof. First consider case (i). Recall that every compact metrizable space is separable (see [160, p.204]). Let $K(x)$ denote the set of limiting values of $\{f^n(x)\}_{n \in \mathbb{N}}$, which is a measurable subset of A. For $\epsilon > 0$, taking $U = \mathbb{N}$ with the discrete topology and applying Theorem A.10.2 with $y(x) = 0$ and

$$k^\epsilon(x,n) = \Big(\inf_{a\in K(x)}\{\|f^{n(x)}(x) - a\|\} - \epsilon\Big)^+,$$

we see that there exists a measurable function $u : X \to \mathbb{N}$ such that $\inf_{a\in K(x)}\{\|f^{u(x)}(x) - a\|\} \leq \epsilon$ for all x. For $\epsilon = 1$, we use this to define the function $n_1 = u$.

Repeating this argument, we can find a measurable function $n_2(x)$ such that, for all x,

$$\inf_{a\in K(x)}\Big\{\|f^{n_2(x)}(x) - a\|\Big\} < 1/2 \text{ and } \|f^{n_2(x)}(x) - f^{n_1(x)}(x)\| < 2.$$

(By the triangle inequality we see that the relevant sets overlap for every x, so this construction is possible.) By induction, we then construct a sequence of functions n_m such that, for all x,

$$\inf_{a\in K(x)}\{\|f^{n_m(x)}(x) - a\|\} < 2^{-m} \text{ and } \|f^{n_m(x)}(x) - f^{n_{m-1}(x)}(x)\| < 2^{-m+2}.$$

From the second property, we see that $f^{n_m(x)}(x)$ converges, uniformly in x, and so we define $g(x) = \lim_m f^{n_m(x)}(x)$, which is a measurable function. Furthermore,

$$\inf_{a\in K(x)}\{\|g(x) - a\|\} \leq 2^{-m} \to 0$$

so $g(x)$ takes values in the closed set $K(x)$, as desired.

To prove (ii), we recall that the Banach–Saks theorem states that, for each x, if $\{f^n(x)\}_{n\in\mathbb{N}}$ is a sequence taking values in A, then there exists a subsequence $f^{n_k(x)}(x)$ such that the arithmetic means $\frac{1}{m}\sum_{k=1}^m f^{n_k(x)}(x)$ converge strongly. Furthermore, a standard proof of this (see, for example, Royden and Fitzpatrick [160, p.175]) allows the subsequence to be explicitly constructed, in particular, $n_k(x)$ is measurable in x. The result then follows from (i) applied to the sequence of arithmetic means. □

Remark A.10.6. Many further results in this direction are possible, under the generic heading of 'measurable selection theorems'. For a classic survey of this area, see Wagner [180, 181].

B

Spaces of Càdlàg Adapted Processes

This appendix presents Figure B.1, which indicates the relationships between many of the different spaces[1] of adapted càdlàg processes we have considered. From Section A.6.1, we write $\mathcal{S}_{\mathrm{Sp}}$ for the special semimartingales, and for simplicity, here we write $\tilde{\mathcal{M}}$ for the space of martingales (as opposed to $\mathcal{M}$, which is the space of uniformly integrable martingales), and $\tilde{\mathcal{M}}^d$ and $\tilde{\mathcal{M}}_b$ for purely discontinuous and bounded martingales respectively. We assume throughout that $p \in [1, \infty[$. A subscript 'preloc' refers to processes prelocally in the space.

An arrow indicates inclusion, that is, the arrow $\mathcal{M} \to \tilde{\mathcal{M}}$ indicates that every uniformly integrable martingale is a martingale, or $\mathcal{M} \subset \tilde{\mathcal{M}}$. In most cases, given an appropriate topology[2] on each of these spaces, the arrow also implies that one topology is stronger than the other, that is $\mathcal{H}^1 \to \mathcal{H}^1_{\mathcal{S}}$ indicates that the $\mathcal{H}^1$ topology is at least as strong as the $\mathcal{H}^1_{\mathcal{S}}$ topology (the latter being restricted to $\mathcal{H}^1$). These results are proven at various points in the text, or are immediate from the definition. In addition, apart from the clear exceptions when $p = 1$, each of these inclusions is strict in general, that is, $\mathcal{M} \to \tilde{\mathcal{M}}$ indicates $\mathcal{M} \neq \tilde{\mathcal{M}}$. Each of the nodes in this tree can be localized, and we have only shown those local classes which are of particular interest. We should also note that $\tilde{\mathcal{M}}_b$ is dense in $\mathcal{H}^1$.

[1] All the classes of processes in the diagram are real vector spaces, except $\mathcal{A}^+$, $\mathcal{A}^+_{\mathrm{loc}}$ and $\mathcal{V}^+$, which are not closed under multiplication by a negative number.

[2] This is only true 'in most cases' as we have not discussed appropriate topologies on some of these spaces, for example on $\mathcal{S}_{\mathrm{Sp}}$, or how a topology naturally weakens to allow for localization of a space.

S.N. Cohen, R.J. Elliott, *Stochastic Calculus and Applications*,
Probability and Its Applications, DOI 10.1007/978-1-4939-2867-5

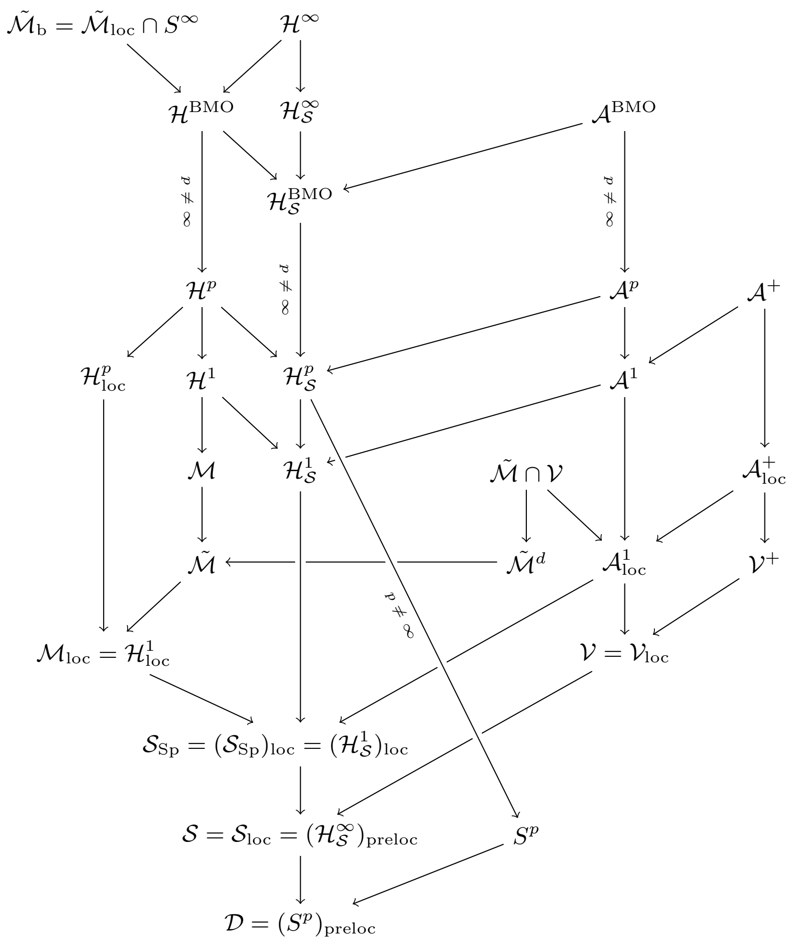

Fig. B.1. A taxonomic tree for adapted càdlàg processes.

When considering this diagram, the following identities are also informative (the first only holding for $p < \infty$)

$$\mathcal{H}^p = \mathcal{M}_{\mathrm{loc}} \cap S^p, \qquad \mathcal{S}_{\mathrm{Sp}} = \mathcal{M}_{\mathrm{loc}} \oplus \mathcal{A}^1_{\mathrm{loc}}, \qquad \mathcal{S} = \mathcal{M}_{\mathrm{loc}} \oplus \mathcal{V},$$

and for $p < p'$, $\mathcal{H}^p \subset \mathcal{H}^{p'}$. We also note that we paid particular attention to the sets of sub- and super-martingales, and to the space $\mathcal{H}^{2,d} = \mathcal{H}^2 \cap \tilde{\mathcal{M}}^d$, which do not appear in the diagram.

References

[1] D. Applebaum, *Lévy Processes and Stochastic Calculus*, 2nd edn (Cambridge University Press, Cambridge, 2009)

[2] A. Bain, D. Crisan, *Fundamentals of Stochastic Filtering* (Springer, New York, 2009)

[3] G. Barles, R. Buckdahn, E. Pardoux, Backward stochastic differential equations and integral-partial differential equations. Stochast. Stochast. Rep. **60**(1), 57–83 (1997)

[4] P. Barrieu, N. El Karoui, Monotone stability of quadratic semimartingales with applications to unbounded general quadratic BSDEs. Ann. Probab. **41**(3B), 1831–1863 (2013)

[5] P. Barrieu and N. El Karoui, Optimal derivatives design under dynamic risk measures, in Mathematics of Finance, Contemporary Mathematics (A.M.S. Proceedings), 13–26, (2004)

[6] M. Beiglböck, P. Siorpaes, Pathwise Versions of the Burkholder–Davis–Gundy Inequality, Bernoulli **21**(1), 360–373 (2015)

[7] M. Beiglböck, P. Siorpaes, Riemann-integration and a new proof of the Bichteler–Dellacherie theorem. Stoch. Process. Appl. **124**(3), 1226–1235 (2014)

[8] R. Bellman, The theory of dynamic programming. Bull. Am. Math. Soc. **60**(6), 503–515 (1954)

[9] J.J. Benedetto, *Real Variable and Integration* (Teubner, Stuttgart, 1976)

[10] V.E. Beneš, Existence of optimal strategies based on specified information, for a class of stochastic decision problems. SIAM J. Control **8**, 179–188 (1970)

[11] V.E. Beneš, Existence of optimal stochastic control laws. SIAM J. Control **9**, 446–475 (1971)

[12] V.E. Beneš, Full bang to reduce predicted miss is optimal. SIAM J. Control **15**, 52–83 (1976)

S.N. Cohen, R.J. Elliott, *Stochastic Calculus and Applications*,
Probability and Its Applications, DOI 10.1007/978-1-4939-2867-5

[13] V.E. Beneš, Exact finite dimensional filters for certain diffusions with nonlinear drift. Stochastics **5**, 65–92 (1981)

[14] K. Bichteler, *Stochastic Integration with Jumps* (Cambridge University Press, Cambridge, 2002)

[15] K. Bichteler, J.-B. Gravereaux, J. Jacod, *Malliavin Calculus for Processes with Jumps* (OPA (Amsterdam) B.V., Amsterdam, 1987)

[16] P. Billingsley, *Probability and Measure*, Anniversary edn (Wiley, New York, 2012)

[17] J.M. Bismut, Conjugate convex functions in optimal stochastic control. J. Math. Anal. App. **44**, 384–404 (1973)

[18] D. Blackwell, On a class of probability spaces, in *Proceedings Third Berkeley Symposium on Mathematical Statistics and Probability*, vol. 2 (University of California Press, Berkeley, 1956)

[19] D. Blackwell, L.E. Dubins, On existence and non-existence of proper, regular, conditional distributions. Ann. Probab. **3**(5), 741–752 (1975)

[20] R. Boel, P. Varaiya, Optimal control of jump processes. SIAM J. Control **15**, 92–119 (1977)

[21] V.I. Bogachev, *Measure Theory* (Springer, Berlin, 2007)

[22] B. Bouchard, N. Touzi, Discrete-time approximation and Monte-Carlo simulation of backward stochastic differential equations. Stochast. Process. Appl. **111**(2), 175–206 (2004)

[23] P. Briand, B. Delyon, Y. Hu, E. Pardoux, L. Stoica, l^p solutions of backward stochastic differential equations. Stoch **108**, 109–129 (2003)

[24] P. Briand, R. Elie, A simple constructive approach to quadratic BSDEs with or without delay. Stochast. Process. Appl. **123**, 2921–2939 (2013)

[25] P. Briand, Y. Hu, Quadratic BSDEs with convex generators and unbounded terminal conditions. Probab. Theory Relat. Fields **141**, 543–567 (2008)

[26] R.W. Brockett, Nonlinear systems and nonlinear estimation theory, in *Stochastic Systems*, ed. by M. Hazewinkel, J.C. Willems. NATO Advanced Study Institute, Les Arcs, France (Reidel, Dordrecht, 1980)

[27] B. Bru, M. Yor, Comments on the life and mathematical legacy of Wolfgang Doeblin. Financ. Stochast. **6**, 3–47 (2002)

[28] R.S. Bucy, Nonlinear filtering theory. IEEE Trans. Autom. Control **10**(2), 198 (1965)

[29] M. Capiński, P.E. Kopp, *Measure, Integral and Probability*, 2nd edn (Springer, New York, 2004)

[30] R. Carmona, M. Teranchi, *Interest Rate Models: An Infinite Dimensional Stochastic Analysis Perspective* (Springer, Berlin, 2006)

[31] J.-F. Chassagneux, A. Richou, Numerical simulation of quadratic BSDEs. to appear in Ann. Appl. Probab. (2015)

[32] A.S. Cherny, A.N Shiryaev, Vector stochastic integrals and the fundamental theorems of asset pricing. Proc. Steklov Math. Inst. **237**, 12–56 (2002)

[33] C.S. Chou, P.-A. Meyer, Sur la représentation des martingales comme intégrales stochastiques dans les processus ponctuels. In *Séminaire de Probabilités IX*. Lecture Notes in Mathematics, vol. 465 (Springer, New York, 1975)

[34] S.N. Cohen, R.J. Elliott, Existence, uniqueness and comparisons for BSDEs in general spaces. Ann. Probab. **40**(5), 2264–2297 (2012)

[35] R. Coifman, C. Fefferman, Weighted norm inequalities for maximal functions and singular integrals. Studia Math. **51**, 241–250 (1974)

[36] F. Confortola, Dissipative backward stochastic differential equations in infinite dimensions. Infinite Dimens. Anal. Quantum Probab Relat. Top. **9**, 155–168 (2006)

[37] M.G. Crandall, H. Ishii, P.-L. Lions, User's guide to viscosity solutions of second order partial differential equations. Bull. Am. Math. Soc. **27**, 1–67 (1992)

[38] S. Crépey, *Financial Modeling: A Backward Stochastic Differential Equations Perspective* (Springer, Berlin, 2013)

[39] G. Da Prato, J. Zabczyk, *Stochastic Equations in Infinite Dimensions* (Cambridge University Press, Cambridge, 1992)

[40] R.W.R. Darling, E. Pardoux, Backwards SDE with random terminal time and applications to semilinear elliptic PDE. Ann. Probab. **25**(3), 1135–1159 (1997)

[41] M.H.A. Davis, On a multiplicative functional transformation arising in nonlinear filtering theory. Z. Wahrsch. verw. Gebiete **54**, 125–139 (1980)

[42] M.H.A. Davis, J.M.C. Clark, On predicted miss stochastic control problems. Stochastics **2**, 197–209 (1979)

[43] M.H.A. Davis, Martingale representation and all that. in *Advances in Control, Communication Networks, and Transportation Systems: In Honor of Pravin Varaiya*, ed. by E.H. Abed. Systems and Control: Foundations and Applications (Birkhauser, Boston, 2005)

[44] M.H.A. Davis, On the existence of optimal policies in stochastic control. SIAM J. Control **11**, 507–594 (1973)

[45] M.H.A. Davis, Martingales of Wiener and Poisson processes. J. Lond. Math. Soc. **13**, 336–338 (1976)

[46] M.H.A. Davis, The representation of martingales of jump processes. SIAM J. Control **14**, 623–238 (1976)

[47] M.H.A. Davis, Martingale methods in stochastic control, in *Stochastic Control and Stochastic Differential Systems*. Lecture Notes in Control and Information Sciences, vol. 16 (Springer, Berlin, 1979)

[48] M.H.A. Davis, Pathwise nonlinear filtering, in *Stochastic Systems*, ed. by M. Hazewinkel, J.C. Willems. NATO Advanced Study Institute, Les Arcs, France (Reidel, Dordrecht, 1980)

[49] M.H.A. Davis, R.J. Elliott, Optimal control of a jump process. Z. Wahrsch. verw. Gebiete **40**, 183–202 (1977)

[50] M. De Donno, P. Guasoni, M. Pratelli, Super-replication and utility maximization in large financial markets. Stochast. Process. Appl. **115**(12), 2006–2022 (2005)
[51] F. Delbaen, W. Schachermayer, *The mathematics of arbitrage*, 2nd Printing (Springer, Berlin, 2008)
[52] F. Delbaen, Y. Hu, X. Bao, Backward SDEs with superquadratic growth. Probab. Theory Relat. Fields **150**, 145–192 (2011)
[53] C. Dellacherie, *Capacités et Processus Stochastiques* (Springer, Berlin, 1972)
[54] C. Dellacherie, P.-A. Meyer, *Probabilités et Potentiel*, 2ème edn, chaps. I–IV (Hermann, Paris, 1975)
[55] Ł. Delong, *Backward Stochastic Differential Equations with Jumps and Their Actuarial and Financial Applications* (Springer, London, 2013)
[56] A. Dembo, O. Zeitouni, Parameter estimation of partially observed continuous time stochastic processes via the EM algorithm. Stochast. Process. Appl. **23**, 91–113 (1986)
[57] A.P. Dempster, N.M Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. Ser. B **39**(1), 1–38 (1977)
[58] G. Di Nunno, Y.A. Rozanov, On measurable modification of stochastic functions. Theory Probab. Appl. **46**(1), 122–127 (2002)
[59] G. Di Nunno, B. Øksendal, F. Proske, *Malliavin Calculus for Lévy Processes with Applications to Finance* (Springer, Berlin, 2009)
[60] C. Doléans-Dade, P.-A. Meyer, Une charctérisation de BMO, in *Séminaire de Probabilités XI, Univ. de Strasbourg.* Lecture Notes in Mathematics, vol. 581 (Springer, Berlin, 1977), pp. 383–389
[61] C. Doléans-Dade, On the existence and unicity of solutions of stochastic integral equations. Z. Wahrsch. verw. Gebiete **36**, 93–101 (1976)
[62] J.L. Doob, *Stochastic Processes* (Wiley, New York, 1953)
[63] T.E. Duncan, P. Varaiya, On the solutions of a stochastic control system. SIAM J. Control **9**, 354–371 (1971)
[64] I. Ekren, C. Keller, N. Touzi, J. Zhang, On viscosity solutions of path dependent PDEs. Ann. Probab. **42**, 204–236 (2014)
[65] N. El Karoui, S. Peng, M.-C. Quenez, Backward stochastic differential equations in finance. Math. Financ. **7**(1), 1–71 (1997)
[66] R.J. Elliott, *Viscosity Solutions and Optimal Control.* Pitman Research Notes in Mathematics (Longman Scientific & Technical, London, 1987)
[67] R.J. Elliott, Stochastic integrals for martingales of a jump process with partially accessible jump times. Z. Wahrsch. verw. Gebiete **36**, 213–226 (1976)
[68] R.J. Elliott, Innovation projections of a jump process and local martingales. Proc. Camb. Philos. Soc. **81**, 77–90 (1977)
[69] R.J. Elliott, Lévy functionals and jump process martingales. J. Math. Anal. App. **57**, 638–652 (1977)

[70] R.J. Elliott, M. Kohlmann, The variation principle and stochastic optimal control. Stochastics **3**, 229–241 (1980)
[71] R.J. Elliott, L. Aggoun, J.B. Moore, *Hidden Markov Models: Estimation and Control* (Springer, Berlin, 1994)
[72] M. Émery, Une définition faible de BMO. Ann. Inst. Henri Poincaré Sect. B **21**(1), 59–71 (1985)
[73] M. Émery, Stabilité des solutions des équations diffeérentielles stochastiques application aux intégrales multiplicatives stochastiques. Z. Wahrsch. verw. Gebiete **41**, 241–262 (1978)
[74] M. Émery, Équations différentielles stochastiques Lipschitziennes: étude de la stabilité. *Seminaire de Probabilités XIII* (Springer, Berlin, 1979), pp. 281–293
[75] M. Émery, Une topologie sur l'espace des semimartingales, in *Seminaire des Probabilités (Université de Strasbourg)*, vol. 13 (Springer, Berlin, 1979), pp. 260–280
[76] M. Émery, Compensation de processus à variation finie non localement intégrables. in *Seminaire des Probabilités (Université de Strasbourg)*, vol. 14 (Springer, Berlin, 1980), pp. 152–160
[77] S.N. Ethier, T.G. Kurtz, *Markov Processes: Characterization and Convergence* (Wiley, Hoboken, 1986)
[78] A.F. Filippov, On certain questions in the theory of optimal control. Vestnik Moskov. Univ. Ser. Mat. Meh. Astronom. **2**, 25–42 (1959). English trans. J. Soc. Indust. Appl. Math. Ser. A Control **1**, 76–84 (1962)
[79] W.H. Fleming, H.M. Soner. *Controlled Markov Processes and Viscosity Solutions*, 2nd edn (Springer, New York, 2005)
[80] W.H. Fleming, Optimal continuous parameter stochastic control. SIAM Rev. **11**, 470–509 (1969)
[81] H. Föllmer, A. Schied, *Stochastic Finance: An Introduction in Discrete Time*. Studies in Mathematics, vol. 27 (de Gruyter, Berlin, 2002)
[82] D.H. Fremlin, *Measure Theory* (Torres Fremlin, Colchester, 2001)
[83] M. Fuhrman, Y. Hu, Backward stochastic differential equations in infinite dimensions with continuous driver and applications. Appl. Math. Optim. **56**(2), 265–302 (2007)
[84] M. Fujisaki, G. Kallianpur, H. Kunita, Stochastic differential equations for the nonlinear filtering problem. Osaka J. Math. **1**, 19–40 (1972)
[85] A.M. Garsia, *Martingale Inequalities Seminar Notes on Recent Progress* (W.A. Benjamin Inc, Reading, 1973)
[86] B.R. Gelbaum, J.M.H. Olmsted, *Counterexamples in Analysis* (Dover Publications Inc, New York, 2003). Reprint of original (1965)
[87] H.K. Getoor, M.J. Sharpe, Conformal martingales. Invent. Math. **16**, 271–308 (1972)
[88] I.I. Gihman, A.V. Skorohod, *Stochastic Differential Equations* (Springer, Berlin, 1972)

[89] I.V. Girsanov, On transforming a certain class of stochastic processes by absolutely continuous substitution of measures. Theory Probab. Appl. **5**, 285–301 (1960)

[90] P. Glasserman, *Monte Carlo Methods in Financial Engineering* (Springer, New York, 2003)

[91] M. Hairer, An introduction to stochastic PDEs. arXiv:0907.4178 (2009)

[92] B. Hajek, E. Wong, Multiple stochastic integrals: projection and iteration. Z. Wahrsch. verw. Gebiete **63**, 349–368 (1983)

[93] S. Hamadène, J.-P. Lepeltier, Backward equations, stochastic control and zero-sum stochastic differential games. Stochast. Stochast. Rep. **54**, 221–231 (1995)

[94] S.-W. He, J.-G. Wang, J.-A. Yan, *Semimartingale Theory and Stochastic Calculus* (Science Press and CRC, Beijing, 1992)

[95] J.G. Hocking, G.S. Young, *Topology* (Addison-Wesley, Reading, 1961)

[96] Y. Hu, P. Imkeller, M. Müller, Utility maximization in incomplete markets. Ann. Appl. Probab. **15**(3), 1691–1712 (2005)

[97] Y. Hu, S. Peng, On the comparison theorem for multidimensional BSDEs. C. R. Math. **343**(2), 135–140 (2006)

[98] N. Ikeda, S. Watanabe, *Stochastic Differential Equations and Diffusion Processes* (North-Holland, Amsterdam, 1981)

[99] P. Imkeller, G. dos Reis, Path regularity and explicit convergence rate for BSDE with truncated quadratic growth. Stochast. Process. Appl. **120**, 348–379 (2010)

[100] H. Ishii, P.-L. Lions, Viscosity solutions of fully nonlinear second-order elliptic partial differential equations. J. Differ. Equ. **83**(1), 26–78 (1990)

[101] K. Itô, Differential equations determining a Markoff process. J. Pan-Japan Math. Coll. **244**(1077), 1352–1400 (1942). Original Japanese: Zenkoku Sizyo Sugaku Danwakai-si. Reproduced in translation in Kiyosi Itô, Selected Papers, eds. Stroock and Varadhan, 1987 (Springer)

[102] K. Itô, Stochastic integrals. Proc. Imp. Acad. Tokyo **20**, 519–524 (1944)

[103] M. Izumisawa, N. Kazamaki, Weighted norm inequalities for martingales. Tôhoku Math. J. **29**, 115–124 (1977)

[104] M. Izumisawa, T. Sekiguchi, Y. Shiota, Remark on a characterization of BMO-martingales. Tôhoku Math. J. **31**, 281–284 (1979)

[105] J. Jacod, Multivariate point processes; predictable projection, Radon–Nikodym derivatives, representation of martingales. Z. Wahrsch. verw. Gebiete **31**, 235–253 (1975)

[106] J. Jacod, Un théorème de représentation pour les martingales discontinues. Z. Wahrsch. verw. Gebiete **34**, 225–244 (1976)

[107] J. Jacod, *Calcul stochastique et problèmes de martingales*. Lecture Notes in Mathematics, vol. 714 (Springer, Berlin, 1979)

[108] J. Jacod, Intégrales stochastiques par rapport à une semi-martingale vectorielle et changements de filtration, in *Seminaire de Probabilités XIV* (Springer, Berlin, 1980), pp. 161–172

[109] J. Jacod, M. Yor, Étude des solutions extrémales et représentation intégrale des solutions pour certains problèmes de martingales. Z. Wahrsch. verw. Gebiete **38**, 83–125 (1977)
[110] J. Jacod, A.N. Shiryaev, *Limit Theorems for Stochastic Processes*, 2nd edn (Springer, Berlin, 2002)
[111] G. Johnson, L.L. Helms, Class D supermartingales. Bull. Am. Math. Soc. **69**, 59–62 (1963)
[112] T. Kailath, An innovation approach to least square estimation. Part I: Linear filtering in additive white noise. IEEE Trans. Autom. Control **13**, 646–655 (1968)
[113] O. Kallenberg, *Foundations of Modern Probability* (Springer, New York, 2002)
[114] G. Kallianpur, *Stochastic Filtering Theory* (Springer, Berlin, 1980)
[115] R.E. Kalman, A new approach to linear filtering and prediction problems. J. Basic Eng. ASME **82**, 33–45 (1960)
[116] R.E. Kalman, R.S. Bucy, New results in linear filtering and prediction theory. J. Basic Eng. ASME **83**, 95–108 (1961)
[117] I. Karatzas, S.E. Shreve, *Brownian Motion and Stochastic Calculus*, 2nd edn (Springer, New York, 1991)
[118] P.E. Kloeden, E. Platen, *Numerical Solution of Stochastic Differential Equations*, 3rd printing edition (Springer, Berlin, 1999)
[119] M. Kobylanski, Backward stochastic differential equations and partial differential equations with quadratic growth. Ann. Probab. **2**, 558–602 (2000)
[120] N.V. Krylov, *Controlled Diffusion Processes* (Springer, New York, 1980)
[121] H. Kunita, Cauchy problems for stochastic partial differential equations arising in nonlinear filtering theory. Syst. Control Lett. **1**, 37–41 (1981)
[122] H. Kunita, S. Watanabe, On square integrable martingales. Nagoya Math. J. **30**, 209–245 (1967)
[123] J.-P. Lepeltier, B. Marchal, Problème des martingales et équations différentielles stochastiques associées à un opérateur intégro-différentiel. Ann. Inst. Henri Poincaré, Sect. B **12**(1), 43–103 (1976)
[124] D. Lépingle, J. Mémin, Integrabilité uniforme et dans l^r des martingales exponentielles. Sém. Probab. de Rennes (1978)
[125] D. Lépingle, J. Mémin, Sur l'intégrabilité uniforme des martingales exponentielles. Z. Wahrsch. verw. Gebiete **42**, 175–203 (1978)
[126] R.S. Lipster, A.N. Shiryayev, *Statistics of Random Processes*, vols I and II (Springer, Berlin, 1977)
[127] G. Lowther, *Almost Sure: A Random Mathematical Blog*. http://almostsure.wordpress.com/
[128] R. Mansuy, The origins of the word martingale. Electron. J. Hist. Probab. Stat. **5**(1), 1–10 (2009)
[129] S.I. Marcus, Modeling and approximation of stochastic differential equations driven by semimartingales. Stochastics **4**, 223–245 (1981)

[130] E.J. McShane, R.B. Warfield, Jr., On Filippov's implicit functions lemma. Proc. Am. Math. Soc. **18**, 41–47 (1967)
[131] J. Mémin, Conditions d'optimalité pour un problème de control portant sur une famille dominée de probabilités. Journée de control de Metz, Mai 1976 (1976)
[132] J. Mémin, Espaces de semi martingales et changement de probabilité. Z. Wahrsch. verw. Gebiete **52**(1), 9–39 (1980)
[133] P.-A. Meyer, *Un cours sur les intégrales stochastiques, Seminaire de Probabilités X*. Lecture Notes in Mathematics, vol. 511 (Springer, Berlin, 1976)
[134] P.-A. Meyer, Notes sur les integrales stochastiques. II, in *Seminaire de Probabilités XI* (Springer, Berlin, 1977), pp. 463–464
[135] P.-A. Meyer, Inégalités de normes pour les intégrales stochastiques, in *Seminaire de Probabilités XII* (Springer, Berlin, 1978), pp. 757–762
[136] P.-A. Meyer, A differential geometric formalism for the Ito calculus. in *Stochastic Integrals*, ed. by D. Williams. Lecture Notes in Mathematics, vol. 851 (Springer, Berlin, 1981)
[137] R. Mikulevicius, B.L. Rozovskii, Normalized stochastic integrals in topological vector spaces, in *Seminaire de Probabilités XXXII* (Springer, Berlin, 1998), pp. 137–165
[138] J. Neveu, *Martingales à temps discret* (Masson & Cie, Paris, 1972)
[139] J. Norris, *Markov Chains* (Cambridge University Press, Cambridge, 1997)
[140] D. Nualart, W. Schoutens, Chaotic and predictable representations for Lévy processes. Stochast. Process. Appl. **90**(1), 109–122 (2000)
[141] T. Okada, A criterion for uniform integrability of exponential martingales. Tohoku Math. J. **34**, 485–630 (1982)
[142] B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications*, 5th edn (Springer, New York, 2003)
[143] B. Øksendal, A. Sulem, *Applied Stochastic Control of Jump Diffusions* (Springer, Berlin, 2007)
[144] A. Osekowski, Sharp maximal inequalities for the martingale square bracket. Stochastics **28**, 589–605 (2010)
[145] E. Pardoux, Stochastic partial differential equations and filtering of diffusion processes. Stochastics **3**, 127–167 (1979)
[146] E. Pardoux, S. Peng, Backward stochastic differential equations and quasilinear parabolic partial differential equations, in *Stochastic Partial Differential Equations and Their Applications*. Lecture Notes in Control and Information Sciences, vol. 176 (Springer, Berlin, 1992), pp. 200–217
[147] E. Pardoux, S. Peng, Adapted solution of a backward stochastic differential equation. Syst. Control Lett. **14**, 55–61 (1990)
[148] K.R. Parthasarathy, *Probability Measures on Metric Spaces* (American Mathematical Society, Providence, 2005)
[149] H. Pham, *Continuous-Time Stochastic Control and Optimization with Financial Applications* (Springer, Berlin, 2009)

[150] S.R. Pliska, Controlled jump processes. Stochast. Process. Appl. **3**, 259–282 (1975)

[151] P. Protter, $\mathcal{H}^p$ stability of solutions of stochastic differential equations. Z. Wahrsch. verw. Gebiete **44**, 337–352 (1978)

[152] P. Protter, *Stochastic Integration and Differential Equations*, 2nd edn (Springer, New York, 2000)

[153] P. Protter, K. Shimbo, No arbitrage and general semimartingales, in *Markov Processes and related Topics: A Festschrift for Thomas G. Kurtz*. IMS Collections, vol. 4 (Institute of Mathematical Statistics, Beachwood, 2008), pp. 267–283

[154] M.-C. Quenez, Stochastic control and BSDEs, in *Backward Stochastic Differential Equations*, ed. by N. El Karoui, L. Mazliak (Addison Wesley Longman, Harlow, 1997)

[155] D. Revuz, M. Yor, *Continuous Martingales and Brownian Motion*, 3rd edn (Springer, Berlin, 1999)

[156] R. Rishel, *A Minimum Principle for Controlled Jump Processes*. Lecture Notes in Economics and Mathematics Systems, vol. 107 (Springer, Berlin, 1975)

[157] R.T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, 1996)

[158] L.C.G. Rogers, A guided tour through excursions. Bull. Lond. Math. Soc. **21**(4), 305–341 (1989)

[159] L.C.G. Rogers, D. Williams, *Diffusions, Markov Processes and Martingales*, vols. I and II, 2nd edn (Cambridge University Press, Cambridge, 2000)

[160] H.L. Royden, P.M. Fitzpatrick, *Real Analysis*, 4th edn (Prentice Hall, Upper Saddle River, 2010)

[161] M. Royer, BSDEs with a random terminal time driven by a monotone generator and their links with PDEs. Stochast. Stochast. Rep. **76**(4), 281–307 (2004)

[162] M. Royer, Backward stochastic differential equations with jumps and related non-linear expectations. Stochast. Process. Appl. **116**(10), 1358–1376 (2006)

[163] W. Rudin, *Principles of Mathematical Analysis*, 3rd edn (McGraw-Hill International, New York, 1976)

[164] J. Ruf, M. Larsson, *Convergence of Local Supermartingales and Novikov–Kazamaki Type Conditions for Processes with Jumps*. arXiv:1411.6229 (2014)

[165] K.-I. Sato, *Lévy Processes and Infinitely Divisible Distributions* (Cambridge University Press, Cambridge, 1999)

[166] A.N. Shiryaev, *Probability*, 2nd edn (Springer, New York, 2000)

[167] A.N. Shiryayev, Some new results in the theory of controlled stochastic processes, in *Trans. 4th Prague Conference on Information Theory* (Czech Acad. of Sciences, Prague, 1967)

[168] R. Situ, On solutions of backward stochastic differential equations with jumps and applications. Stochast. Process. Appl. **66**(2), 209–236 (1997)
[169] A. Sokol, Optimal Novikov-type criteria for local martingales with jumps. Electron. Commun. Probab. **18**(39), 1–6 (2013)
[170] R.L. Stratonovich, A new representation for stochastic integrals and equations. SIAM J. Control **4**, 362–371 (1966)
[171] R.L. Stratonovich, *Conditional Markov Processes and Their Application to the Theory of Optimal Control* (Elsevier, New York, 1968)
[172] D.W. Stroock, Applications of Fefferman–Stein type interpolation to probability theory and analysis. Commun. Pure Appl. Math. **26**, 477–495 (1973)
[173] D.W. Stroock, Diffusion processes associated with Lévy generators. Z. Wahrsch. verw. Gebiete **32**, 209–244 (1975)
[174] D.W. Stroock, S.R.S. Varadhan, *Multidimensional Diffusion Processes* (Springer, Berlin, 1979)
[175] S. Tang, X. Li, Necessary conditions for optimal control of stochastic systems with random jumps. SIAM J. Control Optim. **32**, 1447–1475 (1994)
[176] R. Tevzadze, Solvability of backward stochastic differential equations with quadratic growth. Stochast. Process. Appl. **118**(3), 503–515 (2008)
[177] N. Touzi, Optimal stochastic control, stochastic target problems, and backward SDE, in *Fields Institute Monographs*, vol. 29 (Springer, New York, 2013)
[178] B.S. Tsirelson, An example of a stochastic differential equation having no strong solution. Theory Probab. Appl. **20**(2), 416–418 (1975)
[179] J. Ville, *Étude critique de la notion de collectif*. PhD Thesis, Faculté des sciences de Paris (1939)
[180] D.H. Wagner, Survey of measurable selection theorems. SIAM J. Control Optim. **15**(5), 859–903 (1977)
[181] D.H. Wagner, Survey of measurable selection theorems: an update, in *Measure Theory Oberwolfach 1979*. Lecture Notes in Mathematics, vol. 794 (Springer, Berlin, 1980), pp. 176–219
[182] J.B. Walsh, An introduction to stochastic partial differential equations, in *École d'Été de Probabilités de Saint Flour XIV - 1984*. Lecture Notes in Mathematics, vol. 1180 (1986), pp. 265–439
[183] D. Williams, *Probability with Martingales* (Cambridge University Press, Cambridge, 1991)
[184] E. Wong, Recent progress in stochastic processes—a survey. IEEE Trans. Info. Theory **19**(3), 262–275 (1973)
[185] W.N. Wonham, Some applications of stochastic differential equations to optimal nonlinear filtering. SIAM J. Control **2**, 347–369 (1965)
[186] J.-A. Yan, A propos de l'intégrabilité uniforme des martingales exponentielles, in *Seminaire de Probabilités XVI* (Springer, Berlin, 1982), pp. 338–347

[187] C. Yoeurp, Decompositions des martingales locales et formules exponentielles, in *Séminaire de Probabilités*. Lecture Notes in Mathematics, vol. 511 (Springer, Berlin, 1976)

[188] J. Yong, Completeness of security markets and backward stochastic differential equations with unbounded coefficients. Nonlinear Anal. **63**, e2079–e2089 (2005)

[189] K. Yosida, *Functional Analysis* (Springer, Berlin, 1980)

[190] M. Zakai, On the optimal filtering of diffusion processes. Z. Wahrsch. verw. Gebiete **11**, 230–243 (1969)

[191] O. Zeitouni, A. Dembo, Exact filters for the estimation of the number of transitions of finite-state continuous-time Markov processes. IEEE Trans. Info. Theory **34**, 890–893 (1988)

[192] J. Zhang, A numerical scheme for BSDEs. Ann. Appl. Probab. **14**(1), 459–488 (2004)

Notation and Abbreviations

2^S	Power set of S	Page 4
$\int_A$	Integral over a set A	Page 15
$a \vee b$	Maximum of a and b	Page 15
$a \wedge b$	Minimum of a and b	Page 15
$\bigvee_n \mathcal{F}_n$	σ-algebra generated by $\bigcup_n \mathcal{F}_n$	Page 5
f^+, f^-	Positive and negative parts of f	Page 15
$\mathcal{A}$, $\mathcal{A}^+$	Spaces of processes of integrable variation	Page 179
$\mathcal{A}^p$	Space of processes of p-integrable variation	Page 191
$\mathcal{A}^p_{\text{pred}}$	Predictable processes in $\mathcal{A}^p$	Page 406
a.e., a.a.	Almost everywhere, almost all	Page 12
a.s.	Almost surely	Page 51
$\tilde{\mathcal{A}}$	The optional and integrable random measures	Page 307
$\tilde{\mathcal{A}}_\sigma$	The optional and (predictably) σ-integrable random measures	Page 307
$\tilde{\mathcal{A}}^1$, $\tilde{\mathcal{A}}^1_\sigma$	Integer valued random measures in $\tilde{\mathcal{A}}$, $\tilde{\mathcal{A}}_\sigma$	Page 313
$\mathcal{B}(S), \mathcal{B}(\mathbb{R}^n)$	Borel σ-algebras	Page 5
$\overline{\mathcal{B}}(\mathbb{R})$	Lebesgue σ-algebra	Page 10
BMO	Bounded mean oscillation spaces	Page 603
BSDE	Backward stochastic differential equation	Page 468
B(X,Y)	Bounded linear operators $X \to Y$	Page 30
C^2_b	Bounded functions with bounded first and second derivatives	Page 437
C^2_ν	C^2 functions with a global ν-integrability property	Page 442
C_0	continuous functions vanishing at $\pm\infty$	Page 437
C_b	Bounded continuous functions	Page 437
$\mathcal{C}_{\text{loc}}$	Processes locally in a class $\mathcal{C}$	Page 84
càdlàg	Right-continuous with left-limits	Page 78

S.N. Cohen, R.J. Elliott, *Stochastic Calculus and Applications*,
Probability and Its Applications, DOI 10.1007/978-1-4939-2867-5

ℓ^2, ℓ^p	Normed spaces of real sequences	Page 32
$L^2(M)$	Predictable H with $E\big[(H_0M_0)^2 + \int_{[0,\infty]} H_s^2 d\langle M\rangle_s\big] < \infty$	Page 261
$L^2(\nu)$	Functions $g : \mathcal{Z} \to \mathbb{R}^m$ with $\int_{\mathcal{Z}} \|g(\zeta)\|^2 \nu(d\zeta) < \infty$	Page 428
$L^2(\langle W\rangle)$, $L^2(\langle\tilde{\mu}\rangle)$	L^2 spaces associated with $\mathcal{H}^2$ stochastic integrals	Page 470
$L^p(M)$	Predictable H with $E\big[\big(\int_{[0,\infty]} H_s^2 d[M]_s\big)^{p/2}\big] < \infty$	Page 269
$L^p(S, \Sigma, \mu)$	Space of p-integrable functions $S \to \mathbb{R}$	Page 35
$L(X)$	Space of predictable integrands with $H \bullet X$ well defined	Page 272
$\|\mu\|$, $\|d\mu\|$	Total variation measure	Page 41
μ^+, μ^-	Positive and negative parts of a measure	Page 40
μ_p	The compensator of the random measure μ	Page 311
$\mu\|_\Sigma$	Restriction of a measure μ to Σ	Page 9
$\mathcal{M}$	The space of uniformly integrable martingales	Page 132
$\mathcal{M}_{0,\mathrm{loc}}$	The space of local martingales with $M_0 = 0$	Page 132
$\mathcal{M}_{\mathrm{loc}}$	The space of local martingales	Page 132
M_μ	The Doléans measure of μ	Page 308
$\langle M\rangle$	Predictable quadratic variation of M	Page 233
$\langle M, N\rangle$	Predictable quadratic covariation of M and N	Page 236
$[M]$	Optional quadratic variation of X	Page 234
$\mathbb{N}$	Natural numbers $\{1, 2, \dots\}$	Page 3
$\overline{\mathbb{N}}$	Extended natural numbers $\{1, 2, \dots, \infty\}$	Page 3
$\tilde{N}$	The compensated Poisson process	Page 128
$(\Omega, \mathcal{F}, P)$	A general probability space	Page 50
$\tilde{\Omega}$	The product $\Omega \times [0,\infty] \times \mathcal{Z}$	Page 304
$\\| \cdot \\|_p$	The L^p norm	Page 33
$P(\cdot\|A)$	The probability conditional on an event	Page 53
Π_x^*	The dual optional or predictable projection (for $x = o, p$)	Page 188
$\pi_t(\phi)$	The $\{\mathcal{Y}_t\}_{t\in[0,T]}$-optional projection of ϕ	Page 544
Π_x	The optional or predictable projection ($x = o, p$)	Page 167
$\overline{\mathbb{R}}$	Extended real numbers $[-\infty, \infty]$	Page 3
$\sigma(f)$	σ-algebra generated by a function f	Page 13
$\sigma(\mathcal{G})$	σ-algebra generated by a family of sets $\mathcal{G}$	Page 4
Σ^0	Uncompleted σ-algebra	Page 10
$\Sigma_1 \otimes \Sigma_2$	The product σ-algebra	Page 25
$\tilde{\Sigma}_x$	The product $\Sigma_x \otimes \mathcal{Z}$	Page 304

Index

S.N. Cohen, R.J. Elliott, *Stochastic Calculus and Applications*,
Probability and Its Applications, DOI 10.1007/978-1-4939-2867-5