

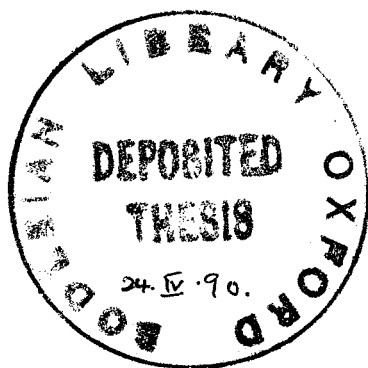
THE ITÔ CALCULUS: A VECTOR-INTEGRAL APPROACH

A thesis submitted for the degree of Doctor of
Philosophy at Oxford University,

BY

PAUL DAVID LING,

St. Catherine's College, Oxford,
July 1989.



THE ITÔ CALCULUS: A VECTOR-INTEGRAL APPROACH

PAUL DAVID LING,
ST. CATHERINE'S COLLEGE, OXFORD.

Thesis submitted for the degree D. Phil.,
Trinity Term 1989.

ABSTRACT

The Itô calculus is the theory of stochastic integrals $\int_0^t X_u dS_u$, where S is a semimartingale, and X is a suitable previsible process. The approach most commonly given in the literature is the 'Strasbourg approach', in which S is taken first to be an L^2 -martingale, and an L^2 -isometry is used to define the integral. Then localization and pathwise Stieltjes integration are used to extend to the case when S is a semimartingale. In this thesis a different, and more direct approach is given. The space L^2 is replaced by L^p for an arbitrary $p \in [0, \infty)$, with a particular interest shown in the case when $p = 0$. The integral $\int X dS$ is defined as an L^p -valued Daniell integral (the 'Daniell' analogue of L^p -valued measures). The necessary theory of such Daniell integrals is given in Chapters 2 and 3. In Chapter 4 the stochastic integral $\int X dS$ is defined and its properties given. It is necessary to place a dominated convergence condition on S . Processes satisfying this condition are called L^p -integrators. A characterization of L^p -integrators in terms of uniform convergence is given at the end of Chapter 4. A proof that all integrators have pathwise càdlàg versions is given in Chapter 5. Localization and processes of finite-variation are also considered in Chapter 5. Quadratic variation, $[S]$, and mutual variation, $[R, S]$, are defined in Chapter 6 by a martingale-free method, and its properties proven. The Itô formula is then stated and proved. Martingales are for the first time introduced in Chapters 7 and 8. Some results from the discrete-time theory (which are presented in Chapter 7) are used in Chapter 8 to prove that all martingales are integrators. This Daniell method of stochastic integration is seen to be consistent with and at least as powerful as the Strasbourg method. Finally in Chapter 9 a proof is presented of the result of Bichteler, Dellacherie, Mokobodzki and Meyer, that the class of all (L^0) -integrators is precisely the class of all semimartingales. As prerequisites for this proof, Doob-Meyer decompositions for L^1 -integrators and local L^1 -integrators are obtained, and the class of all semimartingales is proven to be invariant under change of the probability measure (provided the two probabilities are equivalent).

ACKNOWLEDGEMENTS

I would like to thank Dr. D. A. Edwards, my supervisor, for his help and suggestions. I am grateful for the funding of the Science and Engineering Research Council.

CONTENTS

Abstract	p.2
Acknowledgements	p.3
Contents	pp.4—5
Chapter 1: Introduction	pp.6—19
§1.1 Some Notation and Terminology	
§1.2 L^p Spaces for $0 \leq p \leq \infty$	
Chapter 2: Upper Gauges	pp.20—31
§2.1 Definitions	
§2.2 Φ -null Sets	
§2.3 Φ -finite Functions	
§2.4 Integrability and the Convergence Theorems	
§2.5 Φ -measurability	
§2.6 σ -finite Upper Gauge Spaces	
Chapter 3: Vector-Valued Integrals	pp.32—49
§3.1 Definitions	
§3.2 The Upper Gauge Corresponding to J	
§3.3 The Integral and the Convergence Theorems	
§3.4 Independence of the Choice of Metric on $\mathcal{X}$	
§3.5 Examples and Alternative Characterizations of $\mathcal{X}$ -valued Integral Spaces	
§3.6 The Necessity of the C-space Condition	
§3.7 Examples of C-spaces: L^p -spaces	
Chapter 4: The Stochastic Integral	pp.50—71
§4.1 The Setup and Definitions	
§4.2 The Stochastic Integral	
§4.3 The Convergence Theorems and Other Properties	
§4.4 Some Results in the Theory of L^0 -Stochastic Integration	
§4.5 An Alternative Characterization of L^p -integrators	
Chapter 5: Integrators as Pathwise Càdlàg Processes	pp.72—99
§5.1 An Introduction to Stopping Times	
§5.2 Pathwise Càdlàg Versions of Integrators	
§5.3 Localization	
§5.4 Continuous Integrators and a Pathwise Convergence Result	
§5.5 Increasing and Finite-Variation Processes	
§5.6 Brownian Motion as an Integrator	

Chapter 6: Quadratic Variation and the Itô Formula	pp.100—124
§6.1 Motivation	
§6.2 Quadratic Variation and Mutual Variation	
§6.3 The Mutual Variation of Two Stochastic Integrals	
§6.4 The Itô Formula	
§6.5 The Differential Notation	
Chapter 7: Discrete-Time Martingales	pp.125—138
§7.1 Conditional Expectation	
§7.2 Discrete-Time Processes	
§7.3 The Martingale Transform and Mutual Variation	
§7.4 Some General Results	
§7.5 Some Estimates For Martingale Transforms	
Chapter 8: Martingales and the Stochastic Integral	pp.139—164
§8.1 Uniform Integrability	
§8.2 Uniform Integrability and Conditional Expectation	
§8.3 Continuous-Time Martingales	
§8.4 Some General Results	
§8.5 Local Martingales	
§8.6 Mutual Variation of Martingales and Local Martingales	
§8.7 L^2 -Martingales and the Strasbourg Approach	
Chapter 9: The Characterization of Integrators	
as Semimartingales	pp. 165—179
§9.1 Definitions	
§9.2 Some Characterizations of Local L^1 -integrators and the	
Doob-Meyer Decomposition	
§9.3 Change of Probability	
§9.4 The Proof of the Characterization Theorem	
References	p.180

CHAPTER 1: INTRODUCTION

The purpose of this thesis is to develop the theory of the stochastic integral $\int_0^t X dS$, where X and S are certain random processes of continuous time, and the integral is to be thought of as a Stieltjes integral. (A random process X is to be thought of as a random function $t \mapsto X_t$ from $[0, \infty)$ to $\mathbb{R}$.) We shall define the integral in a way that is different from that commonly given in the literature (the ‘Strasbourg approach’ see below). We shall use a method that gives the integral in one step (unlike the Strasbourg approach).

These integrals arise in the theory of stochastic processes from, for example, the theory of stochastic differential equations. Consider for example an ordinary differential equation

$$(1) \quad \frac{dx_t}{dt} = f(t, x_t),$$

for some suitable function $f: [0, \infty) \times \mathbb{R} \rightarrow \mathbb{R}$. Now it may be desirable to add some ‘noise’ to this equation and obtain the stochastic differential equation

$$(2) \quad dX_t = f(t, X_t) dt + \theta(t, X_t) dB_t,$$

where B_t denotes Brownian motion and the solution X is a random process. But what does (2) mean? The problem being that we can not ‘divide’ by dt , since dB_t/dt fails to exist with probability one. The next attempt is to place an integral sign $\int_0^t$ in front of equation (2) to obtain

$$(3) \quad X_t = X_0 + \int_0^t f(s, X_s) ds + \int_0^t \theta(s, X_s) dB_s.$$

There is no problem interpreting the first integral of (3): just consider it pathwise. But if we try to interpret the second integral as a pathwise Stieltjes integral we run into difficulty because B does not have finite variation. The Itô calculus allows us to define such integrals in another way. We must in some way use the fact that we have an underlying probability space. This restricts the class of integrands that we may use: they must not depend on knowledge of the future. Stochastic differential equations have many uses. They for example allow one to construct the diffusion process associated with an elliptic operator, and in particular Brownian motion on a Riemannian manifold.

We return to considering the integral $\int_0^t X dS$. The first thing to do is to introduce a filtration $(\mathcal{F}_t)_{t \in \mathbb{R}^+}$ on the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$. This is an increasing family of sub σ -fields of $\mathcal{F}$. We shall require that S is adapted. That is, S_t is $\mathcal{F}_t$ -measurable for all t . (In probability theory σ -fields represent knowledge. Thus a filtration corresponds to the increasing of knowledge with time.) Next a class $\mathcal{E}$ of ‘elementary’ processes is defined, and the integral $\int X dS$ is defined for $X \in \mathcal{E}$ (see Definitions (4.5) and (4.6)). The next step is to extend $\int \dots dS$ by taking limits.

The Strasbourg Approach. The approach to this limiting process most commonly given in the literature is one that was worked out for the Séminaire de Probabilités at Strasbourg in the year 1974—75 (see Meyer [Me]). We shall refer to this approach throughout this thesis as the ‘Strasbourg approach’ or the ‘Strasbourg method’. As I have indicated, we shall be using a method different from this, which I believe is much cleaner. We now shall give an outline of the Strasbourg method as background information. If M is an L^2 -martingale (of which Brownian motion is an example), there is an increasing L^1 -process $[M]$ (for a Brownian motion B , $[B]_t = t$) such that

$$(4) \quad \mathbb{E} \left(\int X dM \int Y dM \right) = \mathbb{E} \left(\int_0^\infty XY d[M] \right) \quad \text{for all } X, Y \in \mathcal{E}.$$

The left hand side of (4) is just the inner product of $\int X dM$ and $\int Y dM$ in $L^2(\mathbb{P})$. The right hand side is the inner product of X and Y in $L^2(\mu_M)$, where μ_M is a measure defined by $\mu_M E = \mathbb{E}(\int_0^\infty \mathbf{1}_E d[M])$. So $\int \dots dM$ is an isometry, and thus we can extend its domain to the closure of $\mathcal{E}$ in $L^2(\mu_M)$. This closure is $L^2((0, \infty) \times \Omega, \mathcal{P}, \mu_M)$, where $\mathcal{P}$ is the previsible σ -field which will be defined in Chapter 4 (Definition(4.3)(b)).

Now having defined $\int X dM$, we define $(X \cdot M)_t = \int_0^t X dM$ to be $\int \mathbf{1}_{(0,t] \times \Omega} X dM$ when this exists. This defines an L^2 -martingale $X \cdot M$. This however does not give the most general integral possible. To generalise the integral further it is necessary, in the Strasbourg approach, to use the method of localization. This increases the class of integrands to the set of previsible processes X such that there exists a sequence of stopping times $(\sigma_n)_n$ such that firstly $\sigma_n \uparrow \infty$, and secondly $\mathbf{1}_{(0,\sigma_n]} X$ is integrable as above with respect to M on $(0, t]$ for all t . Then $X \cdot M$ is defined by the formula $(X \cdot M)_{\sigma_n \wedge t} = ((\mathbf{1}_{(0,\sigma_n]} X) \cdot M)_t$. Another similar localization generalizes the class of integrands to the class of local L^2 -martingales. Finally we can generalize still further by considering processes S that are the sum of a local L^2 -martingale and a process with paths of finite variation A . Define $X \cdot S = X \cdot M + X \cdot A$, where the last term is defined as a pathwise Stieltjes integral. A full development of the Strasbourg method can be found for example in Rogers & Williams [RW].

The problem with the Strasbourg approach is that using L^2 places restrictions on the processes in the Ω direction of the product $[0, \infty) \times \Omega$. One way we might hope to get around this problem is to replace L^2 by something larger such as L^0 (the space of all random variables with the topology of convergence in probability). It is useful however to use L^p for arbitrary $p \in [0, \infty)$ (we exclude L^∞ for technical reasons). The case when $p = 2$ ties in nicely with the Strasbourg approach. The case when $p = 1$ is useful when considering martingales and local martingales.

We shall start therefore with an adapted vector-valued function $S: \mathbb{R}^+ \rightarrow L^p$ and

define $\int X dS$ for $X \in \mathcal{E}$ as above. We could think of the function S as defining an L^p -valued Stieltjes measure on $(0, \infty) \times \Omega$; but since L^p (when $p < 1$) is not a Banach space this does not seem pleasant. The way we shall take is based on that given in Bichteler [Bi2]. We define something that is like an L^1 -norm on $(0, \infty) \times \Omega$. That something is an upper gauge.

In Chapter 2 we shall define what is meant by an upper gauge on an arbitrary set, and prove analogues for upper gauges of the main theorems of integration theory. This concept is based on something of the same name given in Bichteler [Bi1]. The definition that we shall give of measurability is different from that of Bichteler; but it is sufficient for our purposes.

In Chapter 3 we shall consider linear maps $J: \mathcal{E} \rightarrow \mathcal{X}$, where $\mathcal{E}$ is a certain type of subspace of real-valued functions on a set Γ and $\mathcal{X}$ is an F-space. Then we shall use the theory of upper gauges to extend J to an ' L^1 -closure' of $\mathcal{E}$. But in order to define the upper gauge it is necessary firstly that $\mathcal{X}$ be a C-space (defined in Chapter 3), and secondly J satisfies certain properties. This extension is precisely what is needed to define the stochastic integral. In Chapter 3 it is shown that L^p is a C-space for $0 \leq p < \infty$; but L^∞ in general is not. It is this that prevents us using L^∞ in the definition of the stochastic integral.

Then in Chapter 4 we apply this theory to the problem of extending $\int \dots dS$ from $\mathcal{E}$ to something useful. This places restrictions on S , and the concept of an integrator results (in fact we have the term ' L^p -integrator' for each $p \in [0, \infty)$). Various properties of the integral are then proven; for example $X \bullet (H \bullet S) = (XH) \bullet S$. In §4.4 the power of using L^0 begins to be apparent. For instance we see that for each càd càg process X (see (1.4) below), $X \bullet S$ is defined for all integrators S . Also $\int_0^t X dS$ is expressed as a limit in probability of Riemann sums. Then a characterization of integrators, given in Bichteler [Bi2] (Theorem 2.6), in terms of uniform convergence of integrands is proved (thereby filling an apparent gap in Bichteler's account). This

is more useful functional analytically, because uniform convergence is easier to cope with than dominated convergence (as is used in the original definition).

Up to this point integrators have been considered as functions from $[0, \infty)$ to L^p . But in Chapter 5 we shall introduce stopping times and prove the result, given in Bichteler [Bi2], that an integrator has a pathwise càdlàg version. This is done using the method of upcrossings. The theory of localization is also developed in this chapter, and is shown not to lead to any generalization of the stochastic integral (unlike the Strasbourg approach). We also show that finite-variation processes are integrators and also that Brownian motion is an integrator.

Next we come to a fundamental result, the Itô formula. This is proved in Chapter 6. The one dimensional version of this result states that

$$(5) \quad f(S_t) - f(S_0) = \int_0^t f'(S) dS + \frac{1}{2} \int_0^t f''(S) d[S],$$

when S is a continuous integrator, $f: \mathbb{R} \rightarrow \mathbb{R}$ is a C^2 function, and $[S]$ denotes the quadratic variation process of S . The definition of this process and that of the mutual variation process $[R, S]$ of two integrators R and S is also given in this chapter. The important fact that $[X \cdot R, Y \cdot S] = (XY) \cdot [R, S]$, is given, and proved using upper-gauge methods. The chapter finishes with a formalization of the differential notation such as used in equation (2) above. All this is done without ever using the concept of a martingale!

However martingales are important in the Itô calculus. The rest of this thesis therefore considers the connection between martingales and the Itô calculus. We show that all local martingales, local submartingales, local supermartingales and local quasimartingales are integrators. It is also proved that $H \cdot M$ is a local martingale when M is a local martingale and H is a locally bounded previsible process. By proving that a martingale is an integrator we can use a result of Chapter 5 to see that all martingales have pathwise càdlàg versions; thus avoiding the need for another upcrossing argument.

This is the work of Chapter 8. In order to carry this out we shall need some results from the discrete-time case; particularly some inequalities concerning the martingale transform (the discrete-time analogue of $H \cdot X$). The discrete-time theory needed is given in Chapter 7. In that chapter discrete-time analogues of $H \cdot X$ and $[X, Y]$ are defined. This gives a foretaste of what is to follow in the continuous-time case. This use of discrete-time theory is an extension of some work done by D. A. Edwards, see [Ed]. It allows us to avoid the use at this stage of the continuous-time Doob-Meyer decomposition. We give at the end of Chapter 8 results that tie up the method of this thesis with the Strasbourg approach. Finally we present a proof in Chapter 9 of the result due to Bichteler, Dellacherie, Mokobodzki and Meyer that the class of integrators is precisely the class of semimartingales. Also given in that chapter are Doob-Meyer decomposition results for L^1 -integrators, local L^1 -integrators and submartingales. In order to prove the characterization theorem, we give a proof that the class of all semimartingales is invariant under change of probability measure, providing the two probabilities are equivalent. This is made easy by the work of previous chapters.

There have been books written on the subject of the Itô calculus using vector-valued measures instead of the Strasbourg approach; see for example Métivier & Pellaumail [MP]. The method presented in this thesis is completely different to that given in these books (and I believe simpler).

Note. I have tried to make this thesis as self-contained as possible, even if this has sometimes entailed giving more or less routine proofs of standard results. I have done this partly because I hope the work will later be useful as a text for beginners in stochastic analysis. But I have also sought to make it a manifesto for this particular way of developing the theory; and I believe that in the end the provision of all essential proofs may be the quickest method of convincing the sceptical reader that there is no sleight of hand in my attempt to persuade him or her that this is indeed a direct,

clean, and attractive way of establishing the theory.

§1.1 SOME NOTATION AND TERMINOLOGY

We include in this section some remarks about some terminology and notation that we shall be using throughout this thesis.

(1.1)Number Sets. We shall assume throughout the following conventions for number sets.

$$\begin{aligned}\mathbb{N} &= \{0, 1, 2, \dots\}, & \mathbb{R}^+ &= [0, \infty), & \overline{\mathbb{R}} &= [-\infty, \infty], \\ \mathbb{Q}^+ &= \mathbb{Q} \cap [0, \infty), & \mathbb{R}^{++} &= (0, \infty), & \overline{\mathbb{R}}^+ &= [0, \infty].\end{aligned}$$

In general a superscript ‘+’ on a set X such as a set of numbers, or a function space, will denote the set of non-negative elements of that set. We shall also use conventions such as $\mathbb{N}_{>0}$ for the set $\{1, 2, 3, \dots\}$ (i.e. the elements of $\mathbb{N}$ that are greater than zero).

(1.2)Set Theory. When A and B are sets, $A \triangle B$ will denote the symmetric difference, $(A \setminus B) \cup (B \setminus A)$, of A and B . B^A will denote the set of functions from A to B .

When $P(s_1, \dots, s_n)$ is an n -ary predicate on a set Γ , and $f_1, \dots, f_n: \Omega \rightarrow \Gamma$ for some set Ω , we will use $[P(f_1, \dots, f_n)]$ to denote $\{\omega \in \Omega \mid P(f_1(\omega), \dots, f_n(\omega))\}$. So for example if $f, g: \Omega \rightarrow \mathbb{R}$, then $[f < g]$ denotes the set of points $\omega \in \Omega$ such that $f(\omega) < g(\omega)$.

When A is a subset of some set Ω , we shall use $\mathbf{1}_A$ to denote the real-valued function on Ω that takes the value 1 on A and 0 on $\Omega \setminus A$.

(1.3)Maxima, Minima and Vector Sublattices. If $x, y \in \overline{\mathbb{R}}$ then we shall use $x \vee y$ and $x \wedge y$ to denote the maximum and the minimum respectively of x and y .

We also shall use this notation pointwise when applied to $\overline{\mathbb{R}}$ -valued functions. For each $x \in \overline{\mathbb{R}}$, we shall use x^+ and x^- to denote $x \vee 0$ and $(-x) \vee 0$ respectively. So $x = x^+ - x^-$ and $|x| = x^+ + x^-$. Again we shall use this notation pointwise when it is applied to $\overline{\mathbb{R}}$ -valued functions. If Γ is a set then by a *vector sublattice* of $\mathbb{R}^\Gamma$ we shall mean a real subspace of $\mathbb{R}^\Gamma$ that is closed under the pointwise operations $\vee$ and $\wedge$.

(1.4) Càdlàg and Làdcàg Functions. Suppose that Γ is a regular Hausdorff topological space. A function $f: \mathbb{R}^+ \rightarrow \Gamma$ is said to be *càdlàg* if and only if f is right continuous and the limit $f(t-) = \lim_{s \uparrow t} f(s)$ exists in Γ for all $t \in \mathbb{R}^{++}$. A function $f: \mathbb{R}^{++} \rightarrow \Gamma$ is said to be *làdcàg* if and only if it is left continuous and the limit $f(t+) = \lim_{s \downarrow t} f(s)$ exists in Γ for all $t \in \mathbb{R}^+$. Note here it is assumed that $f(0+)$ exists even though $f(0)$ is not defined. This is to ensure that certain boundedness properties hold for f .

The terms ‘càdlàg’ and ‘làdcàg’ originate from the French expressions ‘continu à droite et limité à gauche’ (continuous on the right and limits on the left) and ‘limité à droite et continu à gauche’ (limits on the right and continuous on the left) respectively. The first is in common usage, but the second is my own invention. I thought of using ‘càglàd’ for làdcàg but it sounds and looks too much like ‘càdlàg’.

Now it is true that real-valued càdlàg functions are bounded on each $[0, b]$. To prove such results it is convenient to consider càdlàg functions as continuous functions on some topological space. Let R as a set be the disjoint union of $\mathbb{R}^+$ and a copy of $\mathbb{R}^{++}$. When $t > 0$ let t_- denote the corresponding element of the copy of $\mathbb{R}^{++}$. Now totally order R so that $0 < s_- < s < t_- < t$ whenever $0 < s < t$ and give R the order topology. Now if $f_0: R \rightarrow \Gamma$ is continuous, then $f = f_0|_{\mathbb{R}^+}$ is a càdlàg function. Also $f(t-) = f_0(t_-)$ for all $t > 0$. Now if conversely $f: \mathbb{R}^+ \rightarrow \Gamma$ is càdlàg, then by defining $f_0(t) = f(t)$ and $f_0(t_-) = f(t-)$, we obtain a continuous function $f_0: R \rightarrow \Gamma$ such that $f = f_0|_{\mathbb{R}^+}$ (it is here that the regularity of Γ is used). Thus we see that continuous

functions on R are in one to one correspondence with càdlàg functions on $\mathbb{R}^+$. Now it is relatively easy to prove that the interval $[0, b]_R$ of R is a compact subset of R for each $b \in \mathbb{R}^+$. So a continuous image of it is also compact. Thus for each càdlàg function $f: \mathbb{R}^+ \rightarrow \Gamma$ and $b \in \mathbb{R}^+$, the set $\{f(t) | 0 \leq t \leq b\} \cup \{f(t-) | 0 < t \leq b\}$ is compact.

We can do a similar thing with làdcàg functions. I.e. take L to be $\mathbb{R}^{++}$ disjoint union with a copy of $\mathbb{R}^+$; writing elements of the copy of $\mathbb{R}^+$ as t_+ etc. Now R and L are order isomorphic via the map from R to L that sends t to t_+ and t_- to t . Thus càdlàg functions and làdcàg functions are in one to one correspondence. The correspondence is as follows: the càdlàg function f corresponds to the làdcàg function f_- defined by $f_-(t) = f(t-)$. Conversely the làdcàg function g corresponds to the càdlàg function g_+ defined by $g_+(t) = g(t+)$.

(1.5) Simple Functions. Suppose that X and Y are sets. Then a function $f: X \rightarrow Y$ is said to be *simple* if and only if its range $f(X)$ is a finite subset of Y . This terminology is most useful when X is a measure space and Y is a set of numbers such as $\mathbb{R}$.

(1.6) Partial Derivatives. Suppose that $f: U \rightarrow \mathbb{C}$, where U is an open subset of $\mathbb{R}^n$. We shall use $\partial_k f$ to denote the partial derivative of f with respect to the k^{th} variable when it exists. When $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n$ we shall let ∂^α denote the operator $\partial_1^{\alpha_1} \dots \partial_n^{\alpha_n}$. We shall let $|\alpha| = \alpha_1 + \dots + \alpha_n$ so that $|\alpha|$ is the order of ∂^α . It is also convenient to use x^α to denote the monomial $x_1^{\alpha_1} \dots x_n^{\alpha_n}$ when $x = (x_1, \dots, x_n) \in \mathbb{R}^n$.

(1.7) F-Spaces. An *F-space* is a topological vector space such that its topology can be given by a complete translation-invariant metric d . (A metric d is said to be translation invariant if and only if $d(x+z, y+z) = d(x, y)$ for all x, y and z .) Now for such a translation-invariant d , we can write $d(x, y) = \rho(x - y)$, where $\rho(x) = d(x, 0)$. We shall also refer to ρ as the metric. A subset E of a topological vector space X is said

to be *bounded* if and only if for each neighbourhood V of zero in X , there is a $t_0 \in \mathbb{R}^+$ such that $E \subseteq tV$ for all $t \geq t_0$. When X is an F-space with (translation-invariant) metric ρ , this condition is equivalent to $\sup_{x \in E} \rho(\lambda x) \rightarrow 0$ as $\lambda \downarrow 0$.

§1.2 L^p SPACES FOR $0 \leq p \leq \infty$

Suppose that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space. We shall be using the spaces $L^p(\Omega, \mathcal{F}, \mathbb{P})$ for $0 \leq p \leq \infty$. The spaces $L^p(\Omega, \mathcal{F}, \mathbb{P})$ for $1 \leq p \leq \infty$ are familiar, but for $0 \leq p < 1$ the spaces $L^p(\Omega, \mathcal{F}, \mathbb{P})$ can also be defined. These are F-spaces but not in general Banach spaces (they are not in general even locally convex: see Rudin [Ru] (1.47) for the case $0 < p < 1$). We are particularly interested in the space L^0 , since as we shall see it contains all the random variables, has the topology of convergence in probability, and gives the most general class of integrators for stochastic integration.

(1.8) Definition. Suppose that $(\Omega, \mathcal{F}, \mu)$ is a measure space. When $0 < p < \infty$, $\mathcal{L}^p$ denotes the set of all measurable $f: \Omega \rightarrow \mathbb{R}$ such that $\int_{\Omega} |f|^p d\mu < \infty$. $\mathcal{L}^0$ denotes the set of all measurable functions $f: \Omega \rightarrow \mathbb{R}$. $\mathcal{L}^{\infty}$ denotes the set of all measurable functions $f: \Omega \rightarrow \mathbb{R}$ such that $|f| \leq c$ a.e. for some $c \in \mathbb{R}^+$. We let L^p ($0 \leq p \leq \infty$) denote the set of all equivalence classes in $\mathcal{L}^p$ of the equivalence relation of equality almost everywhere. For $0 < p < \infty$ we define

$$\rho_p(f) = \left(\int_{\Omega} |f|^p d\mu \right)^{1/(1 \vee p)} \quad \text{and} \quad \|f\|_p = \left(\int_{\Omega} |f|^p d\mu \right)^{1/p} \quad \text{for all } f \in \mathcal{L}^p.$$

Then ρ_p is a metric on L^p that makes it into an F-space. For $f \in \mathcal{L}^{\infty}$, we let $\rho_{\infty}(f) = \|f\|_{\infty}$ denote the least $c \in \mathbb{R}^+$ such that $|f| \leq c$ a.e. This is a norm on L^{∞} making it into a Banach space. When μ is a probability we shall make L^0 into an F-space by defining the metric ρ_0 on L^0 by

$$\rho_0(f) = \int_{\Omega} (1 \wedge |f|) d\mu \quad \text{for all } f \in \mathcal{L}^0.$$

When necessary we shall denote the dependence of L^p on $(\Omega, \mathcal{F}, \mu)$ by $L^p(\Omega, \mathcal{F}, \mu)$, $L^p(\Omega)$, $L^p(\mathcal{F})$, $L^p(\mu)$ etc., and similarly for $\mathcal{L}^p$. For $1 \leq p \leq \infty$, $\rho_p = \|\cdot\|_p$ is the usual L^p -norm, so L^p is a Banach space in this case. $\square$

We should here give the proof that ρ_p for $0 < p < 1$ is a metric. If $s, t \geq 0$ and $s + t = 1$ then $s^p \geq s$ and $t^p \geq t$. Therefore $1 = s + t \leq s^p + t^p$. For any $x, y \geq 0$ put $s = x/(x + y)$ and $t = y/(x + y)$ (unless $x = y = 0$) to give that $(x + y)^p \leq x^p + y^p$. It is now easy to see that $\rho_p(f) = \int_{\Omega} |f|^p d\mathbb{P}$ defines a metric. The completeness of L^p for $0 \leq p < 1$ is proved in just the same way as for $1 \leq p < \infty$ by showing first that $\sum_{n=1}^{\infty} \rho_p(f_n) < \infty$ implies that $\sum_{n=1}^{\infty} f_n$ converges a.e. and in L^p . We also see that if $f_n \rightarrow f$ in L^p then a subsequence $(f_{n_k})_k$ converges a.e. to f .

(1.9) Proposition. *Suppose that $(\Omega, \mathcal{F}, \mu)$ is a measure space. Then the following hold.*

- (a) *If $0 < p, q, r \leq \infty$ satisfy $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$, then $\|fg\|_r \leq \|f\|_p \|g\|_q$ for all $f \in \mathcal{L}^p(\mu)$ and $g \in \mathcal{L}^q(\mu)$.*
- (b) *If μ is a probability and $0 \leq q \leq p \leq \infty$, then $\mathcal{L}^p(\mu) \subseteq \mathcal{L}^q(\mu)$ and the inclusion $L^p(\mu) \hookrightarrow L^q(\mu)$ is continuous. In fact the following are true.*
 - (i) *If $q > 0$ then $\|f\|_q \leq \|f\|_p$ for all $f \in \mathcal{L}^p(\mu)$.*
 - (ii) *If $q = 0$ then $\rho_0(f) \leq \rho_p(f)$ for all $f \in \mathcal{L}^p(\mu)$.*

Proof:

(a) The cases where $p = \infty$ or $q = \infty$ are easy; so we shall suppose $p, q, r < \infty$. Since $\frac{1}{p/r} + \frac{1}{q/r} = 1$, we have by Hölder's inequality that

$$\int_{\Omega} |uv| d\mathbb{P} \leq \left(\int_{\Omega} |u|^{p/r} d\mathbb{P} \right)^{r/p} \left(\int_{\Omega} |v|^{q/r} d\mathbb{P} \right)^{r/q} \quad \text{for } u \in \mathcal{L}^{p/r} \text{ and } v \in \mathcal{L}^{q/r}.$$

Now put $u = |f|^r$ and $v = |g|^r$ and raise both sides to the power $1/r$ to give the result.

(b) This will clearly follow from (i) and (ii) (for the cases $q > 0$ and $q = 0$ respectively).

(i) Clearly we may suppose that $q < p$. Let $0 < r < \infty$ such that $\frac{1}{q} = \frac{1}{p} + \frac{1}{r}$ then by (a) $\|uv\|_q \leq \|u\|_p \|v\|_r$ for $u \in \mathcal{L}^p$ and $v \in \mathcal{L}^r$. Let $u = f$ and $v = 1$ to give the result ($\|1\|_r = 1$).

(ii) We can suppose that $p > 0$. There are two cases to consider: $0 < p \leq 1$ and $1 < p \leq \infty$. Suppose first that $0 < p \leq 1$. Then $1 \wedge |f| \leq (1 \wedge |f|)^p \leq |f|^p$; so integrating over Ω gives the result. Now suppose that $1 < p \leq \infty$. Then $\rho_0(f) \leq \rho_1(f) \leq \rho_p(f)$, where the first inequality follows from the case when $0 < p \leq 1$ and the second is from (i). $\square$

Now as we remarked, one of the useful facts about L^0 is that it has the topology of convergence in probability. Here is a proof of that fact.

(1.10)Proposition. *Suppose that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space. Then the following hold.*

(i) *The topology of $L^0(\mathbb{P})$ is that of ‘convergence in probability’ i.e. if $(X_\alpha)_\alpha$ is a net in $L^0(\mathbb{P})$ and $X \in L^0(\mathbb{P})$ then*

$$X_\alpha \rightarrow X \text{ in } L^0(\mathbb{P}) \iff \mathbb{P}[|X_\alpha - X| > \lambda] \rightarrow 0 \text{ for all } \lambda > 0.$$

(ii) *$L^0(\mathbb{P})$ is a topological algebra under the obvious pointwise definition of multiplication.*

Proof:

(i) Clearly without loss of generality we may assume that $X = 0$. Suppose then that $X_\alpha \rightarrow 0$ in L^0 . We have for $\lambda > 0$ that $\mathbf{1}_{[|X_\alpha| > \lambda]} \leq \frac{1}{1 \wedge \lambda} (1 \wedge |X_\alpha|)$. Thus taking the expectation of both sides gives that $\mathbb{P}[|X_\alpha| > \lambda] \leq \frac{1}{1 \wedge \lambda} \rho_0(X_\alpha)$. But $\rho_0(X_\alpha) \rightarrow 0$, so $\mathbb{P}[|X_\alpha| > \lambda] \rightarrow 0$ as required.

Now suppose conversely that $\mathbb{P}[|X_\alpha| > \lambda] \rightarrow 0$ for all $\lambda > 0$. Then given $\varepsilon > 0$, there exists an α_0 such that $\mathbb{P}[|X_\alpha| > \frac{\varepsilon}{2}] < \frac{\varepsilon}{2}$ for all $\alpha \geq \alpha_0$. So for $\alpha \geq \alpha_0$ we have that

$$\begin{aligned}\rho_0(X_\alpha) &= \int_{[|X_\alpha| \leq \varepsilon/2]} (1 \wedge |X_\alpha|) d\mathbb{P} + \int_{[|X_\alpha| > \varepsilon/2]} (1 \wedge |X_\alpha|) d\mathbb{P} \\ &\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.\end{aligned}$$

So $\rho_0(X_\alpha) \rightarrow 0$ as required.

(ii) We have to show that multiplication is continuous. Suppose that $(X_\alpha)_\alpha$ and $(Y_\alpha)_\alpha$ are nets in L^0 over the same directed set and $X, Y \in L^0$ are such that $X_\alpha \rightarrow X$ and $Y_\alpha \rightarrow Y$ in L^0 . We have to show that $X_\alpha Y_\alpha \rightarrow XY$ in L^0 , i.e. $\mathbb{P}[|X_\alpha Y_\alpha - XY| > \lambda] \rightarrow 0$ for each $\lambda > 0$.

So suppose that $\lambda > 0$ and that $\varepsilon > 0$. Take $M > 0$ such that $\mathbb{P}[|X| > M] < \frac{\varepsilon}{4}$ and $\mathbb{P}[|Y| > M] < \frac{\varepsilon}{4}$. Let $\eta > 0$ be such that $2M\eta + \eta^2 \leq \lambda$, and let α_0 be such that $\mathbb{P}[|X_\alpha - X| > \eta] < \frac{\varepsilon}{4}$ and $\mathbb{P}[|Y_\alpha - Y| > \eta] < \frac{\varepsilon}{4}$ for all $\alpha \geq \alpha_0$. Then on the set $\Omega \setminus ([|X| > M] \cup [|Y| > M] \cup [|X_\alpha - X| > \eta] \cup [|Y_\alpha - Y| > \eta])$ we have that

$$\begin{aligned}|X_\alpha Y_\alpha - XY| &\leq |X_\alpha - X||Y| + |X||Y_\alpha - Y| + |X_\alpha - X||Y_\alpha - Y| \\ &\leq 2M\eta + \eta^2 \leq \lambda.\end{aligned}$$

So we have that

$$[|X_\alpha Y_\alpha - XY| > \lambda] \subseteq [|X| > M] \cup [|Y| > M] \cup [|X_\alpha - X| > \eta] \cup [|Y_\alpha - Y| > \eta],$$

and therefore taking probability, it follows that $\mathbb{P}[|X_\alpha Y_\alpha - XY| > \lambda] < \frac{\varepsilon}{4} + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} = \varepsilon$ for all $\alpha \geq \alpha_0$. So $\mathbb{P}[|X_\alpha Y_\alpha - XY| > \lambda] \rightarrow 0$ as required. $\square$

(1.11) Proposition. Suppose that $(\Omega, \mathcal{F}, \mu)$ is a measure space, and that $0 < p < \infty$. Then the following are true.

- (i) For each $f \in \mathcal{L}^p(\mu)$ and $t > 0$, $t^p \mu[|f| > t] \leq \|f\|_p^p$.
- (ii) For each measurable $f: \Omega \rightarrow \mathbb{R}$,

$$\|f\|_p^p = p \int_0^\infty t^{p-1} \mu[|f| > t] dt,$$

(where $\|f\|_p = \infty$ if $f \notin \mathcal{L}^p$).

Proof:

- (i) We have that $t^p \mathbf{1}_{[|f| > t]} \leq |f|^p$. So integrate to give the result.
- (ii) Define $f: \mathbb{R}^+ \times \Omega \rightarrow \mathbb{R}$ by $f(t, \omega) = p t^{p-1} \mathbf{1}_{[|f| > t]}(\omega)$. Now the result follows from the fact that $\int_\Omega \int_0^\infty f(t, \omega) dt \mu(d\omega) = \int_0^\infty \int_\Omega f(t, \omega) \mu(d\omega) dt$. □

CHAPTER 2: UPPER GAUGES

§2.1 DEFINITIONS

As we have mentioned in the introduction, we shall require an abstraction of the L^1 -norm, called an upper gauge. In this chapter we shall define this concept and develop the theory as far as is needed for later chapters. This is based on some of the contents of Bichteler [Bi1]. We shall drop his requirement that $\Phi(\lambda f) = |\lambda|\Phi(f)$, because this will not be true in our application.

(2.1)Definition. Suppose that Γ is a set.

- (a) A *Stone lattice* on Γ is a vector sublattice (see (1.3)) $\mathcal{E}$ of $\mathbb{R}^\Gamma$ such that $s \in \mathcal{E} \implies s \wedge 1 \in \mathcal{E}$.
- (b) Suppose that $\mathcal{E}$ is a Stone lattice on Γ . Then an *upper gauge* on the pair $(\Gamma, \mathcal{E})$ is a function $\Phi: (\overline{\mathbb{R}}^+)^{\Gamma} \rightarrow \overline{\mathbb{R}}^+$ such that the following hold.
 - (1) For $s \in \mathcal{E}^+$ (where $\mathcal{E}^+ = \{s \in \mathcal{E} \mid s \geq 0\}$), $\Phi(\lambda s) \rightarrow 0$ as $\lambda \downarrow 0$.
 - (2) $\Phi(f) \leq \Phi(g)$ whenever $f, g: \Gamma \rightarrow \overline{\mathbb{R}}^+$ such that $f \leq g$.
 - (3) If $f_1, f_2, \dots: \Gamma \rightarrow \overline{\mathbb{R}}^+$, then $\Phi(\sum_{n=1}^{\infty} f_n) \leq \sum_{n=1}^{\infty} \Phi(f_n)$.
 - (4) If $s_1, s_2, \dots \in \mathcal{E}^+$ are such that $s_1 \leq s_2 \leq \dots$ and $\sup_n \Phi(\lambda s_n) \rightarrow 0$ as $\lambda \downarrow 0$, then $\Phi(|s_n - s_m|) \rightarrow 0$ as $n, m \rightarrow \infty$.
- (c) An *upper gauge space* is a triple $(\Gamma, \mathcal{E}, \Phi)$, where Γ is a set, $\mathcal{E}$ is a Stone lattice on Γ and Φ is an upper gauge on $(\Gamma, \mathcal{E})$. □

When $(\Gamma, \mathcal{E}, \Phi)$ is an upper gauge space and $f: \Gamma \rightarrow \overline{\mathbb{R}}$, we shall take $\Phi(f)$ to mean $\Phi(|f|)$. Now we will suppose until further notice that $(\Gamma, \mathcal{E}, \Phi)$ is an upper gauge space.

Note.

1. (1) of Definition(2.1)(b) gives that $\Phi(0) = 0$. Therefore (3) of the same definition gives that $\Phi(f_1 + \dots + f_n) \leq \Phi(f_1) + \dots + \Phi(f_n)$. Putting $f_1 = \dots = f_n = f$ gives that $\Phi(nf) \leq n\Phi(f)$.

2. The sum of two upper gauges is clearly also an upper gauge.

§2.2 Φ -NULL SETS

(2.2)Definition. A set $E \subseteq \Gamma$ is said to be Φ -null if and only if $\Phi(\mathbf{1}_E) = 0$. □

The expressions ' Φ -a.e.' (' Φ -almost everywhere') and 'for Φ -a.a. x ' ('for Φ -almost all x ') are defined in terms of this definition in the usual way. From (2) and (3) of Definition(2.1)(b) we obtain the following proposition.

(2.3)Proposition.

(i) If $E \subseteq F$ and F is Φ -null, then E is also Φ -null.

(ii) If $E_1, E_2, \dots$ are Φ -null, then $\bigcup_{n=1}^{\infty} E_n$ is also Φ -null. □

Now concerning equality Φ -a.e. we have the following.

(2.4)Proposition.

(i) Suppose that $f, g: \Gamma \rightarrow \overline{\mathbb{R}}$ and $f = g$ Φ -a.e. Then $\Phi(f) = \Phi(g)$.

(ii) Suppose that $f: \Gamma \rightarrow \overline{\mathbb{R}}$. Then $f = 0$ Φ -a.e. if and only if $\Phi(f) = 0$.

Proof:

(i) Let $E = [f \neq g]$. So we have $|f| \leq |g| + \sum_{n=1}^{\infty} \mathbf{1}_E$. Now E is Φ -null, so $\Phi(\mathbf{1}_E) = 0$ and thus we have $\Phi(|f|) \leq \Phi(|g|) + \sum_{n=1}^{\infty} \Phi(\mathbf{1}_E) = \Phi(|g|)$. I.e. $\Phi(f) \leq \Phi(g)$. The reverse inequality is similar.

(ii) Suppose that $\Phi(f) = 0$, and let $E = [f \neq 0]$. We have that $\mathbf{1}_E \leq \sum_{n=1}^{\infty} |f|$. Applying Φ gives that $\Phi(\mathbf{1}_E) \leq \sum_{n=1}^{\infty} \Phi(|f|) = 0$. Therefore E is Φ -null as required. For the converse apply (i) with $g = 0$. □

§2.3 Φ -FINITE FUNCTIONS

(2.5)Definition. A function $f: \Gamma \rightarrow \overline{\mathbb{R}}$ is said to be Φ -finite if and only if $\Phi(\lambda f) \rightarrow 0$ as $\lambda \rightarrow 0$ (or equivalently $\Phi(\lambda|f|) \rightarrow 0$ as $\lambda \downarrow 0$). $\square$

We see that (1) of Definition(2.1)(b) says that all elements of $\mathcal{E}$ are Φ -finite.

(2.6)Proposition. Suppose that $f: \Gamma \rightarrow \overline{\mathbb{R}}$ is Φ -finite, then $\Phi(f) < \infty$ and $|f| < \infty$ Φ -a.e.

Note. $\Phi(f) < \infty$ is not necessarily sufficient for f to be Φ -finite (cf. L^0 -integrators later). However it is sufficient in the case when $\Phi(\lambda g) = |\lambda|\Phi(g)$ for all g and λ .

Proof: By considering $|f|$ we may assume without loss of generality that $f \geq 0$. Now $\Phi(\frac{1}{n}f) \rightarrow 0$ as $n \rightarrow \infty$. So for some n , $\Phi(\frac{1}{n}f) < \infty$. Thus $\Phi(f) \leq n\Phi(\frac{1}{n}f) < \infty$. I.e. $\Phi(f) < \infty$. Now let $E = [f = \infty]$. Then $\mathbf{1}_E \leq \lambda f$ for all $\lambda > 0$. Thus applying Φ gives that $\Phi(\mathbf{1}_E) \leq \Phi(\lambda f) \rightarrow 0$ as $\lambda \downarrow 0$. So $\Phi(\mathbf{1}_E) = 0$ and E is Φ -null as required. $\square$

(2.7)Definition. $\mathcal{F}(\Phi)$ will denote the set of all Φ -finite $f: \Gamma \rightarrow \mathbb{R}$ (note real valued: this is no real loss of generality by Proposition(2.6)). We have that $\mathcal{F}(\Phi)$ is a Stone lattice on Γ . $F(\Phi)$ will denote the set of equivalence classes in $\mathcal{F}(\Phi)$ of the equivalence relation of equality Φ -a.e. $\square$

We shall now proceed to show that $F(\Phi)$ is an F-space with Φ as the metric.

(2.8)Lemma.

- (i) If $f_1, f_2, \dots \in \mathcal{F}(\Phi)$, $f: \Gamma \rightarrow \mathbb{R}$ and $\Phi(f_n - f) \rightarrow 0$ as $n \rightarrow \infty$, then $f \in \mathcal{F}(\Phi)$.
- (ii) If $f_1, f_2, \dots \in \mathcal{F}(\Phi)$ and $\sum_{n=1}^{\infty} \Phi(f_n) < \infty$, then $\sum_{n=1}^{\infty} |f_n| < \infty$ Φ -a.e. If further we let $f: \Gamma \rightarrow \mathbb{R}$ be such that $f = \sum_{n=1}^{\infty} f_n$ Φ -a.e., then $f \in \mathcal{F}(\Phi)$ and $\Phi(f - \sum_{n=1}^N f_n) \rightarrow 0$ as $N \rightarrow \infty$.

Proof:

(i) Given $\varepsilon > 0$, let N be such that $\Phi(f_N - f) < \frac{\varepsilon}{2}$. Then if $0 < \lambda \leq 1$, we have that $\Phi(\lambda f) \leq \Phi(\lambda f_N) + \Phi(f - f_N) < \Phi(\lambda f_N) + \frac{\varepsilon}{2}$. But $\Phi(\lambda f_N) < \frac{\varepsilon}{2}$ for λ sufficiently small ($f_N \in \mathcal{F}(\Phi)$), and so it follows that $\Phi(\lambda f) < \varepsilon$ for λ sufficiently small. Thus $\Phi(\lambda f) \rightarrow 0$ as $\lambda \downarrow 0$ as required.

(ii) Let $E = [\sum_{n=1}^{\infty} |f_n| = \infty]$. Then $\mathbf{1}_E \leq \sum_{n=N}^{\infty} |f_n|$. Apply Φ to give that $\Phi(\mathbf{1}_E) \leq \sum_{n=N}^{\infty} \Phi(f_n) \rightarrow 0$ as $N \rightarrow \infty$. Thus $\Phi(\mathbf{1}_E) = 0$, and so E is Φ -null as required.

Now let f be as given in the statement. Then $|f - \sum_{n=1}^N f_n| \leq \sum_{n=N+1}^{\infty} |f_n|$ Φ -a.e. So applying Φ gives that $\Phi(f - \sum_{n=1}^N f_n) \leq \sum_{n=N+1}^{\infty} \Phi(f_n) \rightarrow 0$ as $N \rightarrow \infty$ as required. That $f \in \mathcal{F}(\Phi)$ follows now from (i). $\square$

(2.9) Theorem. $F(\Phi)$ is an F -space with metric Φ (This is finite by Proposition(2.6)).

Proof: The only non-trivial things to check are the following properties.

(i) Scalar multiplication is continuous.

(ii) The metric is complete.

Let us now prove these.

(i) Suppose that $\alpha, \alpha_0 \in \mathbb{R}$ and $f, f_0 \in \mathcal{F}(\Phi)$. Then

$$\Phi(\alpha f - \alpha_0 f_0) \leq \Phi((\alpha - \alpha_0)f_0) + \Phi(\alpha(f - f_0)).$$

Now since f_0 is Φ -finite it follows that $\Phi((\alpha - \alpha_0)f_0) \rightarrow 0$ as $\alpha \rightarrow \alpha_0$. Take $N \in \mathbb{N}$ such that $|\alpha_0| < N$. Then for $|\alpha| < N$, $\Phi(\alpha(f - f_0)) \leq \Phi(N(f - f_0)) \leq N\Phi(f - f_0)$. So $\Phi(\alpha(f - f_0)) \rightarrow 0$ as $\alpha \rightarrow \alpha_0$ and $\Phi(f - f_0) \rightarrow 0$. Therefore $\Phi(\alpha f - \alpha_0 f_0) \rightarrow 0$ as $\alpha \rightarrow \alpha_0$ and $\Phi(f - f_0) \rightarrow 0$. I.e. scalar multiplication is continuous.

(ii) Suppose that $f_1, f_2, \dots \in \mathcal{F}(\Phi)$ are such that $\Phi(f_n - f_m) \rightarrow 0$ as $n, m \rightarrow \infty$. Then pick $n_1 < n_2 < \dots$ such that

$$(\dagger) \quad \Phi(f_n - f_m) < 2^{-k} \quad \text{for } n, m \geq n_k.$$

Then applying Lemma(2.8)(ii) to the sequence $f_{n_1}, f_{n_2} - f_{n_1}, f_{n_3} - f_{n_2}, \dots$ we see that there is an $f \in \mathcal{F}(\Phi)$ such that $\Phi(f_{n_k} - f) \rightarrow 0$ as $k \rightarrow \infty$. Now suppose that $\varepsilon > 0$ is given, and take k such that $\Phi(f_{n_k} - f) < \frac{\varepsilon}{2}$, and $2^{-k} < \frac{\varepsilon}{2}$. We have, by ($\dagger$), that for each $n \geq n_k$, $\Phi(f_n - f) \leq \Phi(f_n - f_{n_k}) + \Phi(f_{n_k} - f) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$. So $\Phi(f_n - f) \rightarrow 0$ as $n \rightarrow \infty$. $\square$

(2.10)Proposition. Suppose that $f_1, f_2, \dots, f \in \mathcal{F}(\Phi)$ and $\Phi(f_n - f) \rightarrow 0$ as $n \rightarrow \infty$. Then there is a subsequence $(f_{n_k})_k$ of $(f_n)_n$ such that $f_{n_k} \rightarrow f$ Φ -a.e.

Proof: Pick $n_1 < n_2 < \dots$ such that $\sum_{k=1}^{\infty} \Phi(f_{n_k} - f) < \infty$. Then by Lemma(2.8)(ii) $\sum_{k=1}^{\infty} |f_{n_k} - f| < \infty$ Φ -a.e. So $f_{n_k} \rightarrow f$ Φ -a.e. $\square$

§2.4 INTEGRABILITY AND THE CONVERGENCE THEOREMS

(2.11)Definition. We say that a function $f: \Gamma \rightarrow \mathbb{R}$ is Φ -integrable if and only if for each $\varepsilon > 0$ there is an $s \in \mathcal{E}$ such that $\Phi(f - s) < \varepsilon$ (or equivalently there is a sequence $(s_n)_n$ in $\mathcal{E}$ such that $\Phi(f - s_n) \rightarrow 0$). $\mathcal{L}^1(\Phi)$ will denote the set of all Φ -integrable functions, so that $\mathcal{L}^1(\Phi)$ is the closure of $\mathcal{E}$ in $\mathcal{F}(\Phi)$ ($\mathcal{L}^1(\Phi) \subseteq \mathcal{F}(\Phi)$ by Lemma(2.8)(i)). We let $L^1(\Phi)$ denote the set of equivalence classes in $\mathcal{L}^1(\Phi)$ of the equivalence relation of equality Φ -a.e. $\square$

We have that $L^1(\Phi)$ is a closed vector subspace of $F(\Phi)$ (and so is complete), and that $\mathcal{L}^1(\Phi)$ is a Stone lattice on Γ . These are easily proved: for example if $f, g \in \mathcal{L}^1(\Phi)$. Let $(s_n)_n$ and $(t_n)_n$ be ‘approximating’ sequences in $\mathcal{E}$ for f and g respectively. Then $(s_n \vee t_n)_n$ is an ‘approximating’ sequence for $f \vee g$ (see Lemma(2.12) below), and so $f \vee g \in \mathcal{L}^1(\Phi)$. That $\mathcal{L}^1(\Phi)$ is closed under addition, scalar multiplication, taking minima etc. is similar.

We shall proceed to give statements and proofs of the promised convergence theorems. First we state and prove a lemma.

(2.12) Lemma. Suppose that $x_1, \dots, x_n, y_1, \dots, y_n \in \mathbb{R}$. Then

$$|(x_1 \vee \dots \vee x_n) - (y_1 \vee \dots \vee y_n)| \leq |x_1 - y_1| \vee \dots \vee |x_n - y_n|,$$

and $|(x_1 \wedge \dots \wedge x_n) - (y_1 \wedge \dots \wedge y_n)| \leq |x_1 - y_1| \vee \dots \vee |x_n - y_n|.$

Proof: Let η denote the right hand side. Then $x_k - \eta \leq y_k \leq x_k + \eta$ for $k = 1, \dots, n$. Now take the maximum as k runs from 1 to n to give that $(x_1 \vee \dots \vee x_n) - \eta \leq (y_1 \vee \dots \vee y_n) \leq (x_1 \vee \dots \vee x_n) + \eta$. So $|(x_1 \vee \dots \vee x_n) - (y_1 \vee \dots \vee y_n)| \leq \eta$ as required. The other formula is similar. $\square$

(2.13) The Monotone Convergence Theorem.

Suppose that $f_1, f_2, \dots \in \mathcal{L}^1(\Phi)$ are such that $0 \leq f_1 \leq f_2 \leq \dots$ and $\sup_n \Phi(\lambda f_n) \rightarrow 0$ as $\lambda \downarrow 0$. Then there exists an $f \in \mathcal{L}^1(\Phi)$ such that $f = \sup_n f_n$ Φ -a.e. and $\Phi(f - f_n) \rightarrow 0$ as $n \rightarrow \infty$. $\square$

Proof: Given $\varepsilon > 0$ take $\tilde{s}_n \in \mathcal{E}$ such that $\Phi(f_n - \tilde{s}_n) < \varepsilon 2^{-n}$. By replacing $\tilde{s}_n$ by $\tilde{s}_n \vee 0$ we may assume without loss of generality that $\tilde{s}_n \geq 0$. Let $s_n = \tilde{s}_1 \vee \dots \vee \tilde{s}_n \in \mathcal{E}^+$, so $s_1 \leq s_2 \leq \dots$. Now $f_n = f_1 \vee \dots \vee f_n$, and so by Lemma(2.12), we see that

$$(\dagger) \quad |f_n - s_n| \leq |f_1 - \tilde{s}_1| \vee \dots \vee |f_n - \tilde{s}_n| \leq \sum_{k=1}^n |f_k - \tilde{s}_k|.$$

Thus applying Φ gives that

$$(\ddagger) \quad \Phi(f_n - s_n) \leq \sum_{k=1}^n \Phi(f_k - \tilde{s}_k) < \varepsilon.$$

For $0 < \lambda \leq 1$ we have by $(\ddagger)$ that

$$\sup_n \Phi(\lambda |f_n - s_n|) \leq \sum_{k=1}^{\infty} \Phi(\lambda |f_k - \tilde{s}_k|) \leq \sum_{k=1}^N \Phi(\lambda |f_k - \tilde{s}_k|) + \sum_{k=N+1}^{\infty} \Phi(|f_k - \tilde{s}_k|).$$

So given $\eta > 0$, there exists an N making the final term $< \frac{\eta}{2}$. Then for λ sufficiently small the previous term is also $< \frac{\eta}{2}$. So we see that $\sup_n \Phi(\lambda|f_n - s_n|) \rightarrow 0$ as $\lambda \downarrow 0$. Therefore $\sup_n \Phi(\lambda s_n) \rightarrow 0$, since $\sup_n \Phi(\lambda s_n) \leq \sup_n \Phi(\lambda|f_n - s_n|) + \sup_n \Phi(\lambda f_n)$. So we see that (4) of Definition(2.1)(b) applies to the sequence $(s_n)_n$. Thus $\Phi(s_n - s_m) \rightarrow 0$ as $n, m \rightarrow \infty$. Take N such that $\Phi(s_n - s_m) < \varepsilon$ for all $n, m \geq N$. Then by (‡), $\Phi(f_n - f_m) < 3\varepsilon$ for all $n, m \geq N$. Thus $\Phi(f_n - f_m) \rightarrow 0$ as $n, m \rightarrow \infty$, and so since $L^1(\Phi)$ is complete, it follows that there is an $f \in \mathcal{L}^1(\Phi)$ such that $\Phi(f_n - f) \rightarrow 0$ as $n \rightarrow \infty$. By Proposition(2.10), we see that a subsequence of $(f_n)_n$ converges to f Φ -a.e. and so $f = \sup_n f_n$ Φ -a.e. $\square$

(2.14) The Dominated Convergence Theorem.

Suppose that $f_1, f_2, \dots \in \mathcal{L}^1(\Phi)$, $f: \Gamma \rightarrow \mathbb{R}$, $f_n \rightarrow f$ Φ -a.e. as $n \rightarrow \infty$, and for some non-negative $g \in \mathcal{F}(\Phi)$, $|f_n| \leq g$ Φ -a.e. for all n . Then $f \in \mathcal{L}^1(\Phi)$ and $\Phi(f_n - f) \rightarrow 0$ as $n \rightarrow \infty$.

Proof: By replacing f_n , f and g by $\mathbf{1}_{\Gamma \setminus E} f_n$, $\mathbf{1}_{\Gamma \setminus E} f$ and $\mathbf{1}_{\Gamma \setminus E} g$ respectively, where E is the Φ -null set outside which $f_n \rightarrow f$ pointwise and $|f_n| \leq g$ for all n , we can assume without loss of generality that $f_n \rightarrow f$ everywhere and $|f_n| \leq g$ everywhere for all n .

$$\text{Let } h_n = \sup\{|f_k - f_l| \mid k, l \geq n\},$$

$$\text{and } h_{n,m} = \sup\{|f_k - f_l| \mid n \leq k, l \leq m\} \quad \text{for } m \geq n.$$

Then $h_{n,m} \in \mathcal{L}^1(\Phi)$, $0 \leq h_n \leq 2g$ and $h_{n,m} \uparrow h_n$ as $m \rightarrow \infty$. Thus we see that $h_n \in \mathcal{L}^1(\Phi)$ by the Monotone Convergence Theorem (2.13). Now $h_n \downarrow 0$ pointwise so $h_1 - h_n \uparrow h_1$. Thus applying the Monotone Convergence Theorem (2.13) again gives that $\Phi(h_n) = \Phi(h_1 - (h_1 - h_n)) \rightarrow 0$ as $n \rightarrow \infty$. But since $|f_n - f_m| \leq h_{n \wedge m}$, it follows that $\Phi(f_n - f_m) \rightarrow 0$ as $n, m \rightarrow \infty$. Therefore by the completeness of $L^1(\Phi)$ that there exists an $\tilde{f} \in \mathcal{L}^1(\Phi)$ such that $\Phi(f_n - \tilde{f}) \rightarrow 0$. Now using Proposition(2.10) we see that $\tilde{f} = f$ Φ -a.e. The theorem now follows. $\square$

§2.5 Φ -MEASURABILITY

We would like to have some concept of measurable function when considering upper gauges. We think of Φ -measurable functions as those functions that would be Φ -integrable if only they were not too big. We make this concept precise by ‘clipping’ the function with an element of $\mathcal{E}^+$ as follows.

(2.15) Definition. A function $f: \Gamma \rightarrow \overline{\mathbb{R}}$ is said to be Φ -measurable if and only if $(f \wedge s) \vee (-s) \in \mathcal{L}^1(\Phi)$ for each $s \in \mathcal{E}^+$. □

We note immediately that all Φ -integrable functions are Φ -measurable. Now we show that Φ -measurable functions that are not too big are Φ -integrable.

(2.16) Proposition. If $f: \Gamma \rightarrow \mathbb{R}$ is Φ -measurable and $|f| \leq g$ for some $g \in \mathcal{L}^1(\Phi)$, then $f \in \mathcal{L}^1(\Phi)$.

Proof: Take $s_n \in \mathcal{E}^+$ such that $\Phi(s_n - g) \rightarrow 0$, and let $f_n = (f \wedge s_n) \vee (-s_n)$. So $f_n \in \mathcal{L}^1(\Phi)$. But $f = (f \wedge g) \vee (-g)$, and so by Lemma(2.12) we have that $|f_n - f| \leq |s_n - g|$. Thus $\Phi(s_n - g) \rightarrow 0$ gives that $\Phi(f_n - f) \rightarrow 0$. But $f_n \in \mathcal{L}^1(\Phi)$, so $f \in \mathcal{L}^1(\Phi)$ as required. □

(2.17) Proposition. If $f_1, f_2, \dots: \Gamma \rightarrow \overline{\mathbb{R}}$ are Φ -measurable, and $f: \Gamma \rightarrow \overline{\mathbb{R}}$ is such that $f_n \rightarrow f$ Φ -a.e., then f is Φ -measurable.

Proof: Suppose that $s \in \mathcal{E}^+$. Then $(f_n \wedge s) \vee (-s) \rightarrow (f \wedge s) \vee (-s)$ Φ -a.e. But also $|(f_n \wedge s) \vee (-s)| \leq s$ and $(f_n \wedge s) \vee (-s) \in \mathcal{L}^1(\Phi)$. Thus it follows by the Dominated Convergence Theorem (2.14) that $(f \wedge s) \vee (-s) \in \mathcal{L}^1(\Phi)$ as required. □

Now we show that Φ -measurability corresponds to the usual measure theoretic idea of measurable function (that defined by the use of σ -fields).

(2.18) **Definition.** If $A \subseteq \Gamma$ then A is said to be Φ -measurable if and only if $\mathbf{1}_A$ is Φ -measurable. $\mathcal{M}(\Phi)$ will denote the set of all Φ -measurable subsets of Γ . $\square$

(2.19) **Proposition.**

- (i) $\mathcal{M}(\Phi)$ is a σ -field on Γ that contains all the Φ -null subsets of Γ .
- (ii) Suppose that $f: \Gamma \rightarrow \overline{\mathbb{R}}$. Then f is Φ -measurable if and only if $f^{-1}(B) \in \mathcal{M}(\Phi)$ for all Borel subsets B of $\overline{\mathbb{R}}$.

Proof:

(i) It is easy to show that Φ -null sets are in $\mathcal{M}(\Phi)$. Thus $\emptyset \in \mathcal{M}(\Phi)$. It remains only to show that the following hold.

$$(a) A \in \mathcal{M}(\Phi) \implies \Gamma \setminus A \in \mathcal{M}(\Phi).$$

$$(b) A_1, A_2, \dots \in \mathcal{M}(\Phi) \implies \bigcup_{n=1}^{\infty} A_n \in \mathcal{M}(\Phi).$$

For (a) suppose that $A \in \mathcal{M}(\Phi)$ and that $s \in \mathcal{E}^+$. Then $\mathbf{1}_{\Gamma \setminus A} \wedge s = 1 \wedge s - \mathbf{1}_A \wedge s \in \mathcal{L}^1(\Phi)$. So $\Gamma \setminus A \in \mathcal{M}(\Phi)$. For (b) let $A = \bigcup_{n=1}^{\infty} A_n$. Then for $s \in \mathcal{E}^+$, $\mathbf{1}_A \wedge s = \lim_{N \rightarrow \infty} \bigvee_{n=1}^N (\mathbf{1}_{A_n} \wedge s) \in \mathcal{L}^1(\Phi)$ by the Monotone Convergence Theorem (2.13) or the Dominated Convergence Theorem (2.14). Thus $A \in \mathcal{M}(\Phi)$ as required.

(ii) Suppose first that f is Φ -measurable. It is enough to show that $[f > \alpha] \in \mathcal{M}(\Phi)$ for all $\alpha \in \mathbb{R}$, since the intervals $(\alpha, \infty]$ generate the σ -field of Borel subsets of $\overline{\mathbb{R}}$.

First we show that $(f - \alpha)^+$ is Φ -measurable. Suppose that $s \in \mathcal{E}^+$, and let $s_n = (ns) \wedge 1$. So $s_n \in \mathcal{E}^+$ and $s_n \uparrow \mathbf{1}_{[s > 0]}$. Let $f_n = (f \wedge ns) \vee (-ns)$ and $g_n = (f_n - \alpha s_n)^+ \wedge s$. So $f_n, g_n \in \mathcal{L}^1(\Phi)$, $|g_n| \leq s$ and $g_n \rightarrow (f - \alpha)^+ \wedge s$. Therefore by the Dominated Convergence Theorem (2.14) we have that $(f - \alpha)^+ \wedge s \in \mathcal{L}^1(\Phi)$. Thus it follows by definition that $(f - \alpha)^+$ is Φ -measurable.

Now let $h_n = (n(f - \alpha)^+) \wedge 1$. Then for $s \in \mathcal{E}^+$, $h_n \wedge s = n((f - \alpha)^+ \wedge \frac{1}{n}(s \wedge 1)) \in \mathcal{L}^1(\Phi)$, since $\frac{1}{n}(s \wedge 1) \in \mathcal{E}^+$. So h_n is Φ -measurable. But $h_n \rightarrow \mathbf{1}_{[f > \alpha]}$, and thus $[f > \alpha] \in \mathcal{M}(\Phi)$ by Proposition (2.17). It now follows that $f^{-1}(B) \in \mathcal{M}(\Phi)$ for all Borel subsets B of $\overline{\mathbb{R}}$ as required.

Now suppose conversely that $f^{-1}(B) \in \mathcal{M}(\Phi)$ for all Borel subsets B of $\overline{\mathbb{R}}$. Then let

$$f_n = \sum_{k=-2^{2n}}^{2^{2n}-1} k 2^{-n} \mathbf{1}_{[k 2^{-n} \leq f < (k+1) 2^{-n}]} - 2^n \mathbf{1}_{[f < -2^n]} + 2^n \mathbf{1}_{[f \geq 2^n]}.$$

Then $f_n \rightarrow f$, and so it is enough by Proposition(2.17) to show that f_n is Φ -measurable. Thus without loss of generality $f = \sum_{k=1}^n \alpha_k \mathbf{1}_{A_k}$, where $\alpha_1, \dots, \alpha_n \in \mathbb{R} \setminus \{0\}$ and $A_1, \dots, A_n$ are disjoint Φ -measurable subsets of Γ . Suppose that $s \in \mathcal{E}^+$. Then $(f \wedge s) \vee (-s) = \sum_{k=1}^n \alpha_k (\mathbf{1}_{A_k} \wedge \frac{1}{|\alpha_k|} s) \in \mathcal{L}^1(\Phi)$ (since $A_1, \dots, A_n$ are disjoint). Thus f is Φ -measurable as required. $\square$

As a corollary we have that the sum, product, maximum, minimum etc. of Φ -measurable functions are Φ -measurable. The following shows that one can replace $\mathcal{E}$ by $\mathcal{L}^1(\Phi)$ in Definition(2.15).

(2.20)Proposition. *Suppose that $f: \Gamma \rightarrow \overline{\mathbb{R}}$. Then f is Φ -measurable if and only if $(f \wedge g) \vee (-g) \in \mathcal{L}^1(\Phi)$ for all $g \in \mathcal{L}^1(\Phi)$ with $g \geq 0$.*

Proof: The ‘if’ part is easy since $\mathcal{E} \subseteq \mathcal{L}^1(\Phi)$. For the ‘only if’ part suppose that f is Φ -measurable and $g \in \mathcal{L}^1(\Phi)$ with $g \geq 0$. Then $(f \wedge g) \vee (-g)$ is Φ -measurable by the above comment and $|(f \wedge g) \vee (-g)| \leq g$. So by Proposition(2.16) it follows that $(f \wedge g) \vee (-g) \in \mathcal{L}^1(\Phi)$ as required. $\square$

§2.6 σ -FINITE UPPER GAUGE SPACES

(2.21)Definition. An upper gauge space $(\Gamma, \mathcal{E}, \Phi)$ is said to be σ -finite if and only if there exist $s_1, s_2, \dots \in \mathcal{E}$ such that $\Gamma = \bigcup_{n=1}^{\infty} [s_n \neq 0]$. $\square$

We note immediately the following result.

(2.22)Proposition. *Suppose that $(\Gamma, \mathcal{E}, \Phi)$ is a σ -finite upper gauge space. Then there exist $s_1, s_2, \dots \in \mathcal{E}^+$ such that $s_n \uparrow \infty$ pointwise as $n \rightarrow \infty$.*

Proof: We have that there are $t_1, t_2, \dots \in \mathcal{E}$ such that $\Gamma = \bigcup_{n=1}^{\infty} [t_n \neq 0]$. Let $s_n = n \sum_{k=1}^n |t_k|$. Then $s_n \uparrow \infty$ as required. $\square$

Now we can give a generalization of Proposition(2.16) in the case of σ -finite upper gauge spaces.

(2.23)Proposition. *Suppose that $(\Gamma, \mathcal{E}, \Phi)$ is a σ -finite upper gauge space, and that $f: \Gamma \rightarrow \mathbb{R}$ is Φ -measurable and Φ -finite. Then f is Φ -integrable.*

Proof: By Proposition(2.22), let $s_1, s_2, \dots \in \mathcal{E}^+$ be such that $s_n \uparrow \infty$. Let $f_n = (f \wedge s_n) \vee (-s_n)$. So that $f_n \in \mathcal{L}^1(\Phi)$, $f_n \rightarrow f$ and $|f_n| \leq |f| \in \mathcal{F}(\Phi)$. Thus by the Dominated Convergence Theorem (2.14), we have $f \in \mathcal{L}^1(\Phi)$ as required. $\square$

In the σ -finite case we have another characterization of $\mathcal{M}(\Phi)$.

(2.24)Proposition. *Suppose that $(\Gamma, \mathcal{E}, \Phi)$ is a σ -finite upper gauge space. Then $\mathcal{M}(\Phi)$ is the smallest σ -field containing all Φ -null subsets of Γ and making all elements of $\mathcal{E}$ measurable.*

Proof: Let $\mathcal{F}$ be the smallest σ -field containing all Φ -null subsets of Γ and making all elements of $\mathcal{E}$ measurable. Now clearly $\mathcal{F} \subseteq \mathcal{M}(\Phi)$.

Suppose that $f \in \mathcal{L}^1(\Phi)$. Then by Definition(2.11) and Proposition(2.10) there exist $s_1, s_2, \dots \in \mathcal{E}$ such that $s_n \rightarrow f$ Φ -a.e. Suppose $s_n \rightarrow f$ outside the Φ -null set E . Let $\tilde{s}_n = \mathbf{1}_{\Gamma \setminus E} s_n$ and $\tilde{f} = \mathbf{1}_{\Gamma \setminus E} f$. Then s_n is $\mathcal{F}$ -measurable, therefore so is $\tilde{s}_n$ (since $\tilde{s}_n = s_n$ Φ -a.e.). But $\tilde{s}_n \rightarrow \tilde{f}$ so $\tilde{f}$ is $\mathcal{F}$ -measurable. Therefore f is $\mathcal{F}$ -measurable (again since $\tilde{f} = f$ Φ -a.e.). We have shown therefore that all Φ -integrable functions are $\mathcal{F}$ -measurable.

To complete the proof suppose that $A \in \mathcal{M}(\Phi)$. We have to show that $A \in \mathcal{F}$. By Proposition(2.22) let $s_1, s_2, \dots \in \mathcal{E}^+$ be such that $s_n \uparrow \infty$. Now $\mathbf{1}_A \wedge s_n \in \mathcal{L}^1(\Phi)$ by the definition of Φ -measurability (Definition(2.15) and (2.18)). Thus by above $\mathbf{1}_A \wedge s_n$

is $\mathcal{F}$ -measurable and so by letting $n \rightarrow \infty$ we see that $\mathbf{1}_A$ is $\mathcal{F}$ -measurable. I.e. $A \in \mathcal{F}$ as required. $\square$

CHAPTER 3: VECTOR-VALUED INTEGRALS

§3.1 DEFINITIONS

In this chapter we shall apply the theory of upper gauges given in Chapter 2 to the following extension problem. We are given a set Γ with a Stone lattice $\mathcal{E}$ on it, an F -space $\mathcal{X}$, and a linear map $J: \mathcal{E} \rightarrow \mathcal{X}$. We require an extension of J to some ' L^1 -closure' of $\mathcal{E}$ (i.e. a closure with respect to an upper gauge). In order to do this we shall have to place restrictions on J and $\mathcal{X}$. The restriction on $\mathcal{X}$ is that it must be a C -space, a concept that we shall now define. (This restriction is necessary in order to prove that property (4) of Definition(2.1)(b) is satisfied by our upper gauge.)

(3.1)Definition. A C -space is an F -space $\mathcal{X}$ such that if $x_1, x_2, \dots \in \mathcal{X}$ are such that the set

$$\mathcal{A} = \{ \sum_{n=1}^N \alpha_n x_n \mid N \in \mathbb{N}_{>0}, \alpha_n \in F \text{ and } |\alpha_n| \leq 1 \text{ for } n = 1, \dots, N \},$$

(where F denotes the scalar field) is bounded in $\mathcal{X}$, then $x_n \rightarrow 0$ in $\mathcal{X}$. □

We note that any closed subspace of a C -space is also a C -space. The spaces that we are interested in, namely L^p ($0 \leq p < \infty$) based on a probability space, will, in §3.7, be shown to be C -spaces. Not all Banach spaces are C -spaces (consider for example L^∞ in Proposition(3.20), also the sequence space c_0 is not a C -space). However all Hilbert spaces are C -spaces. We give here an elementary proof of this fact.

(3.2)Proposition. Suppose that $\mathcal{H}$ is a Hilbert space. Then $\mathcal{H}$ is a C -space. In fact if $x_1, x_2, \dots \in \mathcal{H}$ have the property that for some constant $c \in \mathbb{R}^+$, $\|x_{n_1} + \dots + x_{n_r}\| \leq c$ whenever $r \in \mathbb{N}_{>0}$ and $n_1 < \dots < n_r$, then $x_n \rightarrow 0$.

Proof: By thinking of $\mathcal{H}$ as a real space with the inner product $\text{Re}\langle \cdot, \cdot \rangle$, we may suppose that $\mathcal{H}$ is a real Hilbert space.

Suppose that $y \in \mathcal{H}$ and $N \in \mathbb{N}_{>0}$. Then take $n_1 < \dots < n_r$ to be the elements n of $\{1, \dots, N\}$ such that $\langle x_n, y \rangle > 0$. We have that $\sum_{n=1}^N \langle x_n, y \rangle^+ = \langle x_{n_1} + \dots + x_{n_r}, y \rangle \leq c\|y\|$. Let $N \rightarrow \infty$ to give that $\sum_{n=1}^{\infty} \langle x_n, y \rangle^+ \leq c\|y\|$. Similarly $\sum_{n=1}^{\infty} \langle x_n, y \rangle^- \leq c\|y\|$, and thus adding this to the previous formula gives $\sum_{n=1}^{\infty} |\langle x_n, y \rangle| \leq 2c\|y\|$. Thus we have that

$$(\dagger) \quad \langle x_n, y \rangle \rightarrow 0 \quad \text{as } n \rightarrow \infty \text{ for all } y \in \mathcal{H}.$$

Now suppose for a contradiction that $x_n \nrightarrow 0$. Then there exists $\varepsilon > 0$ such that $\|x_n\| \geq \varepsilon$ for infinitely many n . Define $n_1 < n_2 < \dots$ inductively as follows: let n_1 be such that $\|x_{n_1}\| \geq \varepsilon$, given $n_1 < \dots < n_{k-1}$ with $k > 1$, take using $(\dagger)$ $n_k > n_{k-1}$ such that $\|x_{n_k}\| \geq \varepsilon$ and $|\langle x_{n_j}, x_{n_k} \rangle| < 2^{-k}$ for $j < k$. We have

$$\begin{aligned} \sum_{1 \leq j < k \leq r} |\langle x_{n_j}, x_{n_k} \rangle| &< \sum_{j=1}^{r-1} \sum_{k=j+1}^r 2^{-k} \\ &< \sum_{j=1}^{r-1} 2^{-j} \\ &< 1. \end{aligned}$$

Therefore

$$\begin{aligned} c^2 &\geq \|x_{n_1} + \dots + x_{n_r}\|^2 \\ &= \sum_{k=1}^r \|x_{n_k}\|^2 + 2 \sum_{1 \leq j < k \leq r} \langle x_{n_j}, x_{n_k} \rangle \\ &> r\varepsilon - 2. \end{aligned}$$

Letting $r \rightarrow \infty$ gives the required contradiction. □

(3.3)Definition. Suppose that $\mathcal{X}$ is a real C-space. Then an $\mathcal{X}$ -valued integral space is a triple $(\Gamma, \mathcal{E}, J)$, where Γ is a set, $\mathcal{E}$ is a Stone lattice on Γ , and $J: \mathcal{E} \rightarrow \mathcal{X}$ is linear and satisfies the following.

- (1) $\{J(u) \mid u \in \mathcal{E}, |u| \leq s\}$ is a bounded subset of $\mathcal{X}$ for all $s \in \mathcal{E}^+$.
- (2) $s_1, s_2, \dots \in \mathcal{E}, s_n \downarrow 0$ pointwise $\implies J(s_n) \rightarrow 0$ in $\mathcal{X}$.

J is then known as an $\mathcal{X}$ -valued preintegral on $(\Gamma, \mathcal{E})$. We shall say that $(\Gamma, \mathcal{E}, J)$ is σ -finite if and only if $\Gamma = \bigcup_{n=1}^{\infty} [s_n \neq 0]$ for some $s_1, s_2, \dots \in \mathcal{E}$. $\square$

These axioms correspond to those of Bichteler for an stochastic integrator in [Bi2]. We shall now suppose unless otherwise indicated that $\mathcal{X}$ is a real C-space, and that $(\Gamma, \mathcal{E}, J)$ is an $\mathcal{X}$ -valued integral space. Also in order to construct an upper gauge we shall suppose also that $\mathcal{X}$ comes equipped with a (translation invariant) metric ρ .

§3.2 THE UPPER GAUGE CORRESPONDING TO J

(3.4)Definition. $\mathcal{E}_{\text{inc}}^+$ denotes the set of all functions $u: \Gamma \rightarrow \overline{\mathbb{R}}^+$ such that there exist $s_1, s_2, \dots \in \mathcal{E}^+$ such that $s_n \uparrow u$ pointwise. Now define $\Phi_J: (\overline{\mathbb{R}}^+)^{\Gamma} \rightarrow \overline{\mathbb{R}}^+$ by

$$\begin{aligned}\Phi_J(u) &= \sup\{\rho(J(s)) \mid s \in \mathcal{E}, |s| \leq u\}, & \text{when } u \in \mathcal{E}_{\text{inc}}^+, \\ \Phi_J(f) &= \inf\{\Phi_J(u) \mid f \leq u \in \mathcal{E}_{\text{inc}}^+\}, & \text{when } f: \Gamma \rightarrow \overline{\mathbb{R}}^+.\end{aligned}$$

(It is clear that the first formula defines an increasing function Φ_J on $\mathcal{E}_{\text{inc}}^+$, and thus the second formula extends the first. So we see that the definition of Φ_J is consistent.) Again we put $\Phi_J(f) = \Phi_J(|f|)$ when $f: \Gamma \rightarrow \overline{\mathbb{R}}$. $\square$

(3.5)Theorem. Φ_J is an upper gauge on $(\Gamma, \mathcal{E})$.

In order to prove this we shall require the following result, which is of interest in its own right. First note that it is clear that Φ_J is increasing; i.e. it satisfies (2) of Definition(2.1)(b).

(3.6) Proposition. *If $u_1, u_2, \dots \in \mathcal{E}_{\text{inc}}^+$, $u: \Gamma \rightarrow \overline{\mathbb{R}}^+$ and $u_n \uparrow u$ pointwise, then $u \in \mathcal{E}_{\text{inc}}^+$ and $\Phi_J(u_n) \uparrow \Phi_J(u)$.*

Proof: For each $k = 1, 2, \dots$, take $s_{k,1}, s_{k,2}, \dots \in \mathcal{E}^+$ such that $s_{k,n} \uparrow u_k$ as $n \rightarrow \infty$. Let $s_n = s_{1,n} \vee \dots \vee s_{n,n}$. So $s_n \in \mathcal{E}^+$, and it is easy to check that $s_n \uparrow u$. So $u \in \mathcal{E}_{\text{inc}}^+$.

Now suppose that $0 \leq c < \Phi_J(u)$. Then by definition there exists $t \in \mathcal{E}$ such that $|t| \leq u$ and $\rho(J(t)) > c$. Put $t_n = (t \wedge s_n) \vee (-s_n)$. Then $t_n \in \mathcal{E}$ and $|t_n| \leq s_n \leq u_n$. Now since $t = t^+ - t^-$ and $[t^+ \neq 0] \cap [t^- \neq 0] = \emptyset$ we obtain that $t_n^\pm = t^\pm \wedge s_n \uparrow t^\pm$. Therefore applying (2) of Definition(3.3) to the decreasing sequence $(t^\pm - t_n^\pm)_n$, we obtain that $J(t_n^\pm) \rightarrow J(t^\pm)$ in $\mathcal{X}$. So $J(t_n) \rightarrow J(t)$ in $\mathcal{X}$. Thus $\rho(J(t_n)) \rightarrow \rho(J(t)) > c$, and so there exists n_0 such that $\rho(J(t_{n_0})) > c$. But for $n \geq n_0$, we have that $|t_{n_0}| \leq s_{n_0} \leq u_n$ and so $\Phi_J(u_n) \geq \rho(J(t_{n_0})) > c$ for all $n \geq n_0$. $\Phi_J(u_n) \uparrow \Phi_J(u)$ now follows. $\square$

Proof of Theorem(3.5): We have to prove that (1), (2), (3) and (4) of Definition(2.1)(b) hold.

(1) Suppose that $s \in \mathcal{E}^+$. Put $\mathcal{A} = \{J(t) | t \in \mathcal{E}, |t| \leq s\}$. Then by (1) of Definition(3.3) $\mathcal{A}$ is bounded in $\mathcal{X}$. Thus $\Phi_J(\lambda s) = \sup_{x \in \mathcal{A}} \rho(\lambda x) \rightarrow 0$ as $\lambda \downarrow 0$.

(2) This is easy.

(3) First I claim that

$$(\dagger) \quad \Phi_J(s+t) \leq \Phi_J(s) + \Phi_J(t) \quad \text{for all } s, t \in \mathcal{E}^+.$$

Suppose that $u \in \mathcal{E}$ is such that $|u| \leq s+t$. Let $v = (u \wedge s) \vee (-s)$ and $w = u - v$. So $v, w \in \mathcal{E}$, $|v| \leq s$, $|w| \leq t$ and $u = v + w$. We have that $J(u) = J(v) + J(w)$, and so $\rho(J(u)) \leq \rho(J(v)) + \rho(J(w)) \leq \Phi_J(s) + \Phi_J(t)$. $(\dagger)$ now follows from the definition of Φ_J .

Next I claim that

$$(\ddagger) \quad \Phi_J(u+v) \leq \Phi_J(u) + \Phi_J(v) \quad \text{for all } u, v \in \mathcal{E}_{\text{inc}}^+.$$

Take $s_1, s_2, \dots \in \mathcal{E}^+$ and $t_1, t_2, \dots \in \mathcal{E}^+$ such that $s_n \uparrow u$ and $t_n \uparrow v$. Then by $(\dagger)$, $\Phi_J(s_n + t_n) \leq \Phi_J(s_n) + \Phi_J(t_n)$. Let $n \rightarrow \infty$ and use Proposition(3.6) to give $(\ddagger)$.

Now suppose that $f_1, f_2, \dots: \Gamma \rightarrow \overline{\mathbb{R}}^+$. Then we have to show that

$$(*) \quad \Phi_J\left(\sum_{n=1}^{\infty} f_n\right) \leq \sum_{n=1}^{\infty} \Phi_J(f_n).$$

Clearly we may suppose that the right hand side is finite. Given $\varepsilon > 0$ take $u_n \in \mathcal{E}_{\text{inc}}^+$ such that $f_n \leq u_n$ and $\Phi_J(u_n) < \Phi_J(f_n) + \varepsilon 2^{-n}$. Then

$$\begin{aligned} \Phi_J\left(\sum_{n=1}^{\infty} f_n\right) &\leq \Phi_J\left(\sum_{n=1}^{\infty} u_n\right) \\ &= \lim_{N \rightarrow \infty} \Phi_J\left(\sum_{n=1}^N u_n\right) \quad (\text{by Proposition(3.6)}) \\ &\leq \lim_{N \rightarrow \infty} \sum_{n=1}^N \Phi_J(u_n) \quad (\text{by } (\ddagger)) \\ &< \sum_{n=1}^{\infty} \Phi_J(f_n) + \varepsilon. \end{aligned}$$

And thus $(*)$ now follows.

(4) Suppose that $s_1 \leq s_2 \leq \dots$ in $\mathcal{E}^+$ and $\sup_n \Phi_J(\lambda s_n) \rightarrow 0$ as $\lambda \downarrow 0$. Let $u = \lim_{n \rightarrow \infty} s_n$ so $u \in \mathcal{E}_{\text{inc}}^+$. Now $\Phi_J(u - s_n)$ decreases as n increases. I claim $\Phi_J(u - s_n) \downarrow 0$. For suppose not. Then there exists $\varepsilon > 0$ such that $\Phi_J(u - s_n) > \varepsilon$ for all n . By Proposition(3.6) it follows that for each n , $\Phi_J(s_m - s_n) \uparrow \Phi_J(u - s_n) > \varepsilon$ as $m \rightarrow \infty$ with $m > n$. Thus we must have that for each n , there is an $m > n$ such that $\Phi_J(s_m - s_n) > \varepsilon$. Using this take $n_1 < n_2 < \dots$ such that $\Phi_J(s_{n_{k+1}} - s_{n_k}) > \varepsilon$ for $k = 1, 2, \dots$. By the definition of Φ_J there exists for each k , a $t_k \in \mathcal{E}$ such that $|t_k| \leq s_{n_{k+1}} - s_{n_k}$ and $\rho(J(t_k)) > \varepsilon$. Let $x_k = J(t_k)$, so $\rho(x_k) > \varepsilon$, and put

$$\mathcal{A} = \left\{ \sum_{k=1}^N \alpha_k x_k \mid N \in \mathbb{N}_{>0} \quad \alpha_1, \dots, \alpha_N \in [-1, 1] \right\}.$$

I claim that $\mathcal{A}$ is bounded in $\mathcal{X}$. For if $x = \sum_{k=1}^N \alpha_k x_k \in \mathcal{A}$ ($-1 \leq \alpha_1, \dots, \alpha_N \leq 1$), then $x = J(t)$, where $t = \sum_{k=1}^N \alpha_k t_k$. So $t \in \mathcal{E}$ and $|t| \leq \sum_{k=1}^N |t_k| \leq s_{n_{N+1}}$. Therefore for $\lambda > 0$, we have that $\rho(\lambda x) = \rho(J(\lambda t)) \leq \Phi_J(\lambda s_{n_{N+1}})$, and so $\sup_{x \in \mathcal{A}} \rho(\lambda x) \leq \sup_n \Phi_J(\lambda s_n) \rightarrow 0$ as $\lambda \downarrow 0$. Thus $\mathcal{A}$ is bounded. But $\mathcal{X}$ is a C-space; therefore $x_k \rightarrow 0$ as $k \rightarrow \infty$, which gives a contradiction, since $\rho(x_k) > \varepsilon$ for all k . $\square$

§3.3 THE INTEGRAL AND THE CONVERGENCE THEOREMS

We can now use this upper gauge to extend J to a larger space than $\mathcal{E}$.

(3.7) Definition. Let $\mathcal{L}^1(J)$ denote $\mathcal{L}^1(\Phi_J)$. Then $\int \dots dJ: \mathcal{L}^1(J) \rightarrow \mathcal{X}$ is the unique linear map extending J and satisfying:

$$(\dagger) \quad \rho\left(\int f dJ\right) \leq \Phi_J(f) \quad \text{for all } f \in \mathcal{L}^1(J).$$

Consistency: $(\dagger)$ implies that $\int \dots dJ: \mathcal{L}^1(J) \rightarrow \mathcal{X}$ is continuous (with the topology given by Φ_J) and therefore it is unique if it exists (since $\mathcal{E}$ is dense in $\mathcal{L}^1(J)$). To prove the existence of $\int \dots dJ$ note that $\rho(J(s)) \leq \Phi_J(s)$ for all $s \in \mathcal{E}$ by the definition of Φ_J . So $J: \mathcal{E} \rightarrow \mathcal{X}$ is continuous when $\mathcal{E}$ has the topology given by Φ_J . Therefore (using approximating sequences and the completeness of $\mathcal{X}$) there is a continuous linear extension of J to $\mathcal{L}^1(J)$. To prove $(\dagger)$, suppose that $f \in \mathcal{L}^1(J)$ and take $s_1, s_2, \dots \in \mathcal{E}$ such that $\Phi_J(s_n - f) \rightarrow 0$. Then $J(s_n) \rightarrow \int f dJ$. But $\rho(J(s_n)) \leq \Phi_J(s_n)$. So let $n \rightarrow \infty$ to give $(\dagger)$. $\square$

Now using the Monotone Convergence Theorem (2.13), the Dominated Convergence Theorem (2.14), and the fact that $\Phi_J(f_n - f) \rightarrow 0 \implies \int f_n dJ \rightarrow \int f dJ$ (by $(\dagger)$ of Definition(3.7)), we have immediately the following results.

(3.8) The Monotone Convergence Theorem.

Suppose that $f_1, f_2, \dots \in \mathcal{L}^1(J)$ are such that $0 \leq f_1 \leq f_2 \leq \dots$ and $\sup_n \Phi_J(\lambda f_n) \rightarrow 0$ as $\lambda \downarrow 0$. Then there exists an $f \in \mathcal{L}^1(J)$ such that $f = \sup_n f_n$ Φ_J -a.e., $\Phi_J(f - f_n) \rightarrow 0$ and $\int f_n dJ \rightarrow \int f dJ$ in $\mathcal{X}$. $\square$

(3.9) The Dominated Convergence Theorem.

Suppose that $f_1, f_2, \dots \in \mathcal{L}^1(J)$, $f: \Gamma \rightarrow \mathbb{R}$, $f_n \rightarrow f$ Φ_J -a.e. as $n \rightarrow \infty$, and for some non-negative $g \in \mathcal{F}(\Phi_J)$, $|f_n| \leq g$ Φ_J -a.e. for all n . Then $f \in \mathcal{L}^1(J)$, $\Phi_J(f_n - f) \rightarrow 0$ and $\int f_n dJ \rightarrow \int f dJ$ in $\mathcal{X}$. $\square$

We shall need the following result in the next chapter.

(3.10) Proposition. Suppose that $\mathcal{X}$ is a C -space and that $(\Gamma, \mathcal{E}, J)$ is a σ -finite $\mathcal{X}$ -valued integral space. Let $\mathcal{F}$ be the smallest σ -field making all the elements of $\mathcal{E}$ measurable. Then the following hold.

- (i) If $E \subseteq \Gamma$ is Φ_J -null then there exists an $A \in \mathcal{F}$ such that $E \subseteq A$ and A is Φ_J -null.
- (ii) $\mathcal{M}(\Phi_J)$ is the set of subsets E of Γ such that there exist $A, B \in \mathcal{F}$ such that $A \subseteq E \subseteq B$ and $B \setminus A$ is Φ_J -null.
- (iii) If $f: \Gamma \rightarrow \mathbb{R}$ (resp. $\overline{\mathbb{R}}$) is Φ_J -measurable then there exists $\mathcal{F}$ -measurable function $g: \Gamma \rightarrow \mathbb{R}$ (resp. $\overline{\mathbb{R}}$) such that $f = g$ Φ_J -a.e.

Proof:

(i) We have that $\Phi_J(\mathbf{1}_E) = 0$, and so by Definition(3.4), there exist $u_1, u_2, \dots \in \mathcal{E}_{\text{inc}}^+$ such that $\mathbf{1}_E \leq u_n$ for all n , and $\Phi_J(u_n) \downarrow 0$. Now since $u_n \in \mathcal{E}_{\text{inc}}^+$, we have that u_n is $\mathcal{F}$ -measurable. Let $A = \bigcap_{n=1}^{\infty} [u_n \geq 1]$, so $A \in \mathcal{F}$ and $E \subseteq A$. But $\mathbf{1}_A \leq u_n$, and so $\Phi_J(\mathbf{1}_A) \leq \Phi_J(u_n)$. Let $n \rightarrow \infty$ to give that $\Phi_J(\mathbf{1}_A) = 0$ as required.

(ii) Letting $\mathcal{G}$ denote the set described in the statement, we have to show that $\mathcal{G} = \mathcal{M}(\Phi_J)$. We know by Proposition(2.24) that $\mathcal{M}(\Phi_J)$ is the σ -field generated by $\mathcal{F}$

and the Φ_J -null subsets of Γ . Now $\mathcal{G} \subseteq \mathcal{M}(\Phi_J)$, since if $E \in \mathcal{G}$, take A and B as in the statement. Then E is the union of $A \in \mathcal{F}$ and the Φ_J -null set $E \setminus A$. To prove the reverse inclusion note that $\mathcal{G}$ is a σ -field and that $\mathcal{F} \subseteq \mathcal{G}$. We have only to check that $\mathcal{G}$ contains all the Φ_J -null subsets of Γ . Now if $E \subseteq \Gamma$ is Φ_J -null then by (i) there exists $B \in \mathcal{F}$ such that $E \subseteq B$ and B is Φ_J -null. Take $A = \emptyset$ to show that $E \in \mathcal{G}$.

(iii) Let $R = \mathbb{R}$ (resp. $\overline{\mathbb{R}}$), and let $\mathcal{C}$ be a countable class of subsets of R that generates, as a σ -field, the Borel σ -field of R . For each $C \in \mathcal{C}$, by (ii), let $A_C, B_C \in \mathcal{F}$ be such that $A_C \subseteq f^{-1}(C) \subseteq B_C$ and $B_C \setminus A_C$ is Φ_J -null. Let $E = \bigcup_{C \in \mathcal{C}} (B_C \setminus A_C)$; so E is a Φ_J -null element of $\mathcal{F}$ (since $\mathcal{C}$ is countable). Define $g: \Gamma \rightarrow R$ by

$$g(x) = \begin{cases} f(x), & \text{if } x \in \Gamma \setminus E, \\ 0, & \text{if } x \in E. \end{cases}$$

Then $g = f$ Φ_J -a.e. Now for each $C \in \mathcal{C}$, $g^{-1}(C) = A_C \setminus E \in \mathcal{F}$ if $0 \notin C$, and $g^{-1}(C) = B_C \cup E \in \mathcal{F}$ if $0 \in C$. I.e. $g^{-1}(C) \in \mathcal{F}$ for all $C \in \mathcal{C}$. Now because $\mathcal{C}$ generates the Borel σ -field of R , this gives that g is $\mathcal{F}$ -measurable as required. $\square$

§3.4 INDEPENDENCE OF THE CHOICE OF METRIC ON $\mathcal{X}$

Now we have supposed that our C-space $\mathcal{X}$ comes equipped with a metric ρ . We now consider the effect of changing this metric. The following result shows that things are not changed too much.

(3.11) Proposition. *Suppose that $\mathcal{X}$ is a C-space and that ρ and ρ' are two translation invariant metrics for $\mathcal{X}$. Suppose that $(\Gamma, \mathcal{E}, J)$ is an $\mathcal{X}$ -valued integral space and Φ_J and Φ'_J denote the upper gauges defined using ρ and ρ' respectively. Then for all $\varepsilon > 0$ there exists $\delta > 0$ such that*

$$\Phi_J(f) < \delta \implies \Phi'_J(f) < \varepsilon,$$

and vice versa (i.e. the result of interchanging the roles of Φ_J and Φ'_J). Thus Φ_J and Φ'_J give the same topology on $\mathbb{R}^\Gamma$, and so $\mathcal{F}(\Phi_J) = \mathcal{F}(\Phi'_J)$ and $\mathcal{L}^1(\Phi_J) = \mathcal{L}^1(\Phi'_J)$.

This will follow from the following result by putting $\mathcal{Y} = \mathcal{X}$, T to be the identity and taking $\rho_{\mathcal{X}}$ and $\rho_{\mathcal{Y}}$ to be ρ and ρ' respectively.

(3.12)Proposition. Suppose that $\mathcal{X}$ and $\mathcal{Y}$ are ^{real} C -spaces with translation invariant metrics $\rho_{\mathcal{X}}$ and $\rho_{\mathcal{Y}}$ respectively, and that $T: \mathcal{X} \rightarrow \mathcal{Y}$ is linear and continuous. Suppose that $(\Gamma, \mathcal{E}, J)$ is an $\mathcal{X}$ -valued integral space. Then TJ is a $\mathcal{Y}$ -valued preintegral on $(\Gamma, \mathcal{E})$, and for each $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$\Phi_J(f) < \delta \implies \Phi_{TJ}(f) < \varepsilon.$$

Thus $\mathcal{F}(\Phi_J) \subseteq \mathcal{F}(\Phi_{TJ})$, $\mathcal{L}^1(J) \subseteq \mathcal{L}^1(TJ)$, and $T \int f dJ = \int f d(TJ)$ for all $f \in \mathcal{L}^1(J)$.

Proof: Using the continuity of T it is easy to show that TJ satisfies (1) and (2) of Definition(3.3) and is hence a $\mathcal{Y}$ -valued preintegral. By the continuity of T we can take $\delta_0 > \delta_1 > \dots > 0$ such that $\delta_n \downarrow 0$ and

$$(1) \quad \rho_{\mathcal{X}}(x) < \delta_n \implies \rho_{\mathcal{Y}}(Tx) < 2^{-n}.$$

Now define $\theta: \overline{\mathbb{R}}^+ \rightarrow \overline{\mathbb{R}}^+$ by

$$\theta(\lambda) = \begin{cases} 0, & \text{if } \lambda = 0, \\ 2^{-n}, & \text{if } \delta_{n+1} \leq \lambda < \delta_n, \\ \infty, & \text{if } \lambda \geq \delta_0. \end{cases}$$

Then $\rho_{\mathcal{Y}}(Tx) \leq \theta(\rho_{\mathcal{X}}(x))$ for all $x \in \mathcal{X}$, and θ is increasing and right continuous. Now I claim that $\Phi_{TJ}(f) \leq \theta(\Phi_J(f))$ for all $f: \Gamma \rightarrow \overline{\mathbb{R}}^+$. Consider first $u \in \mathcal{E}_{\text{inc}}^+$. We have

that

$$\begin{aligned}
\Phi_{TJ}(u) &= \sup\{\rho_Y(TJ(s)) \mid s \in \mathcal{E}, \quad |s| \leq u\} \\
&\leq \sup\{\theta(\rho_X(J(s))) \mid s \in \mathcal{E}, \quad |s| \leq u\} \\
&\leq \theta(\Phi_J(u)) \quad (\text{since } \theta \text{ is increasing}).
\end{aligned}$$

For a general function $f: \Gamma \rightarrow \bar{\mathbb{R}}^+$, take $u_1, u_2, \dots \in \mathcal{E}_{\text{inc}}^+$ such that $u_n \geq f$ and $\Phi_J(u_n) \downarrow \Phi_J(f)$. Then $\Phi_{TJ}(f) \leq \Phi_{TJ}(u_n) \leq \theta(\Phi_J(u_n))$ by above. Now let $n \rightarrow \infty$ and use the right continuity of θ to prove that $\Phi_{TJ}(f) \leq \theta(\Phi_J(f))$. Thus

$$\Phi_J(f) < \delta_n \implies \Phi_{TJ}(f) \leq 2^{-n},$$

which gives the result as required. □

§3.5 EXAMPLES AND ALTERNATIVE CHARACTERIZATIONS OF $\mathcal{X}$ -VALUED INTEGRAL SPACES

We have the following characterization of $\mathcal{X}$ -valued preintegrals.

(3.13) Proposition. *Suppose that Γ is any set, $\mathcal{E}$ is any Stone lattice on Γ , $\mathcal{X}$ is a real C -space and that $J: \mathcal{E} \rightarrow \mathcal{X}$ is linear. Then J is a $\mathcal{X}$ -valued preintegral on $(\Gamma, \mathcal{E})$ if and only if*

$$(\dagger) \quad s_1, s_2, \dots \in \mathcal{E}, \quad s_n \rightarrow 0 \text{ pointwise and } |s_n| \leq t \in \mathcal{E}^+ \text{ for all } n \implies J(s_n) \rightarrow 0.$$

Proof: The ‘only if’ part follows from the Dominated Convergence Theorem (3.9). For the ‘if’ part suppose $(\dagger)$ is satisfied. We have to show that (1) and (2) of Definition(3.3) hold. (2) is easy; we consider (1). Thus suppose that $s \in \mathcal{E}^+$ and let $\mathcal{A} = \{J(u) \mid u \in \mathcal{E}, \quad |u| \leq s\}$. We have to show that $\sup_{x \in \mathcal{A}} \rho(\lambda x) \rightarrow 0$ as $\lambda \downarrow 0$; so suppose not. Now $\sup_{x \in \mathcal{A}} \rho(\lambda x) = \sup\{\rho(J(u)) \mid u \in \mathcal{E}, \quad |u| \leq \lambda s\}$, which decreases as $\lambda > 0$ decreases. But since this does not tend to zero as $\lambda \downarrow 0$, it must be

greater than ε for all $\lambda > 0$ for some $\varepsilon > 0$. Taking $\lambda = \frac{1}{n}$ we have that there exists $u_n \in \mathcal{E}$ such that $|u_n| \leq \frac{1}{n}s$ and $\rho(J(u_n)) > \varepsilon$. But by (†) (with s as t) we have that $J(u_n) \rightarrow 0$ which is a contradiction. $\square$

The following is analogous to the Riesz representation theorem.

(3.14)Theorem. *Suppose that Γ is a locally compact Hausdorff space, $\mathcal{X}$ is a real C -space, and that $J: C_0(\Gamma) \rightarrow \mathcal{X}$ is linear and continuous. Then $(\Gamma, C_0(\Gamma), J)$ is an $\mathcal{X}$ -valued integral space.*

Note. Here $C_0(\Gamma)$ denotes the Banach space of all continuous functions, $f: \Gamma \rightarrow \mathbb{R}$, that ‘vanish at infinity’ in the sense that for all $\varepsilon > 0$ there exists a compact subset K of Γ such that $|f| < \varepsilon$ outside K . $C_0(\Gamma)$ has the supremum norm.

Proof: Let $\mathcal{E} = C_0(\Gamma)$; then $\mathcal{E}$ is a Stone lattice on Γ . We have only to check (2) of Definition(3.3); so suppose that $s_n \downarrow 0$ pointwise in $\mathcal{E}$. Given $\varepsilon > 0$ there exists a compact $K \subseteq \Gamma$ such that $s_1 < \varepsilon$ outside K . Now $[s_n < \varepsilon]$ is open and $[s_n < \varepsilon] \uparrow \Gamma$. So by the compactness of K there exists an n_0 such that $[s_{n_0} < \varepsilon] \supseteq K$. Thus $0 \leq s_n < \varepsilon$ on Γ for all $n \geq n_0$. Hence $s_n \rightarrow 0$ uniformly, and therefore by continuity of J , $J(s_n) \rightarrow 0$. $\square$

(3.15)Proposition. *Suppose that Γ is any set and that $\mathcal{E}$ is a Stone lattice on Γ consisting of bounded functions and such that the constant function 1 is an element of $\mathcal{E}$. Suppose further that $\mathcal{X}$ is a real C -space, and that $J: \mathcal{E} \rightarrow \mathcal{X}$ is linear, and continuous when $\mathcal{E}$ has the supremum norm. Then*

$$s_1, s_2, \dots \in \mathcal{E}^+, s_n \downarrow 0 \implies \lim_{n \rightarrow \infty} J(s_n) \text{ exists in } \mathcal{X}.$$

If the limit must always be zero then we see that $(\Gamma, \mathcal{E}, J)$ is an $\mathcal{X}$ -valued integral space.

Note. The continuity of J is equivalent to (1) of Definition(3.3).

Proof: We construct a compact Hausdorff space in a way similar to that used to construct the Stone-Čech compactification. Define $j: \Gamma \rightarrow \mathbb{R}^{\mathcal{E}}$ by $j(x)_s = s(x)$. Give $\mathbb{R}^{\mathcal{E}}$ the product topology and let $K = \overline{j(\Gamma)}$. K is a closed subset of $\prod_{s \in \mathcal{E}} [-\|s\|, \|s\|]$ (where $\|s\| = \sup_{\Gamma} |s|$), and thus we see that K is compact Hausdorff. Also we have that j maps Γ into K with dense range. I claim that the following are true.

- (i) For each $s \in \mathcal{E}$, there exists a unique $\hat{s} \in C(K)$ such that $s = \hat{s} \circ j$.
- (ii) The function $s \mapsto \hat{s}: \mathcal{E} \rightarrow C(K)$ is linear.
- (iii) If $s \leq t$ in $\mathcal{E}$ then $\hat{s} \leq \hat{t}$.
- (iv) $\|\hat{s}\| = \|s\|$ for all $s \in \mathcal{E}$.

To prove (i), define $\hat{s}: K \rightarrow \mathbb{R}$ by $\hat{s} = \pi_s|_K$ where π_s is the s^{th} projection from $\mathbb{R}^{\mathcal{E}}$ onto $\mathbb{R}$. (ii), (iii) and (iv) now follow from (i).

Let $\hat{\mathcal{E}} = \{\hat{s} | s \in \mathcal{E}\}$. Then $\hat{\mathcal{E}}$ is a vector sublattice of $C(K)$ containing the constant functions. Also it follows from the construction of $\hat{s}$ as $\pi_s|_K$, that $\hat{\mathcal{E}}$ separates the points of K . By a well-known theorem, such vector sublattices of $C(K)$ (i.e. those separating points and containing the constant functions) are dense in $C(K)$. (Show that their closures are subalgebras and apply the Stone-Weierstrass approximation theorem or otherwise.)

Now since $\hat{\mathcal{E}}$ is dense in $C(K)$ the mapping $\hat{s} \mapsto J(s): \hat{\mathcal{E}} \rightarrow \mathcal{X}$ (s is determined by $\hat{s}$ by (i)) extends (using (iv)) to a unique continuous linear map $\hat{J}: C(K) \rightarrow \mathcal{X}$. By Theorem(3.14), $(K, C(K), \hat{J})$ is an $\mathcal{X}$ -valued integral space. Now suppose $s_n \downarrow 0$ pointwise in $\mathcal{E}^+$. Then $\hat{s}_n$ decreases as n increases. Let $u = \lim_{n \rightarrow \infty} \hat{s}_n$. Then by the Dominated Convergence Theorem (3.9), $J(s_n) = \hat{J}(\hat{s}_n) = \int \hat{s}_n d\hat{J} \rightarrow \int u d\hat{J}$, and thus $\lim_{n \rightarrow \infty} J(s_n)$ exists. □

§3.6 THE NECESSITY OF THE C-SPACE CONDITION

We shall now show that our assumption that $\mathcal{X}$ is a C-space is necessary.

(3.16)Theorem. Suppose that $\mathcal{X}$ is a real F -space such that whenever $(\Gamma, \mathcal{E}, J)$ is an $\mathcal{X}$ -valued integral space (ignoring the requirement that $\mathcal{X}$ be a C -space), the function $\Phi_J: (\bar{\mathbb{R}}^+)^{\Gamma} \rightarrow \bar{\mathbb{R}}^+$ defined by Definition(3.4) is an upper gauge on $(\Gamma, \mathcal{E})$. Then $\mathcal{X}$ is a C -space.

Proof: Suppose that $x_1, x_2, \dots \in \mathcal{X}$ are such that

$$\mathcal{A} = \{ \sum_{n=1}^N \alpha_n x_n \mid N \in \mathbb{N}_{>0}, \quad \alpha_1, \dots, \alpha_N \in [-1, 1] \},$$

is bounded in $\mathcal{X}$. Take $\Gamma = \mathbb{N}_{>0}$ and $\mathcal{E}$ to be the set of all real-valued functions on Γ with finite support. Define $J: \mathcal{E} \rightarrow \mathcal{X}$ by $J(s) = \sum_{n=1}^{\infty} s(n)x_n$. Now it is easy to show that $(\Gamma, \mathcal{E}, J)$ is an $\mathcal{X}$ -valued integral space.

Let $s_n = \mathbf{1}_{\{1, \dots, n\}}$ for $n = 1, 2, \dots$. Then $s_n \uparrow 1$, and so $1 \in \mathcal{E}_{\text{inc}}^+$. For $\lambda > 0$, $\sup_n \Phi_J(\lambda s_n) \leq \Phi_J(\lambda 1) = \sup_{x \in \mathcal{A}} \rho(\lambda x)$, and so $\sup_n \Phi_J(\lambda s_n) \rightarrow 0$ as $\lambda \downarrow 0$ by the boundedness of $\mathcal{A}$. Thus since Φ_J satisfies (4) of Definition(2.1)(b) we see that $\Phi_J(s_n - s_m) \rightarrow 0$ as $n, m \rightarrow \infty$. So $\rho(x_n) = \rho(J(s_n - s_{n-1})) \leq \Phi_J(s_n - s_{n-1}) \rightarrow 0$ as $n \rightarrow \infty$, i.e. $x_n \rightarrow 0$ as required. $\square$

§3.7 EXAMPLES OF C -SPACES: L^p SPACES

Throughout this section we shall assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space. It remains to give the promised proof that $L^p(\Omega, \mathcal{F}, \mathbb{P})$ is a C -space for $0 \leq p < \infty$. This is required in order to apply the results of this chapter to the theory of stochastic integration in the next chapter. The proof we shall give is based on one given in Rolewicz [Ro] (see Proposition III.8.6). To give that proof we shall need to use the Rademacher functions on $[0, 1]$.

(3.17)Definition. The *Rademacher functions* are the functions $r_1, r_2, \dots: [0, 1] \rightarrow \mathbb{R}$ defined by $r_n(t) = \text{sgn}[\sin(2^n \pi t)]$. $\square$

Thinking about this definition we see that the Rademacher functions are ‘square waves’. It is best to think of $r_1, r_2, \dots$ as independent identically distributed random variables on $[0, 1]$ (with Lebesgue measure as the probability measure), the distribution being that taking ± 1 with equal probability $1/2$. Thinking in this way we see that $(\prod_{n \in S} r_n)_S$, where S runs over all finite subsets of $\mathbb{N}_{>0}$, is an orthonormal set in $L^2[0, 1]$. In fact it is complete, although we shall not need this fact. We give now one half of an important result for Rademacher functions.

(3.18) Theorem (Khintchine’s Inequality).

Let R be the real vector space spanned by the Rademacher functions. Then if $0 \leq p \leq \infty$ and R is given the topology inherited from $L^p[0, 1]$, the inclusion map $R \hookrightarrow L^2[0, 1]$ is continuous. For $0 < p < \infty$ this means that there is a constant $k_p \in (0, \infty)$ such that for all N and all $a_1, \dots, a_N \in \mathbb{R}$ we have

$$\left(\sum_{n=1}^N a_n^2 \right)^{1/2} \leq k_p \left(\int_0^1 \left| \sum_{n=1}^N a_n r_n \right|^p \right)^{1/p}.$$

Proof: Since the inclusion $L^p \hookrightarrow L^0$ is continuous, we may assume without loss of generality that $p = 0$. Suppose that $f = \sum_{n=1}^N a_n r_n \in R$. For $\lambda > 0$, let $A = \{|f| > \lambda\}$, and $\eta = \mu A$, where μ denotes Lebesgue measure. Let I denote $[0, 1]$. We have that

$$\begin{aligned} \lambda^2 &\geq \int_{I \setminus A} \left(\sum_{n=1}^N a_n r_n \right)^2 \\ &= \int_{I \setminus A} \left(\sum_{n=1}^N a_n^2 r_n^2 + 2 \sum_{k < l \leq N} a_k a_l r_k r_l \right) \\ &= (1 - \eta) \sum_{n=1}^N a_n^2 - 2 \sum_{k < l \leq N} a_k a_l \int_A r_k r_l \end{aligned}$$

(since $\int_{I \setminus A} r_k r_l + \int_A r_k r_l = \int_I r_k r_l = 0$ if $k \neq l$ by the orthonormality of the r_k ’s)

$$\geq (1 - \eta) \sum_{n=1}^N a_n^2 - 2 \left(\sum_{k < l \leq N} a_k^2 a_l^2 \right)^{1/2} \left(\sum_{k < l \leq N} \left(\int_A r_k r_l \right)^2 \right)^{1/2}$$

(by the Cauchy-Schwarz inequality).

Now by Bessel's inequality and the fact that $(r_k r_l)_{k < l}$ is orthonormal, we see that

$\sum_{k < l \leq N} \left(\int_A r_k r_l \right)^2 = \sum_{k < l \leq N} \langle \mathbf{1}_A, r_k r_l \rangle^2 \leq \|\mathbf{1}_A\|_2^2 = \eta$. But also we have that $\sum_{k < l \leq N} a_k^2 a_l^2 = \frac{1}{2} \sum_{k \neq l} a_k^2 a_l^2 \leq \frac{1}{2} \left(\sum_{n=1}^N a_n^2 \right)^2$. So we obtain that

$$\lambda^2 \geq (1 - \eta - \sqrt{2\eta}) \sum_{n=1}^N a_n^2 = (1 - \eta - \sqrt{2\eta}) \|f\|_2^2.$$

Now given $\varepsilon > 0$, take $0 < \lambda < \varepsilon$. We have shown that

$$f \in R \implies \|f\|_2 \leq \lambda(1 - \eta - \sqrt{2\eta})^{-1/2} \quad \text{where } \eta = \mu[|f| > \lambda].$$

But as $f \rightarrow 0$ in L^0 , $\eta \rightarrow 0$. Therefore there exists a $\delta > 0$ such that $\|f\|_2 < \varepsilon$ for all $f \in R$ such that $\rho_0(f) < \delta$. This gives the required continuity. $\square$

(3.19) Theorem. For $0 \leq p < \infty$, $L^p(\Omega, \mathcal{F}, \mathbb{P})$ is a C-space. In fact if $X_1, X_2, \dots \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ are such that the set $\mathcal{A} = \{ \sum_{n=1}^N a_n X_n \mid N \in \mathbb{N}_{>0}, a_1, \dots, a_N \in [-1, 1] \}$ is bounded in L^p . Then the following hold.

- (i) $\sum_{n=1}^{\infty} |X_n|^2 < \infty$ a.s.
- (ii) $(\sum_{n=1}^{\infty} |X_n|^2)^{1/2} \in L^p$.
- (iii) $X_n \rightarrow 0$ a.s. and in L^p .

Note. Here we are assuming that $p < \infty$. This is necessary (see next result).

Proof: Consider first the case when $p > 0$. There is a constant $C \in (0, \infty)$ such that $\int_{\Omega} |Y|^p d\mathbb{P} \leq C$ for all $Y \in \mathcal{A}$. Now by Theorem(3.18), we have for $\omega \in \Omega$, that

$$\left(\sum_{n=1}^N X_n(\omega)^2 \right)^{1/2} \leq k_p \left(\int_0^1 \left| \sum_{n=1}^N r_n(t) X_n(\omega) \right|^p dt \right)^{1/p}.$$

Raise both sides to the power p , integrate with respect to ω , and interchange the order of integration on the right hand side to give that

$$\begin{aligned} \int_{\Omega} \left(\sum_{n=1}^N X_n^2 \right)^{p/2} d\mathbb{P} &\leq k_p^p \int_0^1 \int_{\Omega} \left| \sum_{n=1}^N r_n(t) X_n(\omega) \right|^p \mathbb{P}(d\omega) dt \\ &\leq k_p^p \int_0^1 C dt = C k_p^p. \end{aligned}$$

Now let $N \rightarrow \infty$ and use the monotone convergence theorem to obtain that

$$\int_{\Omega} \left(\sum_{n=1}^{\infty} X_n^2 \right)^{p/2} d\mathbb{P} \leq C k_p^p < \infty.$$

So parts (i) and (ii) now follow. For part (iii), let $Y = (\sum_{n=1}^{\infty} X_n^2)^{1/2}$. Then $|X_n| \leq Y \in L^p$, and so $|X_n|^p \leq Y^p \in L^1$. Now $X_n \rightarrow 0$ a.s. Therefore the dominated convergence theorem gives that $\int_{\Omega} |X_n|^p d\mathbb{P} \rightarrow 0$ as required.

Now we consider the case when $p = 0$. Letting $E = [\sum_{n=1}^{\infty} X_n^2 = \infty]$ we must show that $\mathbb{P}(E) = 0$. Suppose not and take δ such that $0 < \delta < \mathbb{P}(E)$. We shall use ρ_0 to denote the metric on $L^0(\Omega, \mathcal{F}, \mathbb{P})$ as usual, but use $\tilde{\rho}_0$ for the corresponding metric on $L^0[0, 1]$. Let R denote the real linear span of the Rademacher functions as in Theorem(3.18). By Theorem(3.18), there is an $\eta > 0$ such that

$$(\dagger) \quad f \in R, \quad \tilde{\rho}_0(f) \leq \eta \implies \|f\|_2 \leq 1.$$

For $\omega \in \Omega$, let $f_{N,\omega} = \sum_{n=1}^N X_n(\omega) r_n$ so $f_{N,\omega} \in R$. Now since $\mathcal{A}$ is bounded in L^0 , there is a $\lambda > 0$ such that

$$(\ddagger) \quad \rho_0\left(\frac{1}{\lambda} Y\right) \leq \eta \delta \quad \text{for all } Y \in \mathcal{A}.$$

Now as $N \rightarrow \infty$, $[\sum_{n=1}^N X_n^2 > \lambda^2] \uparrow [\sum_{n=1}^{\infty} X_n^2 > \lambda^2] \supseteq E$. So there is an N such that $\mathbb{P}[\sum_{n=1}^N X_n^2 > \lambda^2] > \delta$. Let $A = [\sum_{n=1}^N X_n^2 > \lambda^2]$. Then for $\omega \in A$, $\|\frac{1}{\lambda} f_{N,\omega}\|_2 > 1$. So from $(\dagger)$ we see that $\tilde{\rho}_0(\frac{1}{\lambda} f_{N,\omega}) > \eta$. We have thus proved that

$$\mathbf{1}_A(\omega) \leq \frac{1}{\eta} \tilde{\rho}_0\left(\frac{1}{\lambda} f_{N,\omega}\right) \quad \text{for all } \omega \in \Omega.$$

Integrate with respect to ω to give that

$$\begin{aligned}
\mathbb{P}(A) &\leq \frac{1}{\eta} \int_{\Omega} \tilde{\rho}_0\left(\frac{1}{\lambda} f_{N,\omega}\right) \mathbb{P}(d\omega) \\
&= \frac{1}{\eta} \int_0^1 \rho_0\left(\frac{1}{\lambda} \sum_{n=1}^N r_n(t) X_n\right) dt \\
&\quad \text{(by interchanging the order of integration)} \\
&\leq \frac{1}{\eta} \int_0^1 \eta \delta dt \\
&\quad \text{(by (\dagger) since } \sum_{n=1}^N r_n(t) X_n \in \mathcal{A} \text{ for each } t\text{)} \\
&= \delta.
\end{aligned}$$

So we have shown that $\mathbb{P}(A) \leq \delta$, which is a contradiction. Thus $\mathbb{P}(E) = 0$ and we have proved part (i). Parts (ii) and (iii) are now easy in this case. $\square$

Concerning the case $p = \infty$ we have the following result.

(3.20) Proposition. *$L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ is a C -space if and only if it is finite dimensional.*

Note. It follows from this proposition and Theorem(3.16) that L^∞ can not be used as a range space in this theory of vector-valued integrals. We shall not therefore be able to define L^∞ -integrators.

Proof: We say that $A \in \mathcal{F}$ is $(\mathbb{P}-)$ atomic if and only if $\mathbb{P}(A) > 0$ and there does not exist $B \in \mathcal{F}$ such that $B \subseteq A$ and $0 < \mathbb{P}(B) < \mathbb{P}(A)$. I claim that either

(1) $\Omega = A_1 \cup \dots \cup A_N$ for some pairwise disjoint atomic $A_1, \dots, A_N \in \mathcal{F}$;

or

(2) There exist pairwise disjoint $A_1, A_2, \dots \in \mathcal{F}$ such that $\mathbb{P}(A_n) > 0$.

For suppose (1) does not hold. Then Ω is not atomic. Thus we can write $\Omega = A_1 \cup B_1$, with $A_1, B_1 \in \mathcal{F}$ disjoint, and $\mathbb{P}(A_1), \mathbb{P}(B_1) > 0$. We can not have that both A_1 and B_1 are finite disjoint unions of atomic sets; therefore without loss of generality B_1 is

not such a union. Repeat with B_1 as Ω to give $B_1 = A_2 \cup B_2$ with $A_2, B_2 \in \mathcal{F}$ disjoint, $\mathbb{P}(A_2), \mathbb{P}(B_2) > 0$, and B_2 not a finite disjoint union of atomic sets. Continuing by induction gives $A_1, A_2, \dots$ as in (2) and thus proves (2).

If (1) holds then let $\mathcal{G}$ denote the set of unions from $\{A_1, \dots, A_N\}$. So $\mathcal{G}$ is the σ -field generated by $\{A_1, \dots, A_N\}$. We have that for all $F \in \mathcal{F}$ there exists an $A \in \mathcal{G}$ such that $\mathbb{P}(F \triangle A) = 0$. Thus L^∞ is spanned by the subset $\{\mathbf{1}_{A_1}, \dots, \mathbf{1}_{A_N}\}$. I.e. L^∞ is finite dimensional.

If (2) holds then put $X_n = \mathbf{1}_{A_n}$. So $\{\sum_{n=1}^N a_n X_n \mid N \in \mathbb{N}_{>0}, a_1, \dots, a_N \in [-1, 1]\}$ is bounded in L^∞ . But $\|X_n\|_\infty = 1$ for all n , and so $X_n \not\rightarrow 0$ in L^∞ . Thus L^∞ is not a C-space.

Thus if L^∞ is a C-space, it must be finite dimensional. The converse is easy, since any finite-dimensional real vector space is a C-space. $\square$

CHAPTER 4: THE STOCHASTIC INTEGRAL

§4.1 THE SETUP AND DEFINITIONS

(4.1) Definition. Suppose that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space.

- (a) A *filtration* on $(\Omega, \mathcal{F}, \mathbb{P})$ is a family $(\mathcal{F}_t)_{t \in \mathbb{R}^+}$ of σ -fields such that $\mathcal{F}_t \subseteq \mathcal{F}$ for all $t \in \mathbb{R}^+$ and $\mathcal{F}_s \subseteq \mathcal{F}_t$ for all $s \leq t$ in $\mathbb{R}^+$.
- (b) A filtration $(\mathcal{F}_t)_{t \in \mathbb{R}^+}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ is said to be *right continuous* if and only if $\mathcal{F}_t = \bigcap_{s > t} \mathcal{F}_s$ for all $t \in \mathbb{R}^+$.
- (c) A *standard filtration* on $(\Omega, \mathcal{F}, \mathbb{P})$ is a filtration $(\mathcal{F}_t)_{t \in \mathbb{R}^+}$ that is right continuous and such that each $\mathcal{F}_t$ contains all the $\mathbb{P}$ -null sets in $\mathcal{F}$.
- (d) A *standard setup* is a quadruple $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t \in \mathbb{R}^+})$ such that $(\Omega, \mathcal{F}, \mathbb{P})$ is a complete probability space, and $(\mathcal{F}_t)_{t \in \mathbb{R}^+}$ is a standard filtration on $(\Omega, \mathcal{F}, \mathbb{P})$ □

Throughout this chapter, and most of the rest of this thesis, we shall assume that $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t \in \mathbb{R}^+})$ is a standard setup. The intuitive idea behind this definition is that $t \in \mathbb{R}^+$ represents time, and that $\mathcal{F}_t$ is our knowledge at time t . This increases as time t increases. The condition that the filtration is right continuous is just a convenient condition to make the mathematics more pleasant. We shall use $\mathcal{L}^p$ and L^p to denote $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ and $L^p(\Omega, \mathcal{F}, \mathbb{P})$ respectively.

Before we continue we should comment on a convention that we shall use. Given a function $X: \mathbb{R}^+ \times \Omega \rightarrow A$ for some set A , we will consider X also as a function $X: \mathbb{R}^+ \rightarrow A^\Omega$ by putting $X(t)(\omega) = X(t, \omega)$. It is convenient also to write X_t for $X(t)$ (this is standard practice in the theory of random processes). Thus X can be regarded three things: a family $(X_t)_{t \in \mathbb{R}^+}$ of functions from Ω to A , a function $X: \mathbb{R}^+ \rightarrow A^\Omega$ and a function $X: \mathbb{R}^+ \times \Omega \rightarrow A$. We shall use a similar convention for functions defined on the set $\mathbb{R}^{++} \times \Omega$.

We can now make some more definitions.

(4.2) Definition.

- (a) A function $X: \mathbb{R}^+ \times \Omega \rightarrow \mathbb{R}$ (resp. $X: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}$) is said to be an *adapted process* or just a *process* if and only if X_t is $\mathcal{F}_t$ -measurable for all $t \in \mathbb{R}^+$ (resp. $t \in \mathbb{R}^{++}$). We shall assume throughout that all processes are adapted. An L^p -process is a process X such that $X_t \in L^p$ for all t .
- (b) If $X: \mathbb{R}^+ \times \Omega \rightarrow A$ or $X: \mathbb{R}^{++} \times \Omega \rightarrow A$ for some set A , then the functions $(t \mapsto X_t(\omega))_{\omega \in \Omega}$ are known as the *paths* of X .
- (c) A *càdlàg process* (resp. *continuous process*) is a process $X: \mathbb{R}^+ \times \Omega \rightarrow \mathbb{R}$ for which all its paths are càdlàg (resp. continuous). (See (1.4) for the meaning of càdlàg.)
- (d) A *làdcàg process* is an process $X: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}$ for which all its paths are làdcàg. (See (1.4) for the meaning of làdcàg. Note in particular that this insists on the existence for each ω of $\lim_{s \downarrow t} X_s(\omega)$ for all $t \in \mathbb{R}^+$ including $t = 0$.)
- (e) For each càdlàg process X , X_- denotes the làdcàg process given by $(X_-)_t(\omega) = X_{t-}(\omega) = \lim_{s \uparrow t} X_s(\omega)$. For each làdcàg process X , X_+ denotes the càdlàg process given by $(X_+)_t(\omega) = X_{t+}(\omega) = \lim_{s \downarrow t} X_s(\omega)$. The two operations $X \mapsto X_-$ and $X \mapsto X_+$ are mutual inverses (see (1.4)).
- (f) Given two process X and Y we say that X and Y are *indistinguishable* if and only if the paths $t \mapsto X_t(\omega)$ and $t \mapsto Y_t(\omega)$ are the same for a.a. ω . A set $E \subseteq \mathbb{R}^+ \times \Omega$ is said to be *evanescent* if and only if it is a subset of a set of the form $\mathbb{R}^+ \times F$ where $F \subseteq \Omega$ is $\mathbb{P}$ -null. So we see that two processes X and Y are indistinguishable if and only if $[X \neq Y] \subseteq \mathbb{R}^+ \times \Omega$ is evanescent.
- (g) If $X: \mathbb{R}^+ \times \Omega \rightarrow \mathbb{R}$, then we define $X^*: \mathbb{R}^+ \times \Omega \rightarrow \bar{\mathbb{R}}^+$ by $X_t^* = \sup_{0 \leq s \leq t} |X_s|$, where the supremum is evaluated pointwise on Ω . Similarly if $X: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}$, we define $X^*: \mathbb{R}^{++} \times \Omega \rightarrow \bar{\mathbb{R}}^+$ by $X_t^* = \sup_{0 < s \leq t} |X_s|$. We note that some authors omit the modulus sign from the definition of X^* . □

Note.

1. As in (c) and (d) above we shall use the convention that an adjective such as ‘left continuous’, when applied to a process, refers to the paths.

2. If X is a càdlàg (resp. làdcàg) process, then X^* is real-valued and is a càdlàg (resp. làdcàg) process. (In order to see that X_t^* is $\mathcal{F}_t$ -measurable, note that the supremum can be replaced, by the right (resp. left) continuity of X , by a supremum over the countable set $([0, t] \cap \mathbb{Q}) \cup \{t\}$ (resp. $(0, t] \cap \mathbb{Q}$).)

3. Indistinguishable processes are to be thought of as the same process. We shall often write $X = Y$ when X and Y are indistinguishable processes. This should not cause any confusion. (This is part of a general philosophy of ignoring $\mathbb{P}$ -null sets.)

(4.3) Definition.

- (a) A *previsible rectangle* is a subset of $\mathbb{R}^{++} \times \Omega$ of the form $(s, t] \times F$, where $0 \leq s < t < \infty$ and $F \in \mathcal{F}_s$.
- (b) The *previsible σ -field* $\mathcal{P}$ is the σ -field on $\mathbb{R}^{++} \times \Omega$ generated by the set of all previsible rectangles. Elements of $\mathcal{P}$ will be called *previsible sets*.
- (c) A *previsible process* is a function $X: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}$ that is $\mathcal{P}$ -measurable. (Since a previsible process is adapted it is acceptable to use the term ‘process’ here.)

□

(4.4) Proposition.

- (i) Every left-continuous adapted process $X: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}$ is previsible.
- (ii) $\mathcal{P}$ is the smallest σ -field making all the làdcàg processes $X: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}$ measurable.

Proof:

- (i) The function $(t, \omega) \mapsto C(\omega)\mathbf{1}_{(a, b]}(t)$ is previsible when $0 \leq a < b < \infty$ and C is a real-valued $\mathcal{F}_a$ -measurable random variable. (Take C as a limit of simple

random variables.) For a general left-continuous adapted process $X: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}$, let $X^{(n)}: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}$ by

$$X_t^{(n)}(\omega) = \sum_{k=0}^{2^{2n}+1} X_{k2^{-n}}(\omega) \mathbf{1}_{(k2^{-n}, (k+1)2^{-n}]}(t).$$

Then $X^{(n)}$ is previsible and $X^{(n)} \rightarrow X$ pointwise. Thus X is previsible.

(ii) Let $\mathcal{G}$ denote the smallest σ -field making all l  dc  g processes $X: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}$ measurable. Now since all such processes are previsible by (i), it follows that $\mathcal{G} \subseteq \mathcal{P}$. For the reverse inclusion suppose that R is a previsible rectangle. Then $\mathbf{1}_R$ is a l  dc  g process and therefore is $\mathcal{G}$ -measurable. Thus $R \in \mathcal{G}$, and since the previsible rectangles generate $\mathcal{P}$, it follows that $\mathcal{P} \subseteq \mathcal{G}$ as required. $\square$

(4.5) Definition. We shall denote the real vector subspace of $\mathbb{R}^{\mathbb{R}^{++} \times \Omega}$ spanned by $(\mathbf{1}_R)_R$, as R runs over all the previsible rectangles, by $\mathcal{E}$. Elements of $\mathcal{E}$ will be known as *elementary previsible processes* (note that all such objects are previsible processes). Equivalently we see that $\mathcal{E}$ consists of the processes $X: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}$ of the form

$$(†) \quad X_t(\omega) = \sum_{k=0}^{r-1} C_k(\omega) \mathbf{1}_{(t_k, t_{k+1}]}(t),$$

where $r \in \mathbb{N}$, $0 = t_0 < \dots < t_r < \infty$, and for $k = 0, \dots, r-1$, C_k is a simple real-valued $\mathcal{F}_{t_k}$ -measurable random variable.

We shall again use $\mathcal{E}^+$ to denote the set of elements X of $\mathcal{E}$ such that $X \geq 0$. As in Chapter 3, $\mathcal{E}_{\text{inc}}^+$ will denote the set of $F: \mathbb{R}^{++} \times \Omega \rightarrow \overline{\mathbb{R}}^+$ such that there exists a sequence $(X_n)_n$ in $\mathcal{E}^+$ such that $X_n \uparrow F$ pointwise. All elements of $\mathcal{E}_{\text{inc}}^+$ are previsible, since they are pointwise limits of elements of $\mathcal{E}$. $\square$

§4.2 THE STOCHASTIC INTEGRAL

Our aim is to define an integral $\int X dS$ when S is a suitably nice c  dl  g process, known as an integrator, and X a suitable previsible process. To do this we need to

introduce an upper gauge on the set $\mathbb{R}^{++} \times \Omega$. It is easy to see that $\mathcal{E}$ is a Stone lattice on $\mathbb{R}^{++} \times \Omega$. In giving the definition of an integrator it is convenient to define it as an L^p -valued function. We shall see in the next chapter that this leads to a (pathwise) càdlàg process.

(4.6)Definition. Suppose that $0 \leq p < \infty$ and that $S: \mathbb{R}^+ \rightarrow L^p$ is a vector-valued function that is adapted in the sense that S_t (meaning $S(t)$) is $\mathcal{F}_t$ -measurable for all $t \in \mathbb{R}^+$. Then the linear map $\int \dots dS: \mathcal{E} \rightarrow L^p$ is defined by $\int X dS = \sum_{k=0}^{r-1} C_k(S_{t_{k+1}} - S_{t_k})$, when $X \in \mathcal{E}$ is given by (†) of Definition(4.5). It is easy to check that this is well defined.

For $X \in \mathcal{E}$, define also $\int_0^t X dS = \int \mathbf{1}_{[0,t]} X dS$ for all $t \in \mathbb{R}^+$, and define an adapted function $X \cdot S: \mathbb{R}^+ \rightarrow L^p$ by $(X \cdot S)_t = \int_0^t X dS$. $\square$

(4.7)Definition. Suppose that $0 \leq p < \infty$. An L^p -integrator is a vector-valued adapted function $S: \mathbb{R}^+ \rightarrow L^p$ that satisfies the following condition.

(†) If $X_1, X_2, \dots \in \mathcal{E}$, $X_n \rightarrow 0$ pointwise, and $|X_n| \leq Y \in \mathcal{E}^+$ for all n , then $\int X_n dS \rightarrow 0$ in L^p . $\square$

Note. We have not assumed that L^p -integrators are right continuous, since as we shall see (Corollary(4.18)) it will follow from the above definition that they are càdlàg as L^p -valued functions.

Now we see, from Proposition(3.13), that an adapted function $S: \mathbb{R}^+ \rightarrow L^p$ is an L^p -integrator if and only if $\int \dots dS$ is an L^p -valued preintegral on $(\mathbb{R}^{++} \times \Omega, \mathcal{E})$. Thus from Definition(3.3) the following result follows.

(4.8)Proposition. Suppose that $0 \leq p < \infty$. A vector-valued adapted function $S: \mathbb{R}^+ \rightarrow L^p$ is an L^p -integrator if and only if the following hold.

- (1) For each $t \in \mathbb{R}^+$, $\int_0^t \dots dS: \mathcal{E} \rightarrow L^p$ is continuous when $\mathcal{E}$ is given the supremum norm.
- (2) $X_1, X_2, \dots \in \mathcal{E}$, $X_n \downarrow 0$ pointwise $\implies \int X_n dS \rightarrow 0$ in L^p . $\square$

Another characterization of L^p -integrators will be given in §4.5.

By applying the work of Chapter 3 to the L^p -valued integral space $(\mathbb{R}^{++} \times \Omega, \mathcal{E}, \int \dots dS)$, the stochastic integral with respect to an L^p -integrator can now be constructed. We thus make the following definition (cf. Definition(3.4)).

(4.9)Definition. Suppose that $0 \leq p < \infty$ and that S is an L^p -integrator. Then define $\Phi_S^p: (\overline{\mathbb{R}}^+)^{\mathbb{R}^{++} \times \Omega} \rightarrow \overline{\mathbb{R}}^+$ by

$$\begin{aligned} \Phi_S^p(F) &= \sup\{ \rho_p(\int X dS) \mid X \in \mathcal{E}, \quad |X| \leq F \} \quad \text{for all } F \in \mathcal{E}_{\text{inc}}^+, \\ \Phi_S^p(X) &= \inf\{ \Phi_S^p(F) \mid X \leq F \in \mathcal{E}_{\text{inc}}^+ \} \quad \text{for all } X: \mathbb{R}^{++} \times \Omega \rightarrow \overline{\mathbb{R}}^+. \end{aligned}$$

$\square$

The following two results follow at once from Theorem(3.5) and Proposition(3.6).

(4.10)Theorem. Suppose that $0 \leq p < \infty$ and that S is an L^p -integrator. Then $(\mathbb{R}^{++} \times \Omega, \mathcal{E}, \Phi_S^p)$ is a σ -finite upper-gauge space. $\square$

(4.11)Proposition. Suppose that $F_1, F_2, \dots \in \mathcal{E}_{\text{inc}}^+$, $F: \mathbb{R}^{++} \times \Omega \rightarrow \overline{\mathbb{R}}^+$ and that $F_n \uparrow F$ pointwise. Then $F \in \mathcal{E}_{\text{inc}}^+$ and for each L^p -integrator S ($0 \leq p < \infty$), $\Phi_S^p(F_n) \uparrow \Phi_S^p(F)$. $\square$

The following follows from Proposition(3.10) (observing that $\mathcal{P}$ is the smallest σ -field making all the elements of $\mathcal{E}$ measurable).

(4.12) Proposition. Suppose that $0 \leq p < \infty$ and that S is an L^p -integrator. Then the following are true.

- (i) If $E \subseteq \mathbb{R}^{++} \times \Omega$ is Φ_S^p -null, then there exist an $A \in \mathcal{P}$ such that $E \subseteq A$ and A is Φ_S^p -null.
- (ii) $\mathcal{M}(\Phi_S^p)$ is the set of subsets E of $\mathbb{R}^{++} \times \Omega$ such that there exist $A, B \in \mathcal{P}$ such that $A \subseteq E \subseteq B$ and $B \setminus A$ is Φ_S^p -null.
- (iii) If $X: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}(\text{resp. } \overline{\mathbb{R}})$ is Φ_S^p -measurable, then there exists a previsible $Y: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}(\text{resp. } \overline{\mathbb{R}})$ such that $X = Y$ Φ_S^p -a.e. □

Suppose we have an L^p -integrator S . Then the obvious definition for the set of processes integrable with respect to S is $\mathcal{L}^1(\Phi_S^p)$. However it is more convenient to restrict to the previsible processes in this set. This causes no real loss of generality because of Proposition(4.12)(iii) above. Thus restating Definition(3.7) in this special case; we give the following definition of the stochastic integral.

(4.13) Definition. Suppose that $0 \leq p < \infty$ and that S is an L^p -integrator.

- (a) $\mathcal{L}^1(S, L^p)$ will denote the vector space of all previsible elements of $\mathcal{L}^1(\Phi_S^p)$. The (L^p) -stochastic integral is the unique linear map $\int \dots dS: \mathcal{L}^1(S, L^p) \rightarrow L^p$ extending the linear map $\int \dots dS: \mathcal{E} \rightarrow L^p$ given in Definition(4.6) and satisfying,

$$(\dagger) \quad \rho_p \left(\int X dS \right) \leq \Phi_S^p(X) \quad \text{for all } X \in \mathcal{L}^1(S, L^p).$$

We note here that by Proposition(2.23) that $\mathcal{L}^1(S, L^p)$ consists of the previsible elements of $\mathcal{F}(\Phi_S^p)$.

- (b) $\Lambda^1(S, L^p)$ will denote the vector space of all previsible processes $X: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}$ such that $\mathbf{1}_{(0, t] \times \Omega} X \in \mathcal{L}^1(S, L^p)$ for all $t \in \mathbb{R}^+$. For each $X \in \Lambda^1(S, L^p)$, define $\int_0^t X dS = \int \mathbf{1}_{(0, t] \times \Omega} X dS$ for all $t \in \mathbb{R}^+$. Also define $X \bullet S: \mathbb{R}^+ \rightarrow L^p$ by $(X \bullet S)_t = \int_0^t X dS$. □

It is convenient just to note that a consequence of (†) of Definition(4.13)(a) is that $\int \dots dS: \mathcal{L}^1(S, L^p) \rightarrow L^p$ is continuous (with the topology on $\mathcal{L}^1(S, L^p)$ given by Φ_S^p). This is just what we used in its definition in terms of a limit.

We shall be mainly interested in the case when $p = 0$, since this gives the biggest class of integrators and classes of integrable processes as we shall now see.

(4.14)Proposition. *Suppose that $0 \leq q \leq p < \infty$ and that S is an L^p -integrator. Then S is also an L^q -integrator, $\mathcal{L}^1(S, L^p) \subseteq \mathcal{L}^1(S, L^q)$ and the two definitions of $\int \dots dS$ agree on $\mathcal{L}^1(S, L^p)$. Furthermore we have that for each $\varepsilon > 0$ there exists a $\delta > 0$ such that $\Phi_S^q(X) < \varepsilon$ whenever $\Phi_S^p(X) < \delta$.*

Proof: Apply Proposition(3.12) with $(\mathbb{R}^{++} \times \Omega, \mathcal{E}, \int \dots dS)$ as $(\Gamma, \mathcal{E}, J)$ and the inclusion $L^p \hookrightarrow L^q$ as $T: \mathcal{X} \rightarrow \mathcal{Y}$. □

Note.

1. From their definitions in terms of $\mathcal{L}^1(S, L^p)$, $\mathcal{L}^1(S, L^q)$ and $\int \dots dS$, we see that $\Lambda^1(S, L^p) \subseteq \Lambda^1(S, L^q)$ and the two definitions of $X \cdot S$ agree for all $X \in \Lambda^1(S, L^p)$.
2. By taking θ in the proof of Proposition(3.12) to be identity when $q = 0$, and $\lambda \mapsto \lambda^{(1 \wedge q)/(1 \wedge p)}$ when $q > 0$, we can see that actually $(\Phi_S^q)^{1/(1 \wedge q)} \leq (\Phi_S^p)^{1/(1 \wedge p)}$ if $q > 0$, and $\Phi_S^0 \leq \Phi_S^p$ if $q = 0$.

(4.15)Notation. From Proposition(4.14) above we can see that the case $p = 0$ gives the most general definition of stochastic integration. In view of this we shall use the term ‘integrator’ to mean L^0 -integrator, and when S is an integrator, we shall denote $\mathcal{L}^1(S, L^0)$ and $\Lambda^1(S, L^0)$ by $\mathcal{L}^1(S)$ and $\Lambda^1(S)$ respectively. □

§4.3 THE CONVERGENCE THEOREMS AND OTHER PROPERTIES

Using the Monotone Convergence Theorem (3.8) and Dominated Convergence

Theorem (3.9) of Chapter 3 we get immediately the following versions of the convergence theorems.

(4.16) The Monotone Convergence Theorem.

Suppose that $0 \leq p < \infty$, S is an L^p -integrator, $X_1, X_2, \dots: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}^+$ are previsible and satisfy $0 \leq X_1 \leq X_2 \leq \dots$, and $\sup_n \Phi_S^p(\lambda X_n) \rightarrow 0$ as $\lambda \downarrow 0$. Then $X_1, X_2, \dots \in \mathcal{L}^1(S, L^p)$ and there exists $X \in \mathcal{L}^1(S, L^p)$ such that $X = \sup_n X_n$ Φ_S^p -a.e., $\Phi_S^p(X_n - X) \rightarrow 0$, and $\int X_n dS \rightarrow \int X dS$ in L^p . $\square$

(4.17) The Dominated Convergence Theorem.

Suppose that $0 \leq p < \infty$, S is an L^p -integrator, $X_1, X_2, \dots \in \mathcal{L}^1(S, L^p)$, $X: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}$ such that $X_n \rightarrow X$ pointwise, and for some non-negative $Y \in \mathcal{F}(\Phi_S^p)$, $|X_n| \leq Y$ for all n . Then $X \in \mathcal{L}^1(S, L^p)$, $\Phi_S^p(X_n - X) \rightarrow 0$, and $\int X_n dS \rightarrow \int X dS$ in L^p . $\square$

(4.18) Corollary. If $0 \leq p < \infty$ and S is a L^p -integrator, then as a vector-valued function, $S: \mathbb{R}^+ \rightarrow L^p$, it is càdlàg.

Proof: For each $t \in \mathbb{R}^+$, $S_t = S_0 + \int \mathbf{1}_{(0,t]} \times \Omega dS$. So by the Dominated Convergence Theorem (4.17) (with $Y = \mathbf{1}_{(0,b]} \times \Omega$ for some $b \in \mathbb{R}^+$) we have for $t_0 \in \mathbb{R}^+$ that $S_t \rightarrow S_{t_0}$ in L^p as $t \downarrow t_0$ (use a sequence of values for t). Similarly if $0 < t_0 < \infty$, then $S_t \rightarrow S_0 + \int \mathbf{1}_{(0,t_0)} \times \Omega dS$ in L^p as $t \uparrow t_0$. $\square$

(4.19) Proposition (Linearity in the Integrator).

Suppose that $0 \leq p < \infty$, R and S are L^p -integrators and that $\alpha, \beta \in \mathbb{R}$. Then $\alpha R + \beta S$ is also an L^p -integrator, $\mathcal{L}^1(R, L^p) \cap \mathcal{L}^1(S, L^p) \subseteq \mathcal{L}^1(\alpha R + \beta S, L^p)$ and $\Lambda^1(R, L^p) \cap \Lambda^1(S, L^p) \subseteq \Lambda^1(\alpha R + \beta S, L^p)$. For each $X \in \mathcal{L}^1(R, L^p) \cap \mathcal{L}^1(S, L^p)$, $\int X d(\alpha R + \beta S) = \alpha \int X dR + \beta \int X dS$. For each $X \in \Lambda^1(R, L^p) \cap \Lambda^1(S, L^p)$, $X \cdot (\alpha R + \beta S) = \alpha(X \cdot R) + \beta(X \cdot S)$.

Proof: Put $T = \alpha R + \beta S$. It is easy to check that

$$(\dagger) \quad \int X dT = \alpha \int X dR + \beta \int X dS \quad \text{for all } X \in \mathcal{E}.$$

It is easy to verify using $(\dagger)$ that T is an L^p -integrator. Let $\Phi = \Phi_R^p + \Phi_S^p$. Then Φ is an upper gauge on $(\mathbb{R}^{++} \times \Omega, \mathcal{E})$. Suppose now that $X \in \mathcal{L}^1(R, L^p) \cap \mathcal{L}^1(S, L^p)$. Then X is previsible and Φ -finite. Thus we see that $X \in \mathcal{L}^1(\Phi)$, and so there exist $X_1, X_2, \dots \in \mathcal{E}$ such that $\Phi(X_n - X) \rightarrow 0$. But letting $N \in \mathbb{N}$ be such that $|\alpha|, |\beta| \leq N$, we have using $(\dagger)$ that $\Phi_T^p \leq N\Phi_R^p + N\Phi_S^p = N\Phi$. So $\Phi_T^p(X_n - X) \rightarrow 0$ and therefore $X \in \mathcal{L}^1(T, L^p)$. Now also $\Phi_R^p, \Phi_S^p \leq \Phi$; so $\Phi_R^p(X_n - X) \rightarrow 0$ and $\Phi_S^p(X_n - X) \rightarrow 0$. But $(\dagger)$ gives that $\int X_n dT = \alpha \int X_n dR + \beta \int X_n dS$. So letting $n \rightarrow \infty$ we obtain that $\int X dT = \alpha \int X dR + \beta \int X dS$. The rest of the proposition is now easy to prove. $\square$

(4.20)Proposition. Suppose that $0 \leq p < \infty$ and that S is an L^p -integrator. Then every evanescent set $E \subseteq \mathbb{R}^{++} \times \Omega$ is Φ_S^p -null.

Thus the values of $\int X dS$ and $X \cdot S$ when they exist are not altered by replacing X by a previsible process indistinguishable from it.

Proof: By definition $E \subseteq \mathbb{R}^{++} \times F$ for some $\mathbb{P}$ -null subset F of Ω . Let $A_n = (0, n] \times F$. Then $\mathbf{1}_{A_n} \in \mathcal{E}$ and $\mathbf{1}_{A_n} \uparrow \mathbf{1}_{\mathbb{R}^{++} \times F}$. Thus $\mathbf{1}_{\mathbb{R}^{++} \times F} \in \mathcal{E}_{\text{inc}}^+$. Now for each $X \in \mathcal{E}$ such that $|X| \leq \mathbf{1}_{\mathbb{R}^{++} \times F}$, we have that $\int X dS = 0$ outside the null set F , and so $\rho_p(\int X dS) = 0$. Therefore taking the supremum gives that $\Phi_S^p(\mathbf{1}_{\mathbb{R}^{++} \times F}) = 0$. So $\mathbb{R}^{++} \times F$ is Φ_S^p -null and therefore so is E . $\square$

(4.21)Theorem. Suppose that $0 \leq p < \infty$, S is an L^p -integrator, and that $H \in \Lambda^1(S, L^p)$. Then the following are true.

- (i) $H \cdot S$ is an L^p -integrator (and in particular it is adapted).
- (ii) $\mathcal{L}^1(H \cdot S, L^p) = \{ X \mid X: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R} \text{ is previsible and } XH \in \mathcal{L}^1(S, L^p) \}$.

- (iii) If $X \in \mathcal{L}^1(H \bullet S, L^p)$, then $\int X d(H \bullet S) = \int XH dS$ and $\Phi_{H \bullet S}^p(X) = \Phi_S^p(XH)$.
- (iv) $\Lambda^1(H \bullet S, L^p) = \{X \mid X: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R} \text{ is previsible and } XH \in \Lambda^1(S, L^p)\}$.
- (v) If $X \in \Lambda^1(H \bullet S, L^p)$, then $X \bullet (H \bullet S) = (XH) \bullet S$.

Proof: It is clear that (iv) and (v) follow from (ii) and (iii). So we just consider (i), (ii) and (iii). Take $H_1, H_2, \dots \in \mathcal{E}$ such that $\Phi_S^p(H_n - \mathbf{1}_{(0,n] \times \Omega} H) < \frac{1}{n}$. Then $\Phi_S^p(\mathbf{1}_{(0,t] \times \Omega}(H_n - H)) \rightarrow 0$ as $n \rightarrow \infty$ for all $t \in \mathbb{R}^+$. We present the proof as a series of numbered steps.

(1) $H \bullet S$ is adapted and $\int X d(H \bullet S) = \int XH dS$ for all $X \in \mathcal{E}$.

Proof: Clearly $(H_n \bullet S)_t \rightarrow (H \bullet S)_t$ in L^p , and since $(H_n \bullet S)_t$ is $\mathcal{F}_t$ -measurable (by inspection), it follows that $(H \bullet S)_t$ is also $\mathcal{F}_t$ -measurable. Thus $H \bullet S$ is adapted. For the second part, we need only by linearity, consider the case when $X = \mathbf{1}_R$, where R is a previsible rectangle $(s, t] \times F$ ($0 \leq s < t < \infty$ and $F \in \mathcal{F}_s$). In this case $\int \mathbf{1}_R H_n dS = \mathbf{1}_F((H_n \bullet S)_t - (H_n \bullet S)_s)$. Now let $n \rightarrow \infty$, observing that both sides converge in L^p , to give that $\int \mathbf{1}_R H dS = \mathbf{1}_F((H \bullet S)_t - (H \bullet S)_s)$. But the right hand side is just $\int \mathbf{1}_R d(H \bullet S)$, and so (1) is proven.

From (1) it follows that $H \bullet S$ is an L^p -integrator (use the Dominated Convergence Theorem (4.17) to verify $(\dagger)$ of Definition(4.7)).

(2) If $K_1, K_2 \in \Lambda^1(S, L^p)$ and $X \in \mathcal{E}$, then $|\Phi_{K_1 \bullet S}^p(X) - \Phi_{K_2 \bullet S}^p(X)| \leq \Phi_S^p(X(K_1 - K_2))$.

Proof: For each $Y \in \mathcal{E}$ such that $|Y| \leq |X|$, we have by (1), that $\int Y d(K_1 \bullet S) = \int Y d(K_2 \bullet S) + \int Y(K_1 - K_2) dS$, and so

$$\begin{aligned} \rho_p(\int Y d(K_1 \bullet S)) &\leq \rho_p(\int Y d(K_2 \bullet S)) + \rho_p(\int Y(K_1 - K_2) dS) \\ &\leq \Phi_{K_2 \bullet S}^p(X) + \Phi_S^p(X(K_1 - K_2)). \end{aligned}$$

Now take the supremum over all such Y to give that $\Phi_{K_1 \bullet S}^p(X) \leq \Phi_{K_2 \bullet S}^p(X) + \Phi_S^p(X(K_1 - K_2))$. Reverse the roles of K_1 and K_2 to complete the proof.

(3) $\Phi_{H \cdot S}^p(X) = \Phi_S^p(XH)$ for all $X \in \mathcal{E}$.

Proof: Temporarily fix n . Then $H_n \in \mathcal{E}$. Therefore as Y runs over the set $\{U \in \mathcal{E} \mid |U| \leq |X|\}$, YH_n runs over all of the set $\{U \in \mathcal{E} \mid |U| \leq |XH_n|\}$. But $\int Y d(H_n \cdot S) = \int YH_n dS$, and so by applying ρ_p to both sides and taking supremums, we obtain that $\Phi_{H_n \cdot S}^p(X) = \Phi_S^p(XH_n)$. Let $n \rightarrow \infty$ to give the result, since $|\Phi_S^p(XH_n) - \Phi_S^p(XH)| \leq \Phi_S^p(X(H_n - H)) \rightarrow 0$, and $|\Phi_{H_n \cdot S}^p(X) - \Phi_{H \cdot S}^p(X)| \leq \Phi_S^p(X(H_n - H)) \rightarrow 0$ (by (2)).

Now define $\Phi, \Psi: (\overline{\mathbb{R}}^+)^{\mathbb{R}^{++} \times \Omega} \rightarrow \overline{\mathbb{R}}^+$ by $\Phi(X) = \Phi_S^p(X|H|)$ and $\Psi = \Phi_{H \cdot S}^p + \Phi$. Then Φ and Ψ are upper gauges on $(\mathbb{R}^{++} \times \Omega, \mathcal{E})$ (to check (4) of Definition(2.1)(b) for Φ , use the Monotone Convergence Theorem (2.13)).

(4) If $X \in \mathcal{L}^1(H \cdot S, L^p)$ and $XH \in \mathcal{L}^1(S, L^p)$, then $\int X d(H \cdot S) = \int XH dS$ and $\Phi_{H \cdot S}^p(X) = \Phi_S^p(XH)$.

Proof: X is both $\Phi_{H \cdot S}^p$ -finite and Φ -finite and is therefore also Ψ -finite ($\Psi = \Phi_{H \cdot S}^p + \Phi$). But X is previsible and thus it follows that $X \in \mathcal{L}^1(\Psi)$. Take therefore $X_1, X_2, \dots \in \mathcal{E}$ such that $\Psi(X_n - X) \rightarrow 0$. Now since $\Phi_{H \cdot S}^p, \Phi \leq \Psi$, it follows that $\Phi_{H \cdot S}^p(X_n - X) \rightarrow 0$ and $\Phi(X_n - X) \rightarrow 0$. The latter gives that $\Phi_S^p(X_n H - XH) \rightarrow 0$. But (1) and (3) give that $\int X_n d(H \cdot S) = \int X_n H dS$ and that $\Phi_{H \cdot S}^p(X_n) = \Phi_S^p(X_n H)$ respectively. So letting $n \rightarrow \infty$ gives that $\int X d(H \cdot S) = \int XH dS$ and that $\Phi_{H \cdot S}^p(X) = \Phi_S^p(XH)$ as required.

(5) The proof of part (ii) of the theorem.

Let $U_1, U_2, \dots \in \mathcal{E}^+$ be such that $U_n \uparrow \infty$ pointwise. Suppose that $X: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}$ is previsible. Then X is in the right hand side of (ii) if and only if XH is Φ_S^p -finite. This holds if and only if X is Φ -finite. Thus to prove (ii) it is enough to show that X is $\Phi_{H \cdot S}^p$ -finite if and only if X is Φ -finite. Suppose first that X is $\Phi_{H \cdot S}^p$ -finite and put $X_n = (X \wedge U_n) \vee (-U_n)$. So $X_n \in \mathcal{L}^1(H \cdot S, L^p)$, $X_n H \in \mathcal{L}^1(S, L^p)$ and $X_n \rightarrow X$ pointwise. Now $|X_n| \leq |X| \in \mathcal{F}(\Phi_{H \cdot S}^p)$ for all n , and so by the Dominated Convergence Theorem (2.14), we have that $\Phi_{H \cdot S}^p(X_n - X) \rightarrow 0$. So by (4) applied to

$X_n - X_m$, we obtain that $\Phi(X_n - X_m) = \Phi_{H \cdot S}^p(X_n - X_m) \rightarrow 0$ as $n, m \rightarrow \infty$. Thus by the completeness of $F(\Phi)$, there is a $Y \in \mathcal{F}(\Phi)$ such that $\Phi(X_n - Y) \rightarrow 0$. Therefore by Proposition(2.10) a subsequence of $(X_n)_n$ converges to Y Φ -a.e. and so $Y = X$ Φ -a.e. We have therefore that $X \in \mathcal{F}(\Phi)$ as required. The converse is similar.

Now part (iii) of the theorem follows from part (ii) and step (4). $\square$

(4.22)Corollary. *Suppose that $0 \leq p < \infty$, S is an L^p -integrator, and that $X \in \mathcal{L}^1(S, L^p)$. Then*

$$(\dagger) \quad \Phi_S^p(X) = \sup\{\rho_p(\int Y dS) \mid Y \in \mathcal{L}^1(S, L^p), \quad |Y| \leq |X|\}.$$

Proof: That the right hand side of $(\dagger)$ is no bigger than the left hand side follows easily from $(\dagger)$ of Definition(4.13)(a). For the reverse inequality note that $1X = X \in \mathcal{L}^1(S, L^p)$ (here 1 denotes the constant function taking value 1), and so by Theorem(4.21), we have that $1 \in \mathcal{L}^1(X \cdot S, L^p)$. Thus

$$\begin{aligned} \Phi_S^p(X) &= \Phi_{X \cdot S}^p(1) && \text{(by Theorem(4.21)(iii))} \\ &= \sup\{\rho_p(\int Z d(X \cdot S)) \mid Z \in \mathcal{E}, \quad |Z| \leq 1\} && \text{(since } 1 \in \mathcal{E}_{\text{inc}}^+) \\ &= \sup\{\rho_p(\int ZX dS) \mid Z \in \mathcal{E}, \quad |Z| \leq 1\} && \text{(by Theorem(4.21)(iii))} \\ &\leq \sup\{\rho_p(\int Y dS) \mid Y \in \mathcal{L}^1(S, L^p), \quad |Y| \leq |X|\}, \end{aligned}$$

as required. $\square$

(4.23)Proposition. *Suppose that $0 \leq p < \infty$, S is an L^p -integrator, and that $X: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}$ is given by $X_t(\omega) = \sum_{k=0}^{r-1} C_k(\omega) \mathbf{1}_{(t_k, t_{k+1}]}(t)$, where $0 = t_0 < \dots < t_r < \infty$, and for $k = 0, \dots, r-1$, C_k is a bounded $\mathcal{F}_{t_k}$ -measurable real-valued random variable. Then $X \in \mathcal{L}^1(S, L^p)$ and $\int X dS = \sum_{k=0}^{r-1} C_k(S_{t_{k+1}} - S_{t_k})$.*

Proof: Suppose that $M \in \mathbb{R}^+$ is a bound on the absolute values of $C_0, \dots, C_{r-1}$. Put

$$X_t^{(n)}(\omega) = \sum_{k=0}^{r-1} C_{k,n}(\omega) \mathbf{1}_{(t_k, t_{k+1}]}(t),$$

where for $k = 0, \dots, r-1$, $C_{k,n}$ is a simple $\mathcal{F}_{t_k}$ -measurable real-valued random variable, bounded in absolute value by M , and tending to C_k pointwise as $n \rightarrow \infty$. Then $X^{(n)} \in \mathcal{E}$, $X^{(n)} \rightarrow X$ pointwise and $|X^{(n)}| \leq M \mathbf{1}_{(0, t_r] \times \Omega} \in \mathcal{E}^+$ for all n . So by the Dominated Convergence Theorem (4.17), we have that $X \in \mathcal{L}^1(S, L^p)$ and $\int X dS = \lim_{n \rightarrow \infty} \int X_n dS = \sum_{k=0}^{r-1} C_k (S_{t_{k+1}} - S_{t_k})$ as required. $\square$

§4.4 SOME RESULTS IN THE THEORY OF L^0 -STOCHASTIC INTEGRATION

In this section we shall be considering L^0 -stochastic integration and shall prove some results about the existence of stochastic integrals which are not generally true for the other L^p spaces. These results will illustrate the power of considering L^0 -integrators. In the Strasbourg approach these would require the use of localization (see §5.3). We shall be using the notation and terminology of (4.15). Recall also the definition of X^* given in Definition(4.2)(g).

(4.24)Theorem. *Suppose that S is an integrator. Then the following are true.*

- (i) *If $X: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}$ is a l\`adc\`ag process, then $X \in \Lambda^1(S)$.*
- (ii) *For every $b \in \mathbb{R}^{++}$ and $\varepsilon > 0$, there exists a $\delta > 0$ such that $\Phi_S^0(\mathbf{1}_{(0,b] \times \Omega} X) < \varepsilon$ for each l\`adc\`ag process $X: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}$ such that $\rho_0(X_b^*) < \delta$.*

Note. A net $(X^{(\alpha)})_\alpha$ of l\`adc\`ag processes is said to converge *uniformly in probability* to a l\`adc\`ag process X on $(0, b] \times \Omega$ if and only if $\rho_0((X^{(\alpha)} - X)_b^*) \rightarrow 0$. Thus part (ii) shows that in this case $\Phi_S^0(\mathbf{1}_{(0,b] \times \Omega}(X^{(\alpha)} - X)) \rightarrow 0$, and hence also $\int_0^b X^{(\alpha)} dS \rightarrow \int_0^b X dS$ in probability.

Before proving this we need a couple of lemmas.

(4.25)Lemma. *Suppose that S is an integrator, $X \in \mathcal{E}_{\text{inc}}^+$, and that $[X \neq 0] \subseteq \mathbb{R}^{++} \times F$, where $F \in \mathcal{F}$. Then $\Phi_S^0(X) \leq \mathbb{P}(F)$.*

Proof: Suppose that $Y \in \mathcal{E}$ is such that $|Y| \leq X$. Then $[Y \neq 0] \subseteq \mathbb{R}^{++} \times F$. Therefore $\int Y dS$ is zero outside F (see Definition(4.6)), and thus we have that $1 \wedge |\int Y dS| \leq \mathbf{1}_F$. By taking expectation we obtain that $\rho_0(\int Y dS) \leq \mathbb{P}(F)$. Now take the supremum over all such Y to give that $\Phi_S^0(X) \leq \mathbb{P}(F)$ as required. $\square$

(4.26) Lemma.

- (i) Suppose that $X: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}^+$ is given by $X_t(\omega) = \sum_{k=0}^{r-1} C_k(\omega) \mathbf{1}_{(t_k, t_{k+1}]}(t)$, where $0 = t_0 < \dots < t_r < \infty$, and for $k = 0, \dots, r-1$, C_k is a non-negative $\mathcal{F}_{t_k}$ -measurable random variable. Then $X \in \mathcal{E}_{\text{inc}}^+$.
- (ii) If $X: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}^+$ is a non-negative increasing left continuous process, then $X \in \mathcal{E}_{\text{inc}}^+$.

Proof:

(i) For $k = 0, \dots, r-1$, take simple non-negative $\mathcal{F}_{t_k}$ -measurable random variables $C_{k,1}, C_{k,2}, \dots$ such that $C_{k,n} \uparrow C_k$ pointwise as $n \rightarrow \infty$. Let $X_t^{(n)}(\omega) = \sum_{k=0}^{r-1} C_{k,n}(\omega) \mathbf{1}_{(t_k, t_{k+1}]}(t)$. Then $X^{(n)} \in \mathcal{E}^+$ (see Definition(4.5)). But $X^{(n)} \uparrow X$ pointwise; so $X \in \mathcal{E}_{\text{inc}}^+$.

(ii) Define $X^{(n)}: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}^+$ by

$$X_t^{(n)}(\omega) = \sum_{k=0}^{2^{2n}-1} X_{k2^{-n}}(\omega) \mathbf{1}_{(k2^{-n}, (k+1)2^{-n}]}(t).$$

Then $X^{(n)} \uparrow X$ pointwise. But by (i), $X^{(n)} \in \mathcal{E}_{\text{inc}}^+$, and so by Proposition(4.11) it follows that $X \in \mathcal{E}_{\text{inc}}^+$. $\square$

Proof of (4.24): We shall prove part (ii) first.

(ii) Take $\lambda > 0$ such that $\Phi_S^0(\lambda \mathbf{1}_{(0,b] \times \Omega}) < \frac{\varepsilon}{2}$, and then take $\delta > 0$ such that $\mathbb{P}[|W| > \lambda] < \frac{\varepsilon}{2}$, whenever $W \in L^0$ with $\rho_0(W) < \delta$. Now suppose that $X: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}$ is a l\`ad\`c\`ag process such that $\rho_0(X_b^*) < \delta$. $X^* \in \mathcal{E}_{\text{inc}}^+$ by Lemma(4.26)(ii). Put

$Z = \mathbf{1}_{(0,b] \times \Omega} X^*$. We have that

$$(1) \quad Z = \mathbf{1}_{[Z > \lambda]} Z + \mathbf{1}_{[Z \leq \lambda]} Z.$$

Now if $\mathcal{E}^+ \ni Y^{(n)} \uparrow X^*$ then $\mathcal{E}^+ \ni \mathbf{1}_{(0,b] \times \Omega} \mathbf{1}_{[Y^{(n)} > \lambda]} Y^{(n)} \uparrow \mathbf{1}_{[Z > \lambda]} Z$. Thus $\mathbf{1}_{[Z > \lambda]} Z \in \mathcal{E}_{\text{inc}}^+$. But $[Z > \lambda] \subseteq \mathbb{R}^{++} \times [X_b^* > \lambda]$. Therefore by Lemma(4.25) it follows that

$$(2) \quad \Phi_S^0(\mathbf{1}_{[Z > \lambda]} Z) \leq \mathbb{P}[X_b^* > \lambda] < \frac{\varepsilon}{2},$$

(since $\rho_0(X_b^*) < \delta$). Now $|\mathbf{1}_{[Z \leq \lambda]} Z| \leq \lambda \mathbf{1}_{(0,b] \times \Omega}$. So $\Phi_S^0(\lambda \mathbf{1}_{(0,b] \times \Omega}) < \frac{\varepsilon}{2}$ gives that

$$(3) \quad \Phi_S^0(\mathbf{1}_{[Z \leq \lambda]} Z) < \frac{\varepsilon}{2}.$$

From (1), (2) and (3) we deduce that $\Phi_S^0(Z) < \varepsilon$. But $|\mathbf{1}_{(0,b] \times \Omega} X| \leq Z$. Thus $\Phi_S^0(\mathbf{1}_{(0,b] \times \Omega} X) < \varepsilon$ as required.

(i) For $b \in \mathbb{R}^{++}$, we have that $\rho_0((\lambda X)_b^*) = \rho_0(\lambda X_b^*) \rightarrow 0$ as $\lambda \downarrow 0$. Therefore (ii) applied with λX in place of X gives that $\Phi_S^0(\lambda \mathbf{1}_{(0,b] \times \Omega} X) \rightarrow 0$ as $\lambda \downarrow 0$. Thus $\mathbf{1}_{(0,b] \times \Omega} X$ is Φ_S^0 -finite. But it is also previsible. Thus it is an element of $\mathcal{L}^1(S)$. So $X \in \Lambda^1(S)$. □

We shall now show that in the L^0 case the requirement in Proposition(4.23) that $C_0, \dots, C_{r-1}$ be bounded may be removed.

(4.27)Proposition. Suppose that S is an integrator and that $X: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}$ is given by

$$X_t(\omega) = \sum_{k=0}^{r-1} C_k(\omega) \mathbf{1}_{(t_k, t_{k+1}]}(t),$$

where $0 = t_0 < \dots < t_r < \infty$, and for $k = 0, \dots, r-1$, C_k is a $\mathcal{F}_{t_k}$ -measurable real-valued random variable. Then $X \in \mathcal{L}^1(S)$ and $\int X dS = \sum_{k=0}^{r-1} C_k(S_{t_{k+1}} - S_{t_k})$.

Proof: We have that $X \in \Lambda^1(S)$ by Theorem(4.24)(i). Thus $X = \mathbf{1}_{(0, t_r] \times \Omega} X \in \mathcal{L}^1(S)$. Put $X^{(n)} = (X \wedge n) \vee (-n)$. Then $X^{(n)} \rightarrow X$ pointwise and $|X^{(n)}| \leq |X| \in \mathcal{L}^1(S)$

for all n . Thus by the Dominated Convergence Theorem (4.17), $\int X^{(n)} dS \rightarrow \int X dS$ in probability. By Proposition(4.23) we have that $\int X^{(n)} dS = \sum_{k=0}^{r-1} ((C_k \wedge n) \vee (-n))(S_{t_{k+1}} - S_{t_k})$. Thus letting $n \rightarrow \infty$ gives that $\int X dS = \sum_{k=0}^{r-1} C_k(S_{t_{k+1}} - S_{t_k})$ as required. $\square$

(4.28)Proposition. Suppose that S is an integrator and that $X: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}$ is a làdcàg process. Then for each $t \in \mathbb{R}^+$,

$$\sum_{k=0}^{r-1} X_{t_k}(S_{t_{k+1}} - S_{t_k}) \rightarrow \int_0^t X dS \quad \text{in probability,}$$

and

$$\sum_{k=0}^{r-1} X_{t_k+}(S_{t_{k+1}} - S_{t_k}) \rightarrow \int_0^t X dS \quad \text{in probability,}$$

as $\sup_k(t_{k+1} - t_k) \rightarrow 0$ with $0 = t_0 < \dots < t_r = t$. (We have that $X \in \Lambda^1(S)$ by Theorem(4.24)(i).)

Proof: By Theorem(4.24)(i), $X^* \in \Lambda^1(S)$. Suppose that $0 = t_0^{(n)} < \dots < t_{r_n}^{(n)} = t$ and that $\sup_k(t_{k+1}^{(n)} - t_k^{(n)}) \rightarrow 0$ as $n \rightarrow \infty$. Define $Z^{(n)}, W^{(n)}: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}$ by

$$Z_t^{(n)}(\omega) = \sum_{k=0}^{r_n-1} X_{t_k^{(n)}}(\omega) \mathbf{1}_{(t_k^{(n)}, t_{k+1}^{(n)}]}(t),$$

$$W_t^{(n)}(\omega) = \sum_{k=0}^{r_n-1} X_{t_k^{(n)}+}(\omega) \mathbf{1}_{(t_k^{(n)}, t_{k+1}^{(n)}]}(t).$$

Now since X^* has monotone paths we have that

$$\begin{aligned} |Z_t^{(n)}(\omega)|, |W_t^{(n)}(\omega)| &\leq \sum_{k=0}^{r_n-1} X_{t_k^{(n)}+}^*(\omega) \mathbf{1}_{(t_k^{(n)}, t_{k+1}^{(n)}]}(t) \\ &\leq (\mathbf{1}_{(0,t] \times \Omega} X^*)_t(\omega). \end{aligned}$$

I.e. $|Z^{(n)}|, |W^{(n)}| \leq \mathbf{1}_{(0,t] \times \Omega} X^* \in \mathcal{L}^1(S)$ for all n . But $Z^{(n)}, W^{(n)} \rightarrow \mathbf{1}_{(0,t] \times \Omega} X$ pointwise. Thus from the Dominated Convergence Theorem (4.17), we obtain that

$\int Z^{(n)} dS, \int W^{(n)} dS \rightarrow \int_0^t X dS$ in probability (we know that $Z^{(n)}, W^{(n)} \in \mathcal{L}^1(S)$ by Proposition(4.27)). By evaluating $\int Z^{(n)} dS$ and $\int W^{(n)} dS$ using Proposition(4.27), this gives that

$$\sum_{k=0}^{r_n-1} X_{t_k^{(n)}} (S_{t_{k+1}^{(n)}} - S_{t_k^{(n)}}) \rightarrow \int_0^t X dS \quad \text{in probability,}$$

and

$$\sum_{k=0}^{r_n-1} X_{t_k^{(n)}+} (S_{t_{k+1}^{(n)}} - S_{t_k^{(n)}}) \rightarrow \int_0^t X dS \quad \text{in probability.}$$

Now since L^0 is a metric space the consideration of sequences is enough to prove the result. $\square$

We can now give the following corollary to Proposition(4.28).

(4.29)Corollary. *Suppose that S is an integrator and that X is a càdlàg process. Then $X_- \in \Lambda^1(S)$ and for all $t \in \mathbb{R}^+$,*

$$\sum_{k=0}^{r-1} X_{t_k} (S_{t_{k+1}} - S_{t_k}) \rightarrow \int_0^t X_- dS \quad \text{in probability,}$$

as $\sup_k (t_{k+1} - t_k) \rightarrow 0$ with $0 = t_0 < \dots < t_r = t$.

Proof: X_- is a làdcàg process, and therefore is, by Theorem(4.24)(i), an element of $\Lambda^1(S)$. Now since $U \mapsto U_-$ and $U \mapsto U_+$ are mutual inverses, applying Proposition(4.28) to X_- gives the result. $\square$

§4.5 AN ALTERNATIVE CHARACTERIZATION OF L^p -INTEGRATORS

We shall now give a characterization of integrators that will be needed later. This result can be found in Bichteler [Bi2] (Theorem 2.6), though his proof appears to be incorrect (or at best incomplete). Also the proof here for the case $p > 0$ is simpler than his; it uses the ‘Stone-Čech’ argument (see the proof of Proposition(3.15)) rather than uniform spaces.

(4.30)Theorem. Suppose that $0 \leq p < \infty$, and that $S: \mathbb{R}^+ \rightarrow L^p$ is adapted. Then S is an L^p -integrator if and only if the following conditions are satisfied.

- (1) S is right continuous in probability (i.e. $S_u \rightarrow S_t$ in probability as $u \downarrow t$ for all $t \in \mathbb{R}^+$).
- (2) For each $t \in \mathbb{R}^+$, $\int_0^t \dots dS: \mathcal{E} \rightarrow L^p$ is continuous when $\mathcal{E}$ is given the supremum norm.

Proof: The necessity of (1) and (2) follows from Corollary(4.18) and Proposition(4.8) respectively. Now suppose that S satisfies (1) and (2). By Proposition(4.8), we have only to check that (2) of that proposition is satisfied by S . We consider first the case when $p = 0$. Define Φ_S^0 as in Definition(4.9). We do not of course know that Φ_S^0 is an upper gauge; but clearly it satisfies (2) of Definition(2.1)(b). Also, from (2) of the hypothesis, Φ_S^0 satisfies (1) of Definition(2.1)(b). As in the proof of Theorem(3.5), we can prove that $\Phi_S^0(X + Y) \leq \Phi_S^0(X) + \Phi_S^0(Y)$ for all $X, Y \in \mathcal{E}^+$. Now I claim that the following is true.

- (a) For each $t \in \mathbb{R}^+$, $\Phi_S^0(\mathbf{1}_{(t,u] \times \Omega}) \rightarrow 0$ as $u \downarrow t$.

To prove (a), suppose that it fails for some $t \in \mathbb{R}^+$. Then there exists an $\varepsilon > 0$ such that $\Phi_S^0(\mathbf{1}_{(t,u] \times \Omega}) > \varepsilon$ for all $u > t$. Now I claim that for each $u > t$, there exist u' and X such that $t < u' < u$, $X \in \mathcal{E}$, $|X| \leq \mathbf{1}_{(u',u] \times \Omega}$, and $\rho_0(\int X dS) > \varepsilon$. To prove this claim, note that since $\Phi_S^0(\mathbf{1}_{(t,u] \times \Omega}) > \varepsilon$, there must be a $Y \in \mathcal{E}$ such that $|Y| \leq \mathbf{1}_{(t,u] \times \Omega}$, and $\rho_0(\int Y dS) > \varepsilon$. Now as $u' \downarrow t$, $\int \mathbf{1}_{(u',u] \times \Omega} Y dS \rightarrow \int Y dS$ in probability (by (1) and an inspection of Definition(4.6)). Therefore there exists a u' such that $t < u' < u$, and $\rho_0(\int \mathbf{1}_{(u',u] \times \Omega} Y dS) > \varepsilon$. Take $X = \mathbf{1}_{(u',u] \times \Omega} Y$. Now using this define inductively $u_1 > u_2 > \dots > t$ and $X_1, X_2, \dots \in \mathcal{E}$ such that for each n , $|X_n| \leq \mathbf{1}_{(u_{n+1}, u_n] \times \Omega}$, and $\rho_0(\int X_n dS) > \varepsilon$. Let $Q_n = \int X_n dS$, and take

$$\mathcal{A} = \{ \sum_{n=1}^N \alpha_n Q_n \mid N \in \mathbb{N}_{>0}, \quad \alpha_1, \dots, \alpha_N \in [-1, 1] \}.$$

I claim that $\mathcal{A}$ is bounded in L^0 . Suppose that $Q = \sum_{n=1}^N \alpha_n Q_n \in \mathcal{A}$ ($\alpha_1, \dots, \alpha_N \in [-1, 1]$). Then let $X = \sum_{n=1}^N \alpha_n X_n$; so $X \in \mathcal{E}$, $|X| \leq \mathbf{1}_{(t, u_1] \times \Omega}$, and $Q = \int X dS$. Thus $\rho_0(\lambda Q) \leq \Phi_S^0(\lambda \mathbf{1}_{(t, u_1] \times \Omega})$ for all $\lambda > 0$. So $\sup_{Q \in \mathcal{A}} \rho_0(\lambda Q) \leq \Phi_S^0(\lambda \mathbf{1}_{(t, u_1] \times \Omega}) \rightarrow 0$ as $\lambda \downarrow 0$. I.e. $\mathcal{A}$ is bounded in L^0 . Therefore since L^0 is a C-space, $Q_n \rightarrow 0$ in probability. But $\rho_0(Q_n) > \varepsilon$ for all n . This gives the required contradiction, and so (a) is proven.

Now we shall prove that (2) of Proposition(4.8) holds. So suppose that $X_1, X_2, \dots \in \mathcal{E}^+$ are such that $X_n \downarrow 0$ pointwise. Then by scaling, we may assume that $X_1 \leq 1$. Suppose that $\varepsilon > 0$. Write

$$(b) \quad X_n(t, \omega) = \sum_{k=0}^{r_n-1} C_k^{(n)}(\omega) \mathbf{1}_{(t_k^{(n)}, t_{k+1}^{(n)}]}(t),$$

where $0 = t_0^{(n)} < \dots < t_{r_n}^{(n)} < \infty$, and for $k = 0, \dots, r_n - 1$, $C_k^{(n)}$ is a simple $\mathcal{F}_{t_k^{(n)}}$ -measurable random variable with $0 \leq C_k^{(n)} \leq 1$. Now by (a) take $u_k^{(n)} \in (t_k^{(n)}, t_{k+1}^{(n)})$ such that

$$(c) \quad \sum_{k=0}^{r_n-1} \Phi_S^0(\mathbf{1}_{(t_k^{(n)}, u_k^{(n)}] \times \Omega}) < \frac{\varepsilon}{3} 2^{-n}.$$

Then define $Y_n, U_n: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}$ by

$$(d) \quad Y_n(t, \omega) = \sum_{k=0}^{r_n-1} C_k^{(n)}(\omega) \mathbf{1}_{(u_k^{(n)}, t_{k+1}^{(n)}]}(t),$$

$$(e) \quad U_n(t, \omega) = \sum_{k=0}^{r_n-1} C_k^{(n)}(\omega) \mathbf{1}_{[u_k^{(n)}, t_{k+1}^{(n)}]}(t).$$

So $Y_n \in \mathcal{E}^+$ and $Y_n \leq U_n \leq X_n$. Now since

$$X_n(t, \omega) - Y_n(t, \omega) = \sum_{k=0}^{r_n-1} C_k^{(n)}(\omega) \mathbf{1}_{(t_k^{(n)}, u_k^{(n)}]}(t),$$

it follows that $\Phi_S^0(X_n - Y_n) \leq \sum_{k=0}^{r_n-1} \Phi_S^0(\mathbf{1}_{(t_k^{(n)}, u_k^{(n)}] \times \Omega}) < \frac{\varepsilon}{3} 2^{-n}$. Now put $Z_n = Y_1 \wedge \cdots \wedge Y_n$, and $V_n = U_1 \wedge \cdots \wedge U_n$. So $Z_n \in \mathcal{E}$, $Z_n \leq V_n \leq X_n$, $Z_n \downarrow 0$, and $V_n \downarrow 0$. Now $0 \leq X_n - Z_n \leq \sum_{k=0}^n (X_k - Y_k)$. Therefore $\Phi_S^0(X_n - Z_n) \leq \sum_{k=0}^n \Phi_S^0(X_k - Y_k) < \frac{\varepsilon}{3}$, and so

$$(f) \quad \rho_0 \left(\int X_n dS - \int Z_n dS \right) < \frac{\varepsilon}{3} \quad \text{for all } n.$$

Define a random variable $H_n: \Omega \rightarrow \mathbb{R}^+$ by $H_n(\omega) = \sup_{t \in \mathbb{R}^{++}} |Z_n(t, \omega)|$. I claim that $H_n \downarrow 0$ pointwise. Now suppose that $\eta > 0$ is given. For each $\omega \in \Omega$, let $K_n(\omega) = \{t \in \mathbb{R}^{++} \mid V_n(t, \omega) \geq \eta\}$. Then $K_n(\omega) \downarrow \emptyset$, and $K_n(\omega)$ is compact (by the definitions of U_n and V_n). Therefore there exists an n such that $K_n(\omega) = \emptyset$ (n depending on ω). Thus $Z_n(t, \omega) \leq V_n(t, \omega) < \eta$ for all $t \in \mathbb{R}^{++}$, and so $H_n(\omega) \leq \eta$. Thus we have proved that $H_n \downarrow 0$ pointwise as required.

Now take $b \in \mathbb{R}^+$ such that $[X_1 \neq 0] \subseteq (0, b] \times \Omega$, and then, by (2), take $\eta > 0$ such that

$$(g) \quad W \in \mathcal{E} \text{ and } |W| \leq \eta \implies \rho_0 \left(\int_0^b W dS \right) < \frac{\varepsilon}{3},$$

Put $W_n = Z_n \wedge \eta$; then by (g),

$$(h) \quad \rho_0 \left(\int W_n dS \right) < \frac{\varepsilon}{3} \quad \text{for all } n.$$

Now $[H_n > \eta] \downarrow \emptyset$. Take n_0 such that $\mathbb{P}[H_n > \eta] < \frac{\varepsilon}{3}$ for all $n \geq n_0$. But for each n , $Z_n = W_n$ outside $\mathbb{R}^{++} \times [H_n > \eta]$, and thus $\int Z_n dS = \int W_n dS$ outside $[H_n > \eta]$. So by the definition of ρ_0 , we obtain that

$$(i) \quad \rho_0 \left(\int Z_n dS - \int W_n dS \right) \leq \mathbb{P}[H_n > \eta] < \frac{\varepsilon}{3} \quad \text{for all } n \geq n_0.$$

Therefore from (f), (h) and (i) we deduce that $\rho_0(\int X_n dS) < \varepsilon$ for all $n \geq n_0$. We have thus shown that $\int X_n dS \rightarrow 0$ in probability as required.

Now consider the case when $p > 0$. Suppose that $X_1, X_2, \dots \in \mathcal{E}^+$ such that $X_n \downarrow 0$ pointwise. By taking $b \in \mathbb{R}^+$ such that $[X_1 \neq 0] \subseteq (0, b] \times \Omega$, and applying Proposition(3.15) with $(0, b] \times \Omega$ as Γ , the set of restrictions of elements of $\mathcal{E}$ as $\mathcal{E}$, $\int_0^b \dots dS$ as J , and L^p as $\mathcal{X}$, we deduce that $\lim_{n \rightarrow \infty} \int X_n dS$ exists in L^p . But from the case when $p = 0$, we obtain that $\int X_n dS \rightarrow 0$ in probability, and so this limit must be zero. Hence (2) of Proposition(4.8) is verified. $\square$

(4.31)Proposition. *Suppose that $0 \leq p < \infty$, and that $S: \mathbb{R}^+ \rightarrow L^p$ is adapted and such that $\int_0^t \dots dS: \mathcal{E} \rightarrow L^p$ is continuous for all $t \in \mathbb{R}^+$, when $\mathcal{E}$ has the supremum norm. Then for each $t \in \mathbb{R}^+$, the limit $\lim_{u \downarrow t} S_u$ exists in L^p . (So if this limit is S_t for all t , then S is an L^p -integrator.)*

Proof: Fix a $t \in \mathbb{R}^+$, and suppose that $u_n > t$ and $u_n \downarrow t$. Then let $X_n = \mathbf{1}_{(t, u_n] \times \Omega}$. So $X_n \in \mathcal{E}^+$ and $X_n \downarrow 0$ pointwise. Therefore as in the proof of Theorem(4.30) the limit $\lim_{n \rightarrow \infty} \int X_n dS$ exists in L^p . But $\int X_n dS = S_{u_n} - S_t$, and so we must have that $\lim_{n \rightarrow \infty} S_{u_n}$ exists in L^p . Now since L^p is a metric space we will have completed the proof if we show that the value of this limit is independent of the choice of $(u_n)_n$. So suppose also that $v_n > t$ such that $v_n \downarrow t$. Let $w_1 \geq w_2 \geq \dots > t$ be such that $\{w_1, w_2, \dots\} = \{u_1, u_2, \dots\} \cup \{v_1, v_2, \dots\}$. Then $w_n \downarrow t$ and so $\lim_{n \rightarrow \infty} S_{w_n}$ must exist in L^p . So $\lim_{n \rightarrow \infty} S_{u_n}$ and $\lim_{n \rightarrow \infty} S_{v_n}$ must both be equal to this. Thus $\lim_{n \rightarrow \infty} S_{u_n} = \lim_{n \rightarrow \infty} S_{v_n}$ as required. $\square$

CHAPTER 5: INTEGRATORS AS PATHWISE CÀDLÀG PROCESSES

We shall present in this chapter, firstly a proof that an integrator has a pathwise càdlàg version, and then some results that require us to think of integrators as pathwise càdlàg processes.

We shall assume throughout this chapter that $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t \in \mathbb{R}^+})$ is a standard setup, and will use the notation $(\mathcal{E}, \mathcal{P}, \text{etc.})$ of Chapter 4.

§5.1 AN INTRODUCTION TO STOPPING TIMES

(5.1) Definition.

- (a) A *stopping time* is a function $\sigma: \Omega \rightarrow \overline{\mathbb{R}}^+$ such that $[\sigma \leq t] \in \mathcal{F}_t$ for all $t \in \mathbb{R}^+$.
- (b) For $\sigma, \tau: \Omega \rightarrow \overline{\mathbb{R}}^+$ such that $\sigma \leq \tau$, we let $(\sigma, \tau]$ denote the subset $\{(t, \omega) \in \mathbb{R}^+ \times \Omega \mid \sigma(\omega) < t \leq \tau(\omega)\}$ of $\mathbb{R}^+ \times \Omega$. Similarly we define (σ, τ) , $[\sigma, \tau]$ and $[\sigma, \tau)$.
- (c) If σ and τ are stopping times, then $(\sigma, \tau]$, (σ, τ) , $[\sigma, \tau]$ and $[\sigma, \tau)$ are called *stochastic intervals*.
- (d) A stopping time σ is said to be *finite* if and only if $\sigma < \infty$ on Ω . It is said to be *bounded* if and only if $\sigma \leq b$ on Ω for some $b \in \mathbb{R}^+$. It is said to be *simple* if and only if its range $\sigma(\Omega)$ is a finite set. It is said to be *σ -simple* if and only if its range is countable. (I apologize for this double usage of the letter σ .) $\square$

(5.2) Proposition. Suppose that $\sigma: \Omega \rightarrow \overline{\mathbb{R}}^+$. Then the following are equivalent.

- (1) σ is a stopping time.
- (2) $[\sigma < t] \in \mathcal{F}_t$ for all $t \in \mathbb{R}^+$.
- (3) $(\sigma, \infty) \in \mathcal{P}$ (see Definition(4.3)(b)).

Proof: First suppose (1) and deduce (2). We have $[\sigma < t] = \bigcup_{n=1}^{\infty} [\sigma \leq t - \frac{1}{n}] \in \mathcal{F}_t$ for all $t \in \mathbb{R}^+$ as required.

Now suppose (2) holds and deduce (1). For $\varepsilon > 0$ we have $[\sigma \leq t] = \bigcap_{n=1}^{\infty} [\sigma < t + \frac{\varepsilon}{n}] \in \mathcal{F}_{t+\varepsilon}$ and thus $[\sigma \leq t] \in \bigcap_{\varepsilon>0} \mathcal{F}_{t+\varepsilon} = \mathcal{F}_t$ by the right continuity of the filtration. So (1) holds as required.

Now suppose (1) and deduce (3). We have $(\sigma, \infty) = \bigcup_{q \in \mathbb{Q}^+} (q, \infty) \times [\sigma \leq q] \in \mathcal{P}$, since the right hand side is a countable union of previsible sets.

Finally suppose (3) and deduce (2). We note that for $A \in \mathcal{P}$,

$$(\dagger) \quad \{\omega \in \Omega \mid (t, \omega) \in A\} \in \mathcal{F}_t \quad \text{for all } t \in \mathbb{R}^+.$$

This is true because the set of all A satisfying $(\dagger)$ form a σ -field which contains all the previsible rectangles, and therefore it must be a superset of $\mathcal{P}$. Now take A to be the set (σ, ∞) . Then $(\dagger)$ says that $[\sigma < t] \in \mathcal{F}_t$ for all t as required. $\square$

It follows from (3) above that $(\sigma, \tau] \in \mathcal{P}$ when $\sigma \leq \tau$ are stopping times (since $(\sigma, \tau] = (\sigma, \infty) \setminus (\tau, \infty)$).

We shall now see that given a stopping time σ , we have a σ -field $\mathcal{F}_\sigma$ analogous to $\mathcal{F}_t$ for $t \in \mathbb{R}^+$.

(5.3)Definition. Suppose that σ is a stopping time. Then we define $\mathcal{F}_\sigma$ to be the set of all $A \in \mathcal{F}$ such that $A \cap [\sigma \leq t] \in \mathcal{F}_t$ for all $t \in \mathbb{R}^+$. $\square$

It is easy to show that $\mathcal{F}_\sigma$ is a σ -field when σ is a stopping time. Corresponding to Proposition(5.2) we have the following characterizations of $\mathcal{F}_\sigma$.

(5.4)Proposition. Suppose that σ is a stopping time and $A \in \mathcal{F}$. Then the following are equivalent.

- (1) $A \in \mathcal{F}_\sigma$.
- (2) $A \cap [\sigma < t] \in \mathcal{F}_t$ for all $t \in \mathbb{R}^+$.
- (3) $(\mathbb{R}^{++} \times A) \cap (\sigma, \infty) \in \mathcal{P}$.

(4) $\tilde{\sigma}$ is a stopping time, where $\tilde{\sigma}$ is equal to σ on A , but equal to infinity on $\Omega \setminus A$.

Proof: Taking $\tilde{\sigma}$ as in (4), we have that $[\tilde{\sigma} \leq t] = A \cap [\sigma \leq t]$. Thus $A \in \mathcal{F}_\sigma$ if and only if $\tilde{\sigma}$ is a stopping time. I.e. (1) holds if and only if (4) holds. For the rest of the proof apply Proposition(5.2) to $\tilde{\sigma}$, noting that $[\tilde{\sigma} < t] = A \cap [\sigma < t]$ and $(\tilde{\sigma}, \infty) = (\mathbb{R}^{++} \times A) \cap (\sigma, \infty)$. $\square$

(5.5)Notation. Suppose that $\sigma: \Omega \rightarrow \bar{\mathbb{R}}^+$ and that X is a real-valued function defined on a subset of $\bar{\mathbb{R}}^+ \times \Omega$ (e.g. on $\mathbb{R}^+ \times \Omega$). Then we let X_σ denote the real-valued function defined by $X_\sigma(\omega) = X_{\sigma(\omega)}(\omega)$ on the subset of Ω for which the right hand side exists. $\square$

We shall now give many elementary properties of stopping times in the following proposition.

(5.6)Proposition.

- (a) If σ and τ are stopping times such that $\sigma \leq \tau$, then $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$.
- (b) If $t \in \mathbb{R}^+$ and $\tau: \Omega \rightarrow \bar{\mathbb{R}}^+$ by $\tau(\omega) = t$ for all ω , then τ is a stopping time and $\mathcal{F}_\tau = \mathcal{F}_t$.
- (c) If σ and τ are stopping times then so are $\sigma \vee \tau$ and $\sigma \wedge \tau$.
- (d) Suppose that $(\sigma_n)_n$ is a sequence of stopping times and $\sigma_n \downarrow \sigma$ pointwise, where $\sigma: \Omega \rightarrow \bar{\mathbb{R}}^+$. Then σ is also a stopping time and $\mathcal{F}_\sigma = \bigcap_{n=1}^{\infty} \mathcal{F}_{\sigma_n}$.
- (e) Suppose again that $(\sigma_n)_n$ is a sequence of stopping times, but this time $\sigma_n \uparrow \sigma$ pointwise, where $\sigma: \Omega \rightarrow \bar{\mathbb{R}}^+$. Then again σ is a stopping time.
- (f) Suppose that σ is a stopping time. Then there exists a sequence $(\sigma_n)_n$ of stopping times such that $\sigma_n \downarrow \sigma$, and each σ_n satisfies certain properties which are given by the following statements.
 - (i) If σ is finite, we can take each σ_n to be a finite σ -simple stopping time.
 - (ii) If σ is bounded, we can take each σ_n to be a bounded simple stopping time.

(iii) If σ is any stopping time, we can take each σ_n to be a simple (but not necessarily finite) stopping time.

(g) If X is a càdlàg process and σ is a finite stopping time, then X_σ is $\mathcal{F}_\sigma$ -measurable.

Proof:

(a) For $A \in \mathcal{F}_\sigma$ and $t \in \mathbb{R}^+$, $A \cap [\tau \leq t] = (A \cap [\sigma \leq t]) \cap [\tau \leq t] \in \mathcal{F}_t$, since $A \cap [\sigma \leq t], [\tau \leq t] \in \mathcal{F}_t$.

(b) This is easy.

(c) For $t \in \mathbb{R}^+$, $[\sigma \vee \tau \leq t] = [\sigma \leq t] \cap [\tau \leq t] \in \mathcal{F}_t$, and $[\sigma \wedge \tau \leq t] = [\sigma \leq t] \cup [\tau \leq t] \in \mathcal{F}_t$.

(d) We have for $t \in \mathbb{R}^+$, that $[\sigma < t] = \bigcup_{n=1}^{\infty} [\sigma_n < t] \in \mathcal{F}_t$. So σ is a stopping time (here we have used (2) of Proposition(5.2)). Now if $A \in \bigcap_{n=1}^{\infty} \mathcal{F}_{\sigma_n}$ then for $t \in \mathbb{R}^+$, $A \cap [\sigma < t] = \bigcup_{n=1}^{\infty} (A \cap [\sigma_n < t]) \in \mathcal{F}_t$, and so $A \in \mathcal{F}_\sigma$ (using here (2) of Proposition(5.4)). The reverse inclusion is just a consequence of (a).

(e) For $t \in \mathbb{R}^+$, $[\sigma \leq t] = \bigcap_{n=1}^{\infty} [\sigma_n \leq t] \in \mathcal{F}_t$.

(f) For cases (i) and (ii), define σ_n by $\sigma_n(\omega) = \inf\{k2^{-n} \mid k \in \mathbb{N}, k2^{-n} \geq \sigma(\omega)\}$. For case (iii), let $\sigma_n(\omega) = \inf\{t \in D_n \mid t \geq \sigma(\omega)\}$, where $D_n = \{k2^{-n} \mid k = 0, \dots, 2^{2n}\} \cup \{\infty\}$.

(g) First suppose that σ is finite and σ -simple, and let $D \subseteq \mathbb{R}^+$ be the countable range of σ . Then if B is a Borel subset of $\mathbb{R}$, $[X_\sigma \in B] = \bigcup_{s \in D} ([X_s \in B] \cap [\sigma = s]) \in \mathcal{F}$. So for $t \in \mathbb{R}^+$, $[X_\sigma \in B] \cap [\sigma \leq t] = \bigcup_{s \in D: s \leq t} ([X_s \in B] \cap [\sigma = s]) \in \mathcal{F}_t$. Thus $[X_\sigma \in B] \in \mathcal{F}_\sigma$ as required. For a general σ . Let $(\sigma_n)_n$ be as in (f)(i). Then $X_{\sigma_n} \rightarrow X_\sigma$ pointwise. But X_{σ_n} is $\mathcal{F}_{\sigma_n}$ -measurable by the special case. Thus X_σ is $\mathcal{F}_\sigma$ -measurable, since $\mathcal{F}_\sigma = \bigcap_{n=1}^{\infty} \mathcal{F}_{\sigma_n}$ by (d). $\square$

(5.7) Lemma. If σ is a stopping time, then $\mathbf{1}_{(\sigma, \infty)} \in \mathcal{E}_{\text{inc}}^+$.

Proof: Let $A_n = \bigcup_{k=0}^{2^{2n}-1} ((k2^{-n}, (k+1)2^{-n}] \times [\sigma \leq k2^{-n}])$. Then $\mathbf{1}_{A_n} \in \mathcal{E}$ and $\mathbf{1}_{A_n} \uparrow \mathbf{1}_{(\sigma, \infty)}$. Thus $\mathbf{1}_{(\sigma, \infty)} \in \mathcal{E}_{\text{inc}}^+$ as required. $\square$

(5.8)Definition. If $A \subseteq \bar{\mathbb{R}}^+ \times \Omega$, then the *début* of A is the function $\sigma: \Omega \rightarrow \bar{\mathbb{R}}^+$ defined by $\sigma(\omega) = \inf\{t \in \bar{\mathbb{R}}^+ \mid (t, \omega) \in A\}$, where by convention $\inf \emptyset = \infty$. $\square$

The following gives a sufficient condition for the *début* of $[X \in B]$ to be a stopping time. In fact a more general theorem, known as the *Début Theorem*, can be proved. For a statement of this result see Rogers & Williams [RW] Theorem VI.3.4. For a proof of the *Début Theorem* see §IV.50 of Dellacherie & Meyer [DM]. However the following lemma will be sufficient for the requirements of this thesis.

(5.9)Lemma. Suppose that Γ is a topological space and that $X: \mathbb{R}^+ \times \Omega \rightarrow \Gamma$ is an adapted process (with respect to the Borel σ -field on Γ) such that each path of X is either left or right continuous. Then the *début* of $[X \in U]$ is a stopping time for each open subset U of Γ .

Proof: Let σ be the *début* of $[X \in U]$, and suppose that $t \in \mathbb{R}^+$. For $\omega \in \Omega$, $\sigma(\omega) < t$ if and only if $X_s(\omega) \in U$ for some $s \in [0, t)$. Now since the path $t \mapsto X_t(\omega)$ is left or right continuous, this holds if and only if $X_q(\omega) \in U$ for some $q \in [0, t) \cap \mathbb{Q}$. Therefore $[\sigma < t] = \bigcup_{q \in [0, t) \cap \mathbb{Q}} [X_q \in U] \in \mathcal{F}_t$, and so σ is a stopping time. $\square$

§5.2 PATHWISE CÀDLÀG VERSIONS OF INTEGRATORS

The main theorem that we shall give here says that given an integrator S , we can modify each random variable S_t on a null set to get a process that is (pathwise) càdlàg. This theorem is in Bichteler [Bi2] (Proposition 2.5). After we have proved this we shall think of all integrators as (pathwise) càdlàg process. Thus we will have immediately that $X \cdot S$ is a càdlàg process whenever S an integrator and $X \in \Lambda^1(S)$. Recall that the term ‘integrator’ means L^0 -integrator.

(5.10)Theorem. Suppose that S is an integrator. Then there exists a (pathwise) càdlàg process $R: \mathbb{R}^+ \times \Omega \rightarrow \mathbb{R}$ such that $S_t = R_t$ a.s. for all $t \in \mathbb{R}^+$.

Note.

1. The process R must be unique up to indistinguishability. R is known as a *càdlàg version* of S .

2. In view of this theorem and Theorem(4.30), we can see that an L^p -integrator is a càdlàg process S such that for each $t \in \mathbb{R}^+$, $\int_0^t \dots dS: \mathcal{E} \rightarrow L^p$ is continuous when $\mathcal{E}$ is given the supremum norm.

Proof: First we show that there exists a càdlàg process $R: \mathbb{R}^+ \times \Omega \rightarrow \bar{\mathbb{R}}$ such that $S_t = R_t$ a.s. for all $t \in \mathbb{R}^+$. We use the technique of upcrossings that is used in the proof of the martingale convergence theorems. For each $q \in \mathbb{Q}^+$, pick a random variable $\tilde{S}_q: \Omega \rightarrow \mathbb{R}$ such that $\tilde{S}_q = S_q$ a.s. Now for each $a, b \in \mathbb{Q}$ such that $a < b$, and $N \in \mathbb{N}_{>0}$, let $U_N^{(a,b)}$ denote the number of upcrossings of (a, b) by $(\tilde{S}_q)_{q \in [0, N] \cap \mathbb{Q}}$. I.e.

$$U_N^{(a,b)}(\omega) = \sup\{n \in \mathbb{N} \mid \text{there exist } s_1 \leq t_1 \leq \dots \leq s_n \leq t_n \text{ in } [0, N] \cap \mathbb{Q} \\ \text{such that } \tilde{S}_{s_k}(\omega) \leq a \text{ and } \tilde{S}_{t_k}(\omega) \geq b \text{ for each } k\}.$$

We shall now show that $U_N^{(a,b)}$ is measurable and is finite a.s.

So fix N, a and b , and let $F = [0, N] \cap \mathbb{Q}$. Take $F_1, F_2, \dots$ to be finite subsets of F such that $F_n \uparrow F$ and $0, N \in F_n$ for all n . Let U_n denote the number of upcrossings of (a, b) by $(\tilde{S}_q)_{q \in F_n}$. It is an easy exercise to show that U_n is measurable and bounded. Fix n for a moment and define stopping times $\sigma_1, \sigma_2, \dots, \tau_0, \tau_1, \tau_2, \dots$ by

$$\begin{aligned} \tau_0(\omega) &= 0, \\ \sigma_k(\omega) &= N \wedge \inf\{q \in F_n \mid q \geq \tau_{k-1}(\omega) \text{ and } \tilde{S}_q(\omega) \leq a\}, \\ \tau_k(\omega) &= N \wedge \inf\{q \in F_n \mid q \geq \sigma_k(\omega) \text{ and } \tilde{S}_q(\omega) \geq b\} \quad \text{for } k = 1, 2, \dots \end{aligned}$$

It is another easy exercise to check that these are in fact stopping times. Now if $r = U_n(\omega)$, the following are true.

- (i) $\sigma_1(\omega) < \tau_1(\omega) < \dots < \sigma_r(\omega) < \tau_r(\omega)$.
- (ii) $\tilde{S}_{\sigma_k}(\omega) \leq a$ and $\tilde{S}_{\tau_k}(\omega) \geq b$ for $k = 1, \dots, r$.
- (iii) $\tilde{S}_{\sigma_{r+1}}(\omega) \leq a$ or $\sigma_{r+1}(\omega) = N$.
- (iv) $\tau_{r+1}(\omega) = N$ and $\sigma_k(\omega) = \tau_k(\omega) = N$ for all $k > r + 1$.

These give that

$$(1) \quad U_n \leq \frac{1}{b-a} \left(\sum_{k=1}^{\infty} (\tilde{S}_{\tau_k} - \tilde{S}_{\sigma_k}) + (a - \tilde{S}_N)^+ \right),$$

(since $\tilde{S}_{\tau_k}(\omega) - \tilde{S}_{\sigma_k}(\omega) \geq b - a$ when $k \leq r = U_n(\omega)$, $\tilde{S}_{\tau_{r+1}}(\omega) - \tilde{S}_{\sigma_{r+1}}(\omega) \geq -(a - \tilde{S}_N(\omega))^+$, and $\tilde{S}_{\tau_k}(\omega) - \tilde{S}_{\sigma_k}(\omega) = 0$ when $k > r + 1$). The sum is really just a fixed finite sum, since U_n is bounded. Now put $X = \sum_{k=1}^{\infty} \mathbf{1}_{(\sigma_k, \tau_k]}$. So $X \in \mathcal{E}$ and (1) gives that

$$(2) \quad U_n \leq \frac{1}{b-a} \left(\int X dS + (a - \tilde{S}_N)^+ \right) \quad \text{a.s.}$$

But $|X| \leq \mathbf{1}_{(0, N] \times \Omega}$; so for $\lambda > 0$ we have that

$$(3) \quad \rho_0(\lambda U_n) \leq \Phi_S^0 \left(\frac{\lambda}{b-a} \mathbf{1}_{(0, N] \times \Omega} \right) + \rho_0 \left(\frac{\lambda}{b-a} (a - \tilde{S}_N)^+ \right).$$

Now we allow n to go to infinity. We have that $U_n \uparrow U_N^{(a,b)}$. Therefore $U_N^{(a,b)}$ is measurable, and from (3) we deduce that

$$(4) \quad \rho_0(\lambda U_N^{(a,b)}) \leq \Phi_S^0 \left(\frac{\lambda}{b-a} \mathbf{1}_{(0, N] \times \Omega} \right) + \rho_0 \left(\frac{\lambda}{b-a} (a - \tilde{S}_N)^+ \right).$$

So we see that $\rho_0(\lambda U_N^{(a,b)}) \rightarrow 0$ as $\lambda \downarrow 0$. Thus $U_N^{(a,b)} < \infty$ a.s.

Now let $E = \bigcup_{N=1}^{\infty} \bigcup_{a, b \in \mathbb{Q}: a < b} [U_N^{(a,b)} = \infty]$. Then E is $\mathbb{P}$ -null, and therefore by redefining each $\tilde{S}_q$ to be zero on E , we may assume that $E = \emptyset$. So $U_N^{(a,b)} < \infty$ everywhere for all N and all $a < b$ in $\mathbb{Q}$. Now for each $\omega \in \Omega$ and $t \in \mathbb{R}^+$, the

limits $\lim_{q \downarrow t} \tilde{S}_q(\omega)$ and $\lim_{q \uparrow t} \tilde{S}_q(\omega)$ exist in $\overline{\mathbb{R}}$. For if the first does not, there exist $a, b \in \mathbb{Q}$ such that $\liminf_{q \downarrow t} \tilde{S}_q(\omega) < a < b < \limsup_{q \downarrow t} \tilde{S}_q(\omega)$. So if we take $N > t$, we have that $U_N^{(a,b)}(\omega) = \infty$, which is a contradiction. The other limit is similar. Let $R: \mathbb{R}^+ \times \Omega \rightarrow \overline{\mathbb{R}}$ by $R_t(\omega) = \lim_{q \downarrow t} \tilde{S}_q(\omega)$. Then R is a càdlàg adapted process. $R_t = S_t$ a.s. for all t , since S is càdlàg as a function from $\mathbb{R}^+$ to L^0 (by Corollary(4.18)).

We have to show that $\overline{\mathbb{R}}$ can be replaced by $\mathbb{R}$. I.e. R_t and R_{t-} are real valued. Define stopping times $\sigma_1, \sigma_2, \dots$ by letting σ_n be the début of $[|R| > n]$ (use Proposition(5.9)). Now $\sigma_n \uparrow \sigma$ for some stopping time σ . We shall show that $\sigma = \infty$ a.s. Suppose that $N \in \mathbb{N}_{>0}$. Then I claim that

$$(5) \quad R_{\sigma_n \wedge N} = R_0 + \int \mathbf{1}_{(0, \sigma_n \wedge N]} dS \quad \text{a.s.}$$

To prove this take $\tau_1, \tau_2, \dots$ to be finite simple stopping times such that $\tau_k \downarrow \sigma_n \wedge N$ pointwise as $k \rightarrow \infty$ (use Proposition(5.6)(f)(ii)). Now $R_{\tau_k} = R_0 + \int \mathbf{1}_{(0, \tau_k]} dS$. Let $k \rightarrow \infty$, and use the Dominated Convergence Theorem (4.17) to give (5). Now if $\sigma_n(\omega) \leq N$, then we have that $|R_{\sigma_n(\omega)}(\omega)| \geq n$, since R is càdlàg. Therefore $|R_{\sigma_n \wedge N}(\omega)| \geq n$. Thus $\mathbf{1}_{[\sigma_n \leq N]} \leq \frac{1}{n} |R_{\sigma_n \wedge N}|$. Therefore applying ρ_0 and using (5) gives that

$$\begin{aligned} \mathbb{P}[\sigma_n \leq N] &\leq \rho_0\left(\frac{1}{n} R_0\right) + \rho_0\left(\frac{1}{n} \int \mathbf{1}_{(0, \sigma_n \wedge N]} dS\right) \\ &\leq \rho_0\left(\frac{1}{n} R_0\right) + \Phi_S^0\left(\frac{1}{n} \mathbf{1}_{(0, N] \times \Omega}\right) \\ &\rightarrow 0 \quad \text{as } n \rightarrow \infty. \end{aligned}$$

But $\mathbb{P}[\sigma \leq N] \leq \mathbb{P}[\sigma_n \leq N]$, and so it follows that $\mathbb{P}[\sigma \leq N] = 0$. Now $[\sigma < \infty] = \bigcup_{N=1}^{\infty} [\sigma \leq N]$, and thus $[\sigma < \infty]$ is $\mathbb{P}$ -null as promised. Modifying R on the set $\mathbb{R}^+ \times [\sigma < \infty]$ to be zero there, we may assume that $\sigma = \infty$ everywhere. So $\sigma_n \uparrow \infty$ everywhere. Now for $\omega \in \Omega$ and $t \in \mathbb{R}^+$, $\sigma_n(\omega) \uparrow \infty$. Take n such that $\sigma_n(\omega) > t$. Then $|R_s(\omega)| \leq n$ whenever $0 \leq s < \sigma_n(\omega)$. Therefore $|R_t(\omega)|, |R_{t-}(\omega)| \leq n < \infty$; thus showing that R_t and R_{t-} are real valued as required. $\square$

We can now prove a result of which (5) in the above proof is a special case.

(5.11) Proposition. Suppose that S is an integrator and that σ is a finite stopping time. Then $\mathbf{1}_{(0,\sigma]} \in \mathcal{L}^1(S)$ and $\int \mathbf{1}_{(0,\sigma]} dS = S_\sigma - S_0$ (remember that S is now considered as a pathwise càdlàg process).

Proof: Take $\sigma_1, \sigma_2, \dots$ to be finite σ -simple stopping times such that $\sigma_n \downarrow \sigma$ (use Proposition(5.6)(f)(i)). Then supposing the proposition is true for finite σ -simple stopping times, we have that $\mathbf{1}_{(0,\sigma_n]} \in \mathcal{L}^1(S)$ and that $\int \mathbf{1}_{(0,\sigma_n]} dS = S_{\sigma_n} - S_0$ for all n . Now $|\mathbf{1}_{(0,\sigma_n]}| \leq \mathbf{1}_{(0,\sigma_1]}$ for all n . So we may apply the Dominated Convergence Theorem (4.17), and let $n \rightarrow \infty$, to give the required result. Thus without loss of generality, we may assume that σ is a finite σ -simple stopping time. In fact from the proof of Proposition(5.6)(f)(i), we may assume that σ has its range contained in the set $\{k2^{-n} \mid k \in \mathbb{N}\}$ for some n .

We shall now assume that these conditions hold. We have that $\mathbf{1}_{(0,\sigma \wedge N]} \in \mathcal{E}$, and $\int \mathbf{1}_{(0,\sigma \wedge N]} dS = S_{\sigma \wedge N} - S_0$. So letting $N \rightarrow \infty$, and using the Monotone Convergence Theorem (4.16), we see that it is enough to show that $\mathbf{1}_{(0,\sigma]}$ is Φ_S^0 -finite. Now suppose that $\varepsilon > 0$ is given, and then take N such that $\mathbb{P}[\sigma > N] < \frac{\varepsilon}{2}$. Let $X = \mathbf{1}_{(\sigma \wedge N, \sigma]}$. Then since $\mathbf{1}_{(\sigma \wedge N, \sigma \wedge m]}$ is an element of $\mathcal{E}^+$, and increases to X as $m \rightarrow \infty$, we have that $X \in \mathcal{E}_{\text{inc}}^+$. Now $[X \neq 0] \subseteq \mathbb{R}^{++} \times [\sigma > N]$, and so it follows by Lemma(4.25), that $\Phi_S^0(\lambda X) \leq \mathbb{P}[\sigma > N] < \frac{\varepsilon}{2}$ for all $\lambda > 0$. Now writing $\lambda \mathbf{1}_{(0,\sigma]} = \lambda X + \lambda \mathbf{1}_{(0,\sigma \wedge N]}$, and applying Φ_S^0 gives that

$$\begin{aligned} \Phi_S^0(\lambda \mathbf{1}_{(0,\sigma]}) &\leq \Phi_S^0(\lambda X) + \Phi_S^0(\lambda \mathbf{1}_{(0,\sigma \wedge N]}) \\ &< \frac{\varepsilon}{2} + \Phi_S^0(\lambda \mathbf{1}_{(0,N] \times \Omega}) \\ &< \varepsilon, \end{aligned}$$

when $\lambda > 0$ is sufficiently small. I.e. $\Phi_S^0(\lambda \mathbf{1}_{(0,\sigma]}) \rightarrow 0$ as $\lambda \downarrow 0$ as required. $\square$

(5.12) Corollary. Suppose that S is an integrator, $X \in \Lambda^1(S)$ and that σ is a finite stopping time. Then $\mathbf{1}_{(0,\sigma]} X \in \mathcal{L}^1(S)$ and $\int \mathbf{1}_{(0,\sigma]} X dS = (X \cdot S)_\sigma$.

Proof: Replace S by $X \cdot S$ in Proposition(5.11) and apply Theorem(4.21). $\square$

We can now prove two more results of a similar nature.

(5.13)Proposition. Suppose that S is an integrator, $\sigma \leq \tau$ are finite stopping times, and that C is an $\mathcal{F}_\sigma$ -measurable random variable. Then $C\mathbf{1}_{(\sigma, \tau]} \in \mathcal{L}^1(S)$ and $\int C\mathbf{1}_{(\sigma, \tau]} dS = C(S_\tau - S_\sigma)$.

Here of course $(C\mathbf{1}_{(\sigma, \tau]})(t, \omega) = C(\omega)\mathbf{1}_{(\sigma, \tau]}(t, \omega)$.

Proof: Let $X = C\mathbf{1}_{(\sigma, \tau]}$. Suppose first that $C = 1$. Then $X = \mathbf{1}_{(\sigma, \tau]} = \mathbf{1}_{(0, \tau]} - \mathbf{1}_{(0, \sigma]} \in \mathcal{L}^1(S)$, by Proposition(5.11). Also by the same proposition, $\int X dS = S_\tau - S_\sigma$.

Now suppose $C = \mathbf{1}_F$, with $F \in \mathcal{F}_\sigma$. Then let $\tilde{\sigma}$ be the stopping time that is equal to σ on F and equal to τ outside F . Then the above case applied with σ replaced by $\tilde{\sigma}$ gives that $\int X dS = S_\tau - S_{\tilde{\sigma}} = \mathbf{1}_F(S_\tau - S_\sigma)$ as required. So by linearity the case when C is a simple $\mathcal{F}_\sigma$ -measurable random variable is proven.

Now suppose only that $C \geq 0$. Then take simple $\mathcal{F}_\sigma$ -measurable random variables $C_1, C_2, \dots \geq 0$ such that $C_n \uparrow C$. Let $X_n = C_n\mathbf{1}_{(\sigma, \tau]}$. So $X_n \uparrow X$. Now since X is a l  dc  g process, we have by Theorem(4.24)(i), that $X \in \Lambda^1(S)$. So by Corollary(5.12), $X = \mathbf{1}_{(0, \tau]}X \in \mathcal{L}^1(S)$. Now by the previous case $\int X_n dS = C_n(S_\tau - S_\sigma)$. Let $n \rightarrow \infty$, using the Dominated Convergence Theorem (4.17) on the left hand side (with X as the dominating function), to give the result in this case. For the most general case, write $C = C^+ - C^-$ and use linearity. $\square$

This has the following corollary which is a generalization of Proposition(4.27).

(5.14)Corollary. Suppose that S is a integrator and that $X = \sum_{k=0}^{r-1} C_k\mathbf{1}_{(\sigma_k, \sigma_{k+1}]}$, where $0 = \sigma_0 \leq \dots \leq \sigma_r < \infty$ are stopping times, and for $k = 0, \dots, r-1$, C_k is a $\mathcal{F}_{\sigma_k}$ -measurable random variable. Then $X \in \mathcal{L}^1(S)$ and $\int X dS = \sum_{k=0}^{r-1} C_k(S_{\sigma_{k+1}} - S_{\sigma_k})$.

Proof: Just apply Proposition(5.13) to each term of the sum. $\square$

(5.15) Definition.

- (a) If $X: \mathbb{R}^+ \times \Omega \rightarrow \mathbb{R}$ is a càdlàg process and σ is a stopping time, then X^σ will denote the càdlàg process given by $X_t^\sigma = X_{\sigma \wedge t} - X_0$.
- (b) A family $\mathcal{X}$ of càdlàg processes is said to be *stable under stopping* if and only if the following properties hold.
 - (1) If X and Y are indistinguishable càdlàg processes and $X \in \mathcal{X}$, then $Y \in \mathcal{X}$.
 - (2) If $X \in \mathcal{X}$ and σ is a stopping time, then $X^\sigma \in \mathcal{X}$.

Property (1) says that $\mathcal{X}$ can be thought of as a family of equivalence classes of càdlàg processes under the equivalence relation of indistinguishability (cf. Note 2 after Definition(4.2)).

- (c) Suppose that $\mathcal{X}$ is a family of càdlàg processes that is stable under stopping. Then a càdlàg process X is said to be a *local $\mathcal{X}$ -process* if and only if there exist stopping times $\sigma_1, \sigma_2, \dots$ such that $\sigma_n \uparrow \infty$ and $X^{\sigma_n} \in \mathcal{X}$ for all n . The sequence $(\sigma_n)_n$ is known as a *reducing sequence* of X into $\mathcal{X}$. The family of local $\mathcal{X}$ -processes will be denoted by $\mathcal{X}_{\text{loc}}$. Clearly $\mathcal{X} \subseteq \mathcal{X}_{\text{loc}}$. When the elements of $\mathcal{X}$ are described by a name (e.g. ‘martingale’), then the local $\mathcal{X}$ -processes will be described by prefixing that name by the word ‘local’. □

Note. Suppose that $\mathcal{X}$ is a family of càdlàg processes that is stable under stopping, X is a local $\mathcal{X}$ -process with $(\sigma_n)_n$ a reducing sequence of X into $\mathcal{X}$, and $(\tau_n)_n$ is any sequence of stopping times such that $\tau_n \uparrow \infty$. Then $(\sigma_n \wedge \tau_n)_n$ is also a reducing sequence of X into $\mathcal{X}$ (since $X^{\sigma_n \wedge \tau_n} = (X^{\sigma_n})^{\tau_n}$). Thus we see that any finite collection of local $\mathcal{X}$ -processes has a common reducing sequence (take the termwise minimum). This is useful when proving certain facts about $\mathcal{X}_{\text{loc}}$. For example the result that if $\mathcal{X}$ is closed under pointwise addition, then so is $\mathcal{X}_{\text{loc}}$.

The following result says that using localization twice does not give any more processes than those obtained by one application.

(5.16) Proposition. *Suppose that $\mathcal{X}$ is a family of càdlàg processes that is stable under stopping. Then $(\mathcal{X}_{\text{loc}})_{\text{loc}} = \mathcal{X}_{\text{loc}}$.*

Proof: Suppose that $X \in (\mathcal{X}_{\text{loc}})_{\text{loc}}$ and suppose that $(\sigma_n)_n$ is a reducing sequence of X into $\mathcal{X}_{\text{loc}}$. Then $X^{\sigma_n} \in \mathcal{X}_{\text{loc}}$ for all n . For each n , let $(\tau_{n,k})_k$ be a reducing sequence of X^{σ_n} into $\mathcal{X}$. Take k_n such that $\mathbb{P}[\tau_{n,k_n} < \sigma_n - 2^{-n}] < 2^{-n}$, and let $\alpha_n = \inf_{r \geq n} (\sigma_r \wedge \tau_{r,k_r})$. So α_n is a stopping time (since $[\alpha_n < t] = \bigcup_{r=n}^{\infty} [\sigma_r \wedge \tau_{r,k_r} < t] \in \mathcal{F}_t$ for all t). But also $X^{\alpha_n} = ((X^{\sigma_n})^{\tau_{n,k_n}})^{\alpha_n} \in \mathcal{X}$. Now α_n increases with n . We have that

$$[\alpha_n < \sigma_n - 2^{-n}] \subseteq \bigcup_{r=n}^{\infty} [\tau_{r,k_r} < \sigma_r - 2^{-r}],$$

and so taking probability of both sides gives that $\mathbb{P}[\alpha_n < \sigma_n - 2^{-n}] < 2^{-n+1}$. Let $E = \bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} [\alpha_n < \sigma_n - 2^{-n}]$. Then E is $\mathbb{P}$ -null and $\alpha_n \uparrow \infty$ outside E . Thus adjusting α_n to be equal to ∞ on E gives a reducing sequence of X into $\mathcal{X}$ (X^{α_n} is changed to a process indistinguishable from it). Thus $X \in \mathcal{X}_{\text{loc}}$ as required. $\square$

The concept of localization is used in the Strasbourg approach to the Itô calculus. In that approach integration is first done with respect to L^2 -martingales and then one extends to local L^2 -martingales using the formula $(X \cdot S)^{\sigma} = X \cdot (S^{\sigma})$. We ask the question can we increase our class of integrators by considering local integrators? The answer as we shall see is no. But first is the concept meaningful?

(5.17) Proposition. *The family of all integrators is stable under stopping. In fact if S is an integrator, σ is a stopping time and $X \in \Lambda^1(S)$, then S^{σ} is an integrator, $X \in \Lambda^1(S^{\sigma})$ and $(X \cdot S)^{\sigma} = X \cdot (S^{\sigma})$. Also we have that $S^{\sigma} = \mathbf{1}_{(0,\sigma]} \cdot S$.*

Proof: We have that $S^{\sigma} = \mathbf{1}_{(0,\sigma]} \cdot S$ by applying Proposition(5.11) to the stopping time $\sigma \wedge t$, which is finite (actually bounded) even though σ may not be a finite

stopping time. Thus by Theorem(4.21)(i), we obtain that S^σ is an integrator. Given $X \in \Lambda^1(S)$, we have that $\mathbf{1}_{(0,\sigma]}X \in \Lambda^1(S)$. So by Theorem(4.21) parts (iv) and (v), we obtain that $X \in \Lambda^1(S^\sigma)$ and that $X \cdot (S^\sigma) = (X\mathbf{1}_{(0,\sigma]}) \cdot S = \mathbf{1}_{(0,\sigma]} \cdot (X \cdot S) = (X \cdot S)^\sigma$ as required. $\square$

Now we can show that localization introduces no new integrators.

(5.18)Theorem. *If $S: \mathbb{R}^+ \times \Omega \rightarrow \mathbb{R}$ is a local integrator, then S is an integrator.*

Proof: Let $(\sigma_n)_n$ be a reducing sequence of S into the class of all integrators. We have to check that $(\dagger)$ of Definition(4.6) holds. So suppose that X_n and Y are as in this axiom, and suppose that $\lambda > 0$ and $\varepsilon > 0$. Take $b \in \mathbb{R}^+$ such that $[Y \neq 0] \subseteq (0, b] \times \Omega$, and take N such that $\mathbb{P}[\sigma_N < b] < \frac{\varepsilon}{2}$. Now S^{σ_N} is an integrator. So $\int X_n dS^{\sigma_N} \rightarrow 0$ in probability. Therefore there exists an n_0 such that $\mathbb{P}[|\int X_n dS^{\sigma_N}| > \lambda] < \frac{\varepsilon}{2}$ for all $n \geq n_0$. Now $\int X_n dS = \int X_n dS^{\sigma_N}$ on the set $[\sigma_N \geq b]$. Therefore

$$[|\int X_n dS| > \lambda] \subseteq [|\int X_n dS^{\sigma_N}| > \lambda] \cup [\sigma_N < b].$$

Apply $\mathbb{P}$ to give that $\mathbb{P}[|\int X_n dS| > \lambda] < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$ for all $n \geq n_0$. So $\int X_n dS \rightarrow 0$ in probability as required. $\square$

There is another form of localization which is used in the Strasbourg approach to the Itô calculus in order to increase the class of integrands of the stochastic integral. We shall see (Theorem(5.21)) that it does not generalize our stochastic integral. First we shall define this concept.

(5.19)Definition. Suppose that $\mathcal{X}$ is a family of previsible processes such that

$$(\dagger) \quad X \in \mathcal{X}, \sigma \text{ is a stopping time} \implies \mathbf{1}_{(0,\sigma]}X \in \mathcal{X}.$$

Then we shall say that a previsible process X is *locally in $\mathcal{X}$* if and only if there exist stopping times $\sigma_1, \sigma_2, \dots$ such that $\sigma_n \uparrow \infty$ and $\mathbf{1}_{(0, \sigma_n]} X \in \mathcal{X}$ for all n . The set of all previsible processes locally in $\mathcal{X}$ is denoted by $\mathbb{L}\mathcal{X}$. $(\sigma_n)_n$ is known as a *reducing sequence* of X into $\mathcal{X}$. When the elements of $\mathcal{X}$ are described by an adjective (e.g. ‘bounded’), the elements of $\mathbb{L}\mathcal{X}$ will be described by the result of prefixing the adjective by the word ‘locally’. Thus we see for example that a previsible process X is *locally bounded* if and only if there exist stopping times $\sigma_1, \sigma_2, \dots$ such that $\sigma_n \uparrow \infty$ and X is bounded as a function on $(0, \sigma_n]$ for all n . $\square$

The analogous remarks to those given in the note after Definition(5.15) hold here. Also we have the following analogue to Proposition(5.16).

(5.20)Proposition. *Suppose that $\mathcal{X}$ is a family of previsible processes satisfying $(\dagger)$ of Definition(5.19) and that either,*

- (1) $X \in \mathcal{X}$ whenever X is a previsible process indistinguishable from an element of $\mathcal{X}$;

or,

- (2) $\mathbf{1}_{(0, \sigma \vee \tau]} X \in \mathcal{X}$ whenever X is a previsible process, and σ and τ are two stopping times such that $\mathbf{1}_{(0, \sigma]} X \in \mathcal{X}$ and $\mathbf{1}_{(0, \tau]} X \in \mathcal{X}$.

Then $\mathbb{L}\mathcal{X} = \mathcal{X}$.

Proof: If (1) holds then essentially the same proof as for Proposition(5.16) works here; so suppose that (2) holds and suppose that $X \in \mathbb{L}\mathcal{X}$. Then let $(\sigma_n)_n$ be a reducing sequence of X into $\mathcal{X}$. Then for each n , let $(\tau_{n,k})_k$ be a reducing sequence of $\mathbf{1}_{(0, \sigma_n]} X$ into $\mathcal{X}$. We have that $\mathbf{1}_{(0, \sigma_n \wedge \tau_{n,k}]} X \in \mathcal{X}$. Let $\alpha_k = \bigvee_{n=1}^k (\sigma_n \wedge \tau_{n,k})$. Then $\mathbf{1}_{(0, \alpha_k]} X \in \mathcal{X}$ for all k (by (2)), and $\alpha_k \uparrow \infty$ as $k \rightarrow \infty$. Thus $X \in \mathcal{X}$ as required. $\square$

(5.21)Theorem. *If S is an integrator, then $\mathbb{L}\Lambda^1(S) = \Lambda^1(S)$ ($\Lambda^1(S)$ clearly satisfies $(\dagger)$ of Definition(5.19)).*

Proof: Suppose that $X \in \Lambda^1(S)$ and let $(\sigma_n)_n$ be a reducing sequence of X into $\Lambda^1(S)$. Fix $b \in \mathbb{R}^+$. We have to show that $\mathbf{1}_{(0,b] \times \Omega} X \in \mathcal{L}^1(S)$. Given $\varepsilon > 0$, take N such that $\mathbb{P}[\sigma_N < b] < \frac{\varepsilon}{2}$. Define a stopping time τ by

$$\tau(\omega) = \begin{cases} \sigma_N(\omega), & \text{if } \sigma_N(\omega) < b, \\ \infty, & \text{otherwise.} \end{cases}$$

So $[\tau < \infty] = [\sigma_N < b]$, and therefore $\mathbb{P}[\tau < \infty] < \frac{\varepsilon}{2}$. Now by Lemma(5.7), $\mathbf{1}_{(\tau, \infty)} \in \mathcal{E}_{\text{inc}}^+$, and therefore $\infty \mathbf{1}_{(\tau, \infty)} \in \mathcal{E}_{\text{inc}}^+$. So by Lemma(4.25), $\Phi_S^0(\infty \mathbf{1}_{(\tau, \infty)}) \leq \mathbb{P}[\tau < \infty] < \frac{\varepsilon}{2}$. Therefore

$$\begin{aligned} \Phi_S^0(\lambda \mathbf{1}_{(0,b] \times \Omega} X) &\leq \Phi_S^0(\lambda \mathbf{1}_{(0, \sigma_N \wedge b]} X) + \Phi_S^0(\lambda \mathbf{1}_{(\sigma_N \wedge b, b]} X) \\ &\leq \Phi_S^0(\lambda \mathbf{1}_{(0, \sigma_N \wedge b]} X) + \Phi_S^0(\infty \mathbf{1}_{(\tau, \infty)}) \\ &\quad (\text{since } \lambda \mathbf{1}_{(\sigma_N \wedge b, b]} X \leq \infty \mathbf{1}_{(\tau, \infty)}) \\ &\leq \Phi_S^0(\lambda \mathbf{1}_{(0, \sigma_N \wedge b]} X) + \frac{\varepsilon}{2} \\ &< \varepsilon, \end{aligned}$$

for $\lambda > 0$ sufficiently small, since $\mathbf{1}_{(0, \sigma_N \wedge b]} X = \mathbf{1}_{(0,b] \times \Omega} \mathbf{1}_{(0, \sigma_N]} X \in \mathcal{L}^1(S)$ ($\mathbf{1}_{(0, \sigma_N]} X \in \Lambda^1(S)$). So $\mathbf{1}_{(0,b] \times \Omega} X$ is Φ_S^0 -finite and therefore is an element of $\mathcal{L}^1(S)$ as required. $\square$

Recall that in Definition(5.19) we defined the concept of a ‘locally bounded previsible process’. This concept is useful since such processes are always in $\Lambda^1(S)$, and all $\text{l}\grave{\text{a}}\text{d}\text{c}\grave{\text{a}}\text{g}$ processes are locally bounded. This is the content of the next proposition.

(5.22) Proposition.

- (i) If $X: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}$ is any locally bounded previsible process, then $X \in \Lambda^1(S)$ for all integrators S .
- (ii) If $X: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}$ is a $\text{l}\grave{\text{a}}\text{d}\text{c}\grave{\text{a}}\text{g}$ process, then X is a locally bounded previsible process.

Note. We see that by combining (i) and (ii), we obtain an alternative proof of Theorem(4.24)(i).

Proof:

(i) Suppose that S is an integrator. Then any bounded previsible process is in $\Lambda^1(S)$. Thus X is locally in $\Lambda^1(S)$, and so $X \in \Lambda^1(S)$ by Theorem(5.21).

(ii) We have seen already that X is previsible (Proposition(4.3)(i)), so it remains to show that X is locally bounded. Let σ_n be the début of $[|X| > n]$. Then σ_n is a stopping time and by construction $|X| \leq n$ on $(0, \sigma_n)$. But X has left continuous paths. Therefore $|X| \leq n$ on $(0, \sigma_n]$. Clearly $\sigma_n \uparrow \infty$, and so we are done. $\square$

§5.4 CONTINUOUS INTEGRATORS AND A PATHWISE CONVERGENCE RESULT

An important subclass of the class of càdlàg processes is the class of continuous processes. Thus it is important to know whether $X \cdot S$ is continuous when S is. We shall prove that it is.

(5.23)Theorem. Suppose that S is an integrator, $X \in \Lambda^1(S)$ and that $b \in \mathbb{R}^+$.

(i) For each $\lambda \in (0, 1]$, $\mathbb{P}[(X \cdot S)_b^* > \lambda] \leq \frac{1}{\lambda} \Phi_S^0(\mathbf{1}_{(0,b] \times \Omega} X)$.

(ii) $\rho_0((X \cdot S)_b^*) \leq 2 \Phi_S^0(\mathbf{1}_{(0,b] \times \Omega} X)^{1/2}$.

Proof:

(i) Let $\tilde{\sigma}$ be the début of $[|X \cdot S| > \lambda]$, and let $\sigma = \tilde{\sigma} \wedge b$. Now on the set $[(X \cdot S)_b^* > \lambda]$, we have that $|(X \cdot S)_\sigma| \geq \lambda$. So $\mathbf{1}_{[(X \cdot S)_b^* > \lambda]} \leq \frac{1}{\lambda} (1 \wedge |(X \cdot S)_\sigma|)$. Now take expectations to give that $\mathbb{P}[(X \cdot S)_b^* > \lambda] \leq \frac{1}{\lambda} \rho_0((X \cdot S)_\sigma)$. But $(X \cdot S)_\sigma = \int \mathbf{1}_{(0,\sigma]} X dS$. So $\mathbb{P}[(X \cdot S)_b^* > \lambda] \leq \frac{1}{\lambda} \Phi_S^0(\mathbf{1}_{(0,\sigma]} X)$. From this (i) clearly follows.

(ii) Let $c = \Phi_S^0(\mathbf{1}_{(0,b] \times \Omega} X)^{1/2}$. If $c = 0$ then $X \cdot S = 0$ on $(0, b] \times \Omega$, and so the result is trivial. So suppose that $c > 0$. Then from the definitions of ρ_0 and Φ_S^0 we

see that $c \in (0, 1]$. So by (i) with $\lambda = c$ we obtain that $\mathbb{P}[(X \cdot S)_b^* > c] \leq c$. Thus $\rho_0((X \cdot S)_b^*) = \int_{[(X \cdot S)_b^* \leq c]} 1 \wedge (X \cdot S)_b^* d\mathbb{P} + \int_{[(X \cdot S)_b^* > c]} 1 \wedge (X \cdot S)_b^* d\mathbb{P} \leq c + c = 2c$ as required. $\square$

(5.24)Notation. Suppose that $X: \mathbb{R}^+ \times \Omega \rightarrow \mathbb{R}$ is a càdlàg process. Then we define $\Delta X: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}$ by $\Delta X = X - X_-$. I.e. so that $(\Delta X)_t = X_t - X_{t-}$. ΔX is an adapted process. $\square$

(5.25)Theorem. Suppose that S is an integrator and that $X \in \Lambda^1(S)$. Then $\Delta(X \cdot S) = X \Delta S$ (where the equality is that of indistinguishability).

Proof: Let $\mathcal{G}$ be the set of all the $X \in \Lambda^1(S)$ such that $\Delta(X \cdot S) = X \Delta S$. Now I claim that the following holds.

(†) If $X_1, X_2, \dots \in \mathcal{G}$ are such that $X_n \rightarrow X$ pointwise and $|X_n| \leq Y$ for all n , for some $Y \in \Lambda^1(S)$, then $X \in \mathcal{G}$.

To see this, note that $\Phi_S^0(\mathbf{1}_{(0,b] \times \Omega}(X_n - X)) \rightarrow 0$ for all $b \in \mathbb{R}^+$ by the Dominated Convergence Theorem (4.17). Fix $b \in \mathbb{R}^+$. Then $(X_n \cdot S - X \cdot S)_b^* \rightarrow 0$ in probability as $n \rightarrow \infty$ by Theorem(5.23). Let $n_1 < n_2 < \dots$ be such that $(X_{n_k} \cdot S - X \cdot S)_b^* \rightarrow 0$ a.s. as $k \rightarrow \infty$, and let E be a $\mathbb{P}$ -null subset of Ω such that $(X_{n_k} \cdot S - X \cdot S)_b^* \rightarrow 0$ everywhere on $\Omega \setminus E$. We have $\Delta(X_{n_k} \cdot S) = X_{n_k} \Delta S$ on $(0, b] \times (\Omega \setminus E)$. Let $k \rightarrow \infty$ and use that each path converges uniformly on $[0, b]$ to give that $\Delta(X \cdot S) = X \Delta S$ on $(0, b] \times (\Omega \setminus E)$. Now let b run over all of $\mathbb{N}$ to give that $X \in \mathcal{G}$.

Now let $\mathcal{C}$ be the set of all $A \in \mathcal{P}$ such that $\mathbf{1}_A \in \mathcal{G}$. Then $\mathcal{C}$ is a monotone class by (†). It also contains the ring of subsets generated by the set of all previsible rectangles (it is easy to check that $\mathcal{E} \subseteq \mathcal{G}$). Therefore by a monotone class argument $\mathcal{C} = \mathcal{P}$. Thus $\mathcal{G}$ contains all the simple previsible processes. If $X \in \Lambda^1(S)$ with $X \geq 0$ then there exist simple non-negative previsible processes $X_1, X_2, \dots$ such that $X_n \uparrow X$. (†) (with $Y = X$) then shows that $X \in \mathcal{G}$. For a general $X \in \Lambda^1(S)$, $X = X^+ - X^- \in \mathcal{G}$ as required. $\square$

Now since a càdlàg process X is a continuous process if and only if $\Delta X = 0$, we obtain immediately the following corollary.

(5.26)Corollary. If S is a continuous integrator and $X \in \Lambda^1(S)$, then $X \cdot S$ is a continuous process. $\square$

§5.5 INCREASING AND FINITE-VARIATION PROCESSES

In this section we will consider the càdlàg processes A for which the integral $\int X dA$ can be defined pathwise. Three questions arise. Firstly are such processes

integrators? Secondly to what extent do the two sets of processes integrable with respect to A agree? Thirdly do the two possible definitions of $\int X dA$ agree? We shall give some positive answers to these. First we define the class of processes considered.

(5.27) Definition.

- (a) A function $\alpha: \mathbb{R}^+ \rightarrow \mathbb{R}$ is said to have *finite variation* if and only if $V_\alpha(t) < \infty$ for all $t \in \mathbb{R}^+$, where $V_\alpha(t)$ is defined by

$$V_\alpha(t) = \sup \left\{ \sum_{k=0}^{r-1} |\alpha(t_{k+1}) - \alpha(t_k)| \mid r \in \mathbb{N}, \quad 0 = t_0 < \dots < t_r = t \right\}$$

V_α is known as the *variation* of α .

- (b) A càdlàg process is said to have *finite variation* if and only if all its paths have finite variation. If A is such a process, then we define V_A pathwise. We shall use the term '*finite-variation process*' to mean a càdlàg process with finite variation.
- (c) An *increasing process* is a càdlàg process such that all its paths are increasing.

□

Now suppose that $\alpha: \mathbb{R}^+ \rightarrow \mathbb{R}$ is a function of finite variation. Then V_α is an increasing function $\mathbb{R}^+ \rightarrow \mathbb{R}^+$ and $V_\alpha(0) = 0$. It is standard analysis that V_α is càdlàg (resp. continuous) if and only if α is càdlàg (resp. continuous). We let

$$\alpha^{(+)} = \frac{1}{2}(V_\alpha + \alpha), \quad \text{and} \quad \alpha^{(-)} = \frac{1}{2}(V_\alpha - \alpha).$$

So $\alpha = \alpha^{(+)} - \alpha^{(-)}$, $V_\alpha = \alpha^{(+)} + \alpha^{(-)}$, and $\alpha^{(+)}$ and $\alpha^{(-)}$ are increasing functions. When A is a process of finite variation $A^{(+)}$ and $A^{(-)}$ are defined pathwise. It is nice to know that the following holds.

(5.28) Proposition. *If A is a finite-variation process, then V_A is a càdlàg adapted process, and hence so are $A^{(+)}$ and $A^{(-)}$.*

Proof: The only non trivial thing to check is that V_A is adapted. This clearly follows if we show that

$$(\dagger) \quad V_\alpha(t) = \lim_{n \rightarrow \infty} \sum_{k=0}^{\infty} |\alpha(t \wedge (k+1)2^{-n}) - \alpha(t \wedge k2^{-n})|,$$

for all càdlàg functions $\alpha: \mathbb{R}^+ \rightarrow \mathbb{R}$ of finite variation (apply $(\dagger)$ pathwise to A). To prove $(\dagger)$ observe first that the right hand side is evidently no bigger than the left hand side. For the reverse inequality, suppose that we are given a partition $0 = t_0 < \dots < t_r = t$. Then increase $t_0, \dots, t_{r-1}$ by small quantities to give another partition $0 = s_0 < \dots < s_r = t$ with $s_1, \dots, s_{r-1}$ dyadic rationals, and such that $\sum_{k=0}^{r-1} |\alpha(s_{k+1}) - \alpha(s_k)|$ differs by less than a given $\varepsilon > 0$ from $\sum_{k=0}^{r-1} |\alpha(t_{k+1}) - \alpha(t_k)|$ (use the right continuity of α). Now the right hand side of $(\dagger)$ is greater than or equal to $\sum_{k=0}^{r-1} |\alpha(s_{k+1}) - \alpha(s_k)|$, and hence, since $\varepsilon > 0$ was arbitrary, it is also greater than or equal to $\sum_{k=0}^{r-1} |\alpha(t_{k+1}) - \alpha(t_k)|$ as required. $\square$

We are now able to complete the definitions.

(5.29) Definition. Suppose that $0 \leq p < \infty$. Then a càdlàg process A is said to have L^p -variation if and only if A has finite variation, $A_0 \in L^p$ and V_A is an L^p -process. We shall use the term ‘ L^p -variation process’ to mean a càdlàg process with L^p -variation. Also the term ‘integrable variation’ will be taken to mean L^1 -variation. A càdlàg process with *bounded variation* or a *bounded-variation process* is a finite-variation process A such that V_A is bounded as a function on $\mathbb{R}^+ \times \Omega$. It is clear that for each L^p -variation process A , A is an L^p -process, and $A^{(+)}$ and $A^{(-)}$ are increasing L^p -processes. $\square$

Warning. Some authors (e.g. Rogers & Williams [RW]) take the term ‘ L^p -variation’ to mean that $(V_A)_\infty = \lim_{t \rightarrow \infty} (V_A)_t$ is an element of L^p . But it is more convenient for us to use the term as above.

Suppose that $\alpha: \mathbb{R}^+ \rightarrow \mathbb{R}$ is a càdlàg function with finite variation. Then V_α , $\alpha^{(+)}$ and $\alpha^{(-)}$ are càdlàg increasing functions. We have Lebesgue-Stieltjes measures on $\mathbb{R}^{++}$ associated with each of these. Now for $f: \mathbb{R}^{++} \rightarrow \mathbb{R}$ Borel such that $\int_0^\infty |f| dV_\alpha < \infty$, we can and do define

$$\int_0^\infty f d\alpha = \int_0^\infty f d\alpha^{(+)} - \int_0^\infty f d\alpha^{(-)}.$$

This is the Lebesgue-Stieltjes integral. This we can use to define $\int_0^\infty X dA$ pathwise for A a process of finite variation. The following theorem shows that this is consistent with the stochastic integral.

(5.30) Theorem. *Let $\int \dots dA$ denote the stochastic integral and $\int_0^\infty \dots dA$ denote the Stieltjes integral evaluated pathwise. Suppose that $0 \leq p < \infty$.*

(a) *If A is an increasing L^p càdlàg process, then the following hold.*

- (i) *A is an L^p -integrator.*
- (ii) *$\mathcal{L}^1(A, L^p)$ consists of the previsible processes $X: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}$ such that $\int_0^\infty |X| dA$ is finite a.s. and is an element of L^p .*
- (iii) *If $X \in \mathcal{L}^1(A, L^p)$, then*

$$\int X dA = \int_0^\infty X dA \text{ a.s. and } \Phi_A^p(X) = \rho_p \left(\int_0^\infty |X| dA \right).$$

(b) *If A is an L^p -variation process, then the following hold.*

- (i) *A is an L^p -integrator and $\Phi_A^p \leq \Phi_{V_A}^p$.*
- (ii) *$\mathcal{L}^1(V_A, L^p) \subseteq \mathcal{L}^1(A, L^p)$. (Note that by (a)(ii), $\mathcal{L}^1(V_A, L^p)$ is characterized by the pathwise Stieltjes integral with respect to V_A .)*
- (iii) *If $X \in \mathcal{L}^1(V_A, L^p)$, then*

$$\int X dA = \int_0^\infty X dA \text{ a.s. and } \Phi_A^p(X) \leq \rho_p \left(\int_0^\infty |X| dV_A \right).$$

Note. When $p = 0$ part (a) says that for A an increasing process, the stochastic integral $\int \dots dA$ and the pathwise Stieltjes integral $\int_0^\infty \dots dA$ have the same domain, and agree on that domain. Part (b) says that when A is a finite-variation process, the stochastic integral has a domain bigger than or equal to that of the pathwise integral, and they agree on the smaller domain.

Proof:

(a) We have by inspection that

$$(1) \quad \int X dA = \int_0^\infty X dA \quad \text{for all } X \in \mathcal{E}.$$

For (i), suppose that $(X_n)_n$ and Y are as in (†) of Definition(4.7). Then $\rho_p(\int_0^\infty X_n dA) \rightarrow 0$ by using the dominated convergence theorem twice (once for the Stieltjes integral and once for the expectation implicit in ρ_p). But by (1), this means that $\int X_n dA \rightarrow 0$ in L^p . Thus (i) is proven.

Now we consider (ii) and (iii). I claim that

$$(2) \quad \Phi_A^p(F) = \rho_p\left(\int_0^\infty F dA\right) \quad \text{for all } F \in \mathcal{E}_{\text{inc}}^+.$$

For if $X \in \mathcal{E}$ is such that $|X| \leq F$, then by (1), $\rho_p(\int X dA) \leq \rho_p(\int_0^\infty F dA)$. So the left hand side is no bigger than the right hand side. For the reverse inequality, let $X_1, X_2, \dots \in \mathcal{E}^+$ be such that $X_n \uparrow F$. Then $\rho_p(\int X_n dA) \leq \Phi_A^p(F)$. Use (1), let and use the monotone convergence theorem to give that $\rho_p(\int F dA) \leq \Phi_A^p(F)$.

I claim further that

$$(3) \quad \rho_p\left(\int_0^\infty X dA\right) \leq \Phi_A^p(X) \quad \text{for all previsible } X: \mathbb{R}^{++} \times \Omega \rightarrow \overline{\mathbb{R}}^+.$$

For if $X \leq F \in \mathcal{E}_{\text{inc}}^+$, then $\rho_p(\int_0^\infty X dA) \leq \rho_p(\int_0^\infty F dA) = \Phi_A^p(F)$ by (2). Now (3) follows from the definition of $\Phi_A^p(X)$ as an infimum (Definition(4.9)).

Suppose now that $X \in \mathcal{L}^1(A, L^p)$. I claim that $\int_0^\infty |X| dA \in L^p$,

$$(4) \quad \int X dA = \int_0^\infty X dA \quad \text{and} \quad \Phi_A^p(X) = \rho_p\left(\int_0^\infty |X| dA\right).$$

To prove this, note by (3), that $\rho_p(\lambda \int_0^\infty |X| dA) \leq \Phi_A^p(\lambda X) \rightarrow 0$ as $\lambda \downarrow 0$. Thus $\int_0^\infty |X| dA$ is finite a.s. and is an element of L^p . Hence $\int_0^\infty X dA$ exists and is an element of L^p . To prove (4), let $X_1, X_2, \dots \in \mathcal{E}$ be such that $\Phi_A^p(X_n - X) \rightarrow 0$. Then taking $|X_n - X|$ as X in (3) gives that $\int_0^\infty |X_n - X| dA \rightarrow 0$ in L^p . Thus $\int_0^\infty X_n dA \rightarrow \int_0^\infty X dA$ and $\int_0^\infty |X_n| dA \rightarrow \int_0^\infty |X| dA$ in L^p . Now by (1) and (2), $\int X_n dA = \int_0^\infty X_n dA$ and $\Phi_A^p(X_n) = \rho_p(\int_0^\infty |X_n| dA)$. So letting $n \rightarrow \infty$ gives (4).

We have thus proven one half of (ii) and all of (iii). We shall now consider the other half of (ii). So suppose that $X: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}$ is previsible and such that $\int_0^\infty |X| dA \in L^p$. We have to show that $X \in \mathcal{L}^1(A, L^p)$. Take $U_1, U_2, \dots \in \mathcal{E}^+$ such that $U_n \uparrow \infty$, and let $X_n = (X \wedge U_n) \vee (-U_n)$. We have that $X_n \in \mathcal{L}^1(A, L^p)$, $X_n \rightarrow X$ pointwise, and $\rho_p(\int_0^\infty |X_n - X| dA) \rightarrow 0$. By (4), $\Phi_A^p(X_n - X_m) = \rho_p(\int_0^\infty |X_n - X_m| dA) \rightarrow 0$ as $n, m \rightarrow \infty$. So there exists an $\tilde{X} \in \mathcal{L}^1(\Phi_A^p)$ such that $\Phi_A^p(X_n - \tilde{X}) \rightarrow 0$. Now by Proposition(2.10), a subsequence of $(X_n)_n$ converges Φ_A^p -a.e. to $\tilde{X}$. Thus $\tilde{X} = X$ Φ_A^p -a.e. Therefore $X \in \mathcal{L}^1(\Phi_A^p)$ as required.

(b) For (i), note that $A^{(+)}$ and $A^{(-)}$ are L^p -integrators by (a)(i). Thus $A = A^{(+)} - A^{(-)}$ is an L^p -integrator. For $X \in \mathcal{E}$, it is clear that $|\int X dA| \leq \int |X| dV_A$. Therefore by applying ρ_p , and taking the supremum over all $X \in \mathcal{E}$ such that $|X| \leq F$, we obtain that $\Phi_A^p(F) \leq \Phi_{V_A}^p(F)$ for all $F \in \mathcal{E}_{\text{inc}}^+$. Now by taking infima, we deduce that $\Phi_A^p(X) \leq \Phi_{V_A}^p(X)$ for all functions $X: \mathbb{R}^{++} \times \Omega \rightarrow \bar{\mathbb{R}}^+$. Part (ii) clearly follows as a result of the inequality $\Phi_A^p \leq \Phi_{V_A}^p$. To prove (iii), suppose that $X \in \mathcal{L}^1(V_A, L^p)$. Then $X \in \mathcal{L}^1(A, L^p)$ by (ii), and so $X \in \mathcal{L}^1(A^{(\pm)}, L^p)$ (since $A^{(\pm)} = \frac{1}{2}(V_A \pm A)$). Therefore (a)(iii) gives that $\int X dA^{(\pm)} = \int_0^\infty X dA^{(\pm)}$. So subtracting gives the first part of (iii). For the second part, note that the right hand side is, by (a)(iii), just $\Phi_{V_A}^p(X)$. Therefore use the inequality $\Phi_A^p \leq \Phi_{V_A}^p$. $\square$

(5.31)Corollary. Suppose that A is a finite-variation process and $X \in \Lambda^1(V_A)$. Then $X \cdot A$ is also a finite-variation process and $V_{X \cdot A} = |X| \cdot V_A$.

Proof: We have that $X \cdot A = \int_0^t X dA$, and that the integral can be viewed as being pathwise by Theorem(5.30). The result now follows from the theory of Lebesgue-Stieltjes integration. $\square$

§5.6 BROWNIAN MOTION AS AN INTEGRATOR

The only examples of integrators we have so far are the processes of finite variation. Since these could be dealt with by pathwise methods, it is important to give examples that are not of finite variation. In this section we will consider one such example namely that of Brownian motion. We give first the definition.

(5.32)Definition. For $d \in \mathbb{N}_{>0}$, a d -dimensional Brownian motion is a continuous adapted process $B: \mathbb{R}^+ \times \Omega \rightarrow \mathbb{R}^d$ such that the following hold.

- (1) If $s < t$ in $\mathbb{R}^+$, then $B_t - B_s$ is an $N(0, t - s)$ random variable i.e.

$$\mathbb{P}[B_t - B_s \in F] = \left(\frac{1}{2\pi(t-s)} \right)^{d/2} \int_F \exp \left(-\frac{|x|^2}{2(t-s)} \right) dx,$$

for all Borel subsets F of $\mathbb{R}^d$.

- (2) If $s < t$ in $\mathbb{R}^+$, then $B_t - B_s$ is independent of the σ -field $\mathcal{F}_s$ i.e.

$$\mathbb{P}(A \cap [B_t - B_s \in F]) = \mathbb{P}(A)\mathbb{P}[B_t - B_s \in F],$$

for all $A \in \mathcal{F}_s$ and all Borel subsets F of $\mathbb{R}^d$. $\square$

Note.

1. We observe that a d -dimensional Brownian motion satisfies the following properties.

- (1) If $s < t$ in $\mathbb{R}^+$, then $B_t - B_s$ is an $N(0, t - s)$ random variable.
- (2) If $0 \leq t_0 < \dots < t_r < \infty$, then $B_{t_0}, B_{t_1} - B_{t_0}, B_{t_2} - B_{t_1}, \dots, B_{t_r} - B_{t_{r-1}}$ are independent random variables.
- (3) For all $\omega \in \Omega$, the path $t \mapsto B_t(\omega)$ is continuous.

Property (2) follows from (2) of Definition(5.32). A family of $\mathbb{R}^d$ -valued random variables $(B_t)_{t \in \mathbb{R}^+}$ on any probability space $(\Omega, \mathcal{F}, \mathbb{P})$, satisfying (1), (2) and (3) is what is generally known as a d -dimensional Brownian motion. We shall call such an object a raw Brownian motion (since there is no filtration).

2. For this to be an example it is necessary to know that d -dimensional Brownian motions exist for some choices of a standard setup $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_t)$. It is well known that there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a raw d -dimensional Brownian motion B defined on it, and such that B_0 has any distribution desired. Assuming this fact (whose proof is too long to justify including it here), we can first complete $(\Omega, \mathcal{F}, \mathbb{P})$, and therefore assume it is already complete. Then we define a standard filtration $(\mathcal{F}_t)_t$ on it as follows. Let $\mathcal{F}_t^\circ$ be the smallest σ -field that contains all the $\mathbb{P}$ -null sets and makes B_s measurable for $0 \leq s \leq t$. Then take $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s^\circ$. It can then be shown that B is a d -dimensional Brownian motion with respect to the standard setup $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_t)$.

3. If $B = (B^1, \dots, B^d)$ is a d -dimensional Brownian motion and $1 \leq i_1, \dots, i_r \leq d$ are distinct, then $(B^{i_1}, \dots, B^{i_r})$ is an r -dimensional Brownian motion. In particular for $k = 1, \dots, d$, B^k is a 1-dimensional Brownian motion.

4. We shall see in the next chapter that almost all paths of a 1-dimensional Brownian motion fail to have finite variation. Thus pathwise Stieltjes integration will not work for this process.

Our aim now is to show that nevertheless any 1-dimensional Brownian motion is an integrator. By considering $B - B_0$ (which gives the same values for integrals) we

may as well assume that B has zero start; i.e. $B_0 = 0$.

(5.33) Lemma. *Suppose that B is a zero start 1-dimensional Brownian motion. Then $B_t \in \mathcal{L}^2$ for all $t \in \mathbb{R}^+$, and*

$$\mathbb{E} \left(\int X dB \int Y dB \right) = \mathbb{E} \left(\int_0^\infty X_t Y_t dt \right) \quad \text{for all } X, Y \in \mathcal{E}.$$

Proof: B_t is $N(0, t)$ and therefore an element of $\mathcal{L}^2$. To prove the formula suppose that $X, Y \in \mathcal{E}$ and write

$$X_t(\omega) = \sum_{k=0}^{r-1} C_k(\omega) \mathbf{1}_{(t_k, t_{k+1}]}(t), \quad \text{and} \quad Y_t(\omega) = \sum_{k=0}^{r-1} D_k(\omega) \mathbf{1}_{(t_k, t_{k+1}]}(t),$$

where $0 = t_0 < \dots < t_r < \infty$, and for $k = 0, \dots, r-1$, C_k and D_k are simple $\mathcal{F}_{t_k}$ -measurable random variables. Thus

$$\int X dB \int Y dB = \sum_{k,l=0}^{r-1} C_k D_l (B_{t_{k+1}} - B_{t_k})(B_{t_{l+1}} - B_{t_l}).$$

Now if $0 \leq k < l \leq r-1$, $C_k D_l (B_{t_{k+1}} - B_{t_k})$ is $\mathcal{F}_{t_l}$ -measurable and therefore independent of $B_{t_{l+1}} - B_{t_l}$. Thus $\mathbb{E}[C_k D_l (B_{t_{k+1}} - B_{t_k})(B_{t_{l+1}} - B_{t_l})] = \mathbb{E}[C_k D_l (B_{t_{k+1}} - B_{t_k})] \mathbb{E}(B_{t_{l+1}} - B_{t_l}) = 0$. A similar statement is true when $l < k$. For the case when $k = l$, we have that $\mathbb{E}[C_k D_k (B_{t_{k+1}} - B_{t_k})^2] = \mathbb{E}(C_k D_k) \mathbb{E}[(B_{t_{k+1}} - B_{t_k})^2] = \mathbb{E}(C_k D_k)(t_{k+1} - t_k)$. Putting these together we see that

$$\mathbb{E} \left(\int X dB \int Y dB \right) = \sum_{k=0}^{r-1} \mathbb{E}(C_k D_k)(t_{k+1} - t_k) = \mathbb{E} \left(\int_0^\infty X_t Y_t dt \right),$$

as required. □

We now prove a result of more general application than for Brownian motion. It is related to the Strasbourg approach. We shall use it when considering consistency with the Strasbourg approach in Chapter 8.

(5.34) **Theorem.** Suppose that $S: \mathbb{R}^+ \times \Omega \rightarrow \mathbb{R}$ is a L^2 càdlàg process. Further suppose that μ is a measure on $(\mathbb{R}^{++} \times \Omega, \mathcal{P})$ such that $\mu((0, b] \times \Omega) < \infty$ for all $b \in \mathbb{R}^+$, and also that $\mathbb{E}(\int X dS \int Y dS) = \int_{\mathbb{R}^{++} \times \Omega} XY d\mu$ for all $X, Y \in \mathcal{E}$. Then the following hold.

- (i) S is an L^2 -integrator.
- (ii) $\mathcal{L}^1(S, L^2) = \mathcal{L}^2(\mathbb{R}^{++} \times \Omega, \mathcal{P}, \mu)$.
- (iii) $\mathbb{E}(\int X dS \int Y dS) = \int_{\mathbb{R}^{++} \times \Omega} XY d\mu$ for all $X, Y \in \mathcal{L}^1(S, L^2)$.
- (iv) $\Phi_S^2(X) = (\int_{\mathbb{R}^{++} \times \Omega} X^2 d\mu)^{1/2}$ for all previsible $X: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}$.

Proof:

(i) We have only to check that (†) of Definition(4.7) holds. So suppose that $(X_n)_n$ and Y satisfy the conditions given in (†) of Definition(4.7). Then $\mathbb{E}[(\int X_n dS)^2] = \int X_n^2 d\mu \rightarrow 0$ by the dominated convergence theorem of ordinary integration theory. But this is just saying that $\int X_n dS \rightarrow 0$ in L^2 as required.

(ii) First I claim that

$$(1) \quad \Phi_S^2(F) = \left(\int_{\mathbb{R}^{++} \times \Omega} F^2 d\mu \right)^{1/2} \quad \text{for all } F \in \mathcal{E}_{\text{inc}}^+.$$

To prove this, suppose that $X \in \mathcal{E}$ is such that $|X| \leq F$. Then $\rho_2(\int X dS) = (\int X^2 d\mu)^{1/2} \leq (\int F^2 d\mu)^{1/2}$. So the left hand side is no bigger than the right hand side. For the reverse inequality, let $X_1, X_2, \dots \in \mathcal{E}^+$ be such that $X_n \uparrow F$. Then $(\int X_n^2 d\mu)^{1/2} = \rho_2(\int X_n dS) \leq \Phi_S^2(F)$. So letting $n \rightarrow \infty$ and using the monotone convergence theorem gives the required inequality.

Now using a similar argument to that used in the deduction of (3) in the proof of Theorem(5.30), we obtain that

$$(2) \quad \left(\int_{\mathbb{R}^{++} \times \Omega} X^2 d\mu \right)^{1/2} \leq \Phi_S^2(X) \quad \text{for all previsible } X: \mathbb{R}^{++} \times \Omega \rightarrow \overline{\mathbb{R}}.$$

Now suppose that $X \in \mathcal{L}^1(S, L^2)$. Then $\Phi_S^2(X) < \infty$, and so (2) gives that $X \in \mathcal{L}^2(\mu)$. This proves one half. Now I claim that

$$(3) \quad \Phi_S^2(X) = \left(\int_{\mathbb{R}^{++} \times \Omega} X^2 d\mu \right)^{1/2} \quad \text{for all } X \in \mathcal{L}^1(S, L^2).$$

For this is true if $X \in \mathcal{E}$ by (1). For general $X \in \mathcal{L}^1(S, L^2)$, take $X_1, X_2, \dots \in \mathcal{E}$ such that $\Phi_S^2(X_n - X) \rightarrow 0$. Then by (2) with $X_n - X$ as X , we obtain that $X_n \rightarrow X$ in $L^2(\mu)$. But we know that (3) holds with X_n replacing X (by (1)). Let $n \rightarrow \infty$ to give (3) in full generality.

Now suppose that $X \in \mathcal{L}^2(\mu)$. Take $U_1, U_2, \dots \in \mathcal{E}^+$ such that $U_n \uparrow \infty$, and let $X_n = (X \wedge U_n) \vee (-U_n)$. Then $X_n \in \mathcal{L}^1(S, L^2)$, and $X_n \rightarrow X$ pointwise and in $L^2(\mu)$. So by (3), $\Phi_S^2(X_n - X_m) = \left(\int (X_n - X_m)^2 d\mu \right)^{1/2} \rightarrow 0$ as $n, m \rightarrow \infty$. Thus we see that there exists $\tilde{X} \in \mathcal{L}^1(\Phi_S^2)$ such that $\Phi_S^2(X_n - \tilde{X}) \rightarrow 0$. A subsequence of $(X_n)_n$ converges Φ_S^2 -a.e. to $\tilde{X}$ by Proposition(2.10) and so $\tilde{X} = X$ Φ_S^2 -a.e. Thus $X \in \mathcal{L}^1(\Phi_S^2)$ as required.

(iii) Take $X_1, X_2, \dots \in \mathcal{E}$ and $Y_1, Y_2, \dots \in \mathcal{E}$ such that $\Phi_S^2(X_n - X) \rightarrow 0$ and $\Phi_S^2(Y_n - Y) \rightarrow 0$. Then also $X_n \rightarrow X$ and $Y_n \rightarrow Y$ in $L^2(\mu)$ by (2). Now (iii) is true with X and Y replaced by X_n and Y_n respectively. So let $n \rightarrow \infty$ to give the general result.

(iv) We know that this holds for $X \in \mathcal{L}^1(S, L^2)$. But if $X \notin \mathcal{L}^1(S, L^2)$ it follows by (ii) that both sides are infinite, and so (iv) holds in this case also. (Since $\Phi_S^2(\lambda X) = |\lambda| \Phi_S^2(X)$, we have that X is Φ_S^2 -finite if and only if $\Phi_S^2(X) < \infty$.) $\square$

Now by applying this theorem to a zero start Brownian motion with μ as the restriction of the product of Lebesgue measure and $\mathbb{P}$ to $\mathcal{P}$, we obtain the following corollary.

(5.35)Corollary. *Suppose that B is a zero start 1-dimensional Brownian motion. Then the following properties hold.*

(i) B is an L^2 -integrator.

- (ii) $\mathcal{L}^1(B, L^2)$ consists of all the previsible $X: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}$ such that $\mathbb{E}(\int_0^\infty X_t^2 dt) < \infty$.
- (iii) $\mathbb{E}(\int X dB \int Y dB) = \mathbb{E}(\int_0^\infty X_t Y_t dt)$ for all $X, Y \in \mathcal{L}^1(B, L^2)$.
- (iv) $\Phi_B^2(X) = \mathbb{E}(\int_0^\infty X_t^2 dt)^{1/2}$ for all previsible $X: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}$. □

CHAPTER 6: QUADRATIC VARIATION AND THE ITÔ FORMULA

§6.1 MOTIVATION

To evaluate ordinary (Lebesgue) integrals such as $\int_0^t s^2 ds$, we do not do it using the definition in terms of limits and supremums etc. Instead we use the Fundamental Theorem of Calculus. This states that if $f: [a, b] \rightarrow \mathbb{R}$ is a C^1 function, where $-\infty < a < b < \infty$, then

$$(1) \quad f(b) - f(a) = \int_a^b f'(t) dt.$$

Thus in the theory of stochastic integration we would hope for some result analogous to this. To see what form this might take, we observe that if $f: \mathbb{R} \rightarrow \mathbb{R}$ is C^1 and $\alpha: \mathbb{R}^+ \rightarrow \mathbb{R}$ is continuous with finite variation, then

$$(2) \quad f(\alpha(t)) - f(\alpha(0)) = \int_0^t f'(\alpha(s)) d\alpha(s).$$

From this formula we might expect that $f(S_t) - f(S_0) = \int_0^t f'(S) dS$ when $f: \mathbb{R} \rightarrow \mathbb{R}$ is C^1 and S is a continuous integrator. However this is not generally true. Suppose we formally expand in a Taylor series

$$f(S_t + dS_t) = f(S_t) + f'(S_t) dS_t + \frac{1}{2} f''(S_t) (dS_t)^2 + \frac{1}{6} f'''(S_t) (dS_t)^3 + \dots$$

We would expect the term $\frac{1}{2} f''(S_t) (dS_t)^2$ to make no contribution to an integral. But it turns out that ' $\int_0^t (dS)^2$ ' can be made meaningful as an increasing process $[S]$, which is not generally zero. This process is called the quadratic variation of S . So we have ' $(dS_t)^2 = d[S]_t$ '. Putting this in the Taylor series, and putting $(dS_t)^k = 0$ for all $k \geq 3$ hints at the following formula.

$$(3) \quad f(S_t) - f(S_0) = \int_0^t f'(S) dS + \frac{1}{2} \int_0^t f''(S) d[S].$$

This indeed is the true formula; it is valid for all continuous integrators and all C^2 functions $f: \mathbb{R} \rightarrow \mathbb{R}$. (3) is (the 1-dimensional case of) the Itô formula.

§6.2 QUADRATIC VARIATION AND MUTUAL VARIATION

(6.1)Definition. Suppose that R and S are integrators. Then we define

$$[R, S]_t = R_t S_t - R_0 S_0 - \int_0^t R_- dS - \int_0^t S_- dR.$$

I.e. $[R, S] = RS - R_0 S_0 - R_- \cdot S - S_- \cdot R$. So $[R, S]$ is a càdlàg adapted process. It is called the *mutual variation* of R and S . Clearly $[\cdot, \cdot]$ is bilinear. We denote $[S, S]$ by $[S]$ and call it the *quadratic variation* of S . $\square$

Now in what sense is $[R, S]_t$ ‘ $\int_0^t dR dS$ ’? This is answered by the following result.

(6.2)Proposition. Suppose that R and S are integrators and that $t \in \mathbb{R}^+$. Then

$$(1) \quad \sum_{k=0}^{r-1} (R_{t_{k+1}} - R_{t_k})(S_{t_{k+1}} - S_{t_k}) \rightarrow [R, S]_t \quad \text{in probability,}$$

as $\sup_k (t_{k+1} - t_k) \rightarrow 0$ with $0 = t_0 < \dots < t_r = t$. In fact when $s < t$ in $\mathbb{R}^+$,

$$(2) \quad \sum_{k=0}^{r-1} (R_{t_{k+1}} - R_{t_k})(S_{t_{k+1}} - S_{t_k}) \rightarrow [R, S]_t - [R, S]_s \quad \text{in probability,}$$

as $\sup_k (t_{k+1} - t_k) \rightarrow 0$ with $s = t_0 < \dots < t_r = t$.

Proof: We consider (1). We have for $0 = t_0 < \dots < t_r = t$ that

$$\begin{aligned} \sum_{k=0}^{r-1} (R_{t_{k+1}} - R_{t_k})(S_{t_{k+1}} - S_{t_k}) &= \sum_{k=0}^{r-1} (R_{t_{k+1}} S_{t_{k+1}} - R_{t_k} S_{t_k} \\ &\quad - R_{t_k} (S_{t_{k+1}} - S_{t_k}) - S_{t_k} (R_{t_{k+1}} - R_{t_k})) \\ &= R_t S_t - R_0 S_0 - \sum_{k=0}^{r-1} R_{t_k} (S_{t_{k+1}} - S_{t_k}) \\ &\quad - \sum_{k=0}^{r-1} S_{t_k} (R_{t_{k+1}} - R_{t_k}) \\ &\rightarrow R_t S_t - R_0 S_0 - \int_0^t R_- dS - \int_0^t S_- dR, \end{aligned}$$

in probability as $\sup_k(t_{k+1} - t_k) \rightarrow 0$. But by Definition(6.1), the limit is $[R, S]_t$ as required. For (2) suppose that $s = t_0^{(n)} < \dots < t_{r_n}^{(n)} = t$ and $\sup_k(t_{k+1}^{(n)} - t_k^{(n)}) \rightarrow 0$ as $n \rightarrow \infty$. Then choose $0 = u_0^{(n)} < \dots < u_{p_n}^{(n)} = s$ such that $\sup_k(u_{k+1}^{(n)} - u_k^{(n)}) \rightarrow 0$. Let $0 = v_0^{(n)} < \dots < v_{q_n}^{(n)} = t$ be the partition of $[0, t]$ formed by combining those for $[0, s]$ and $[s, t]$ (i.e. $q_n = p_n + r_n$, $v_k^{(n)} = u_k^{(n)}$ for $k = 0, \dots, p_n$, and $v_k^{(n)} = t_{k-p_n}^{(n)}$ for $k = p_n, \dots, q_n$). Then

$$\begin{aligned} \sum_{k=0}^{r_n-1} (R_{t_{k+1}^{(n)}} - R_{t_k^{(n)}})(S_{t_{k+1}^{(n)}} - S_{t_k^{(n)}}) &= \sum_{k=0}^{q_n-1} (R_{v_{k+1}^{(n)}} - R_{v_k^{(n)}})(S_{v_{k+1}^{(n)}} - S_{v_k^{(n)}}) \\ &\quad - \sum_{k=0}^{p_n-1} (R_{u_{k+1}^{(n)}} - R_{u_k^{(n)}})(S_{u_{k+1}^{(n)}} - S_{u_k^{(n)}}) \\ &\rightarrow [R, S]_t - [R, S]_s, \end{aligned}$$

in probability as $n \rightarrow \infty$ as required. $\square$

(6.3)Proposition. *Suppose that R and S are integrators. Then the following properties hold.*

- (i) $[S]$ is an increasing zero start càdlàg process.
- (ii) $[R, S]$ is a finite-variation process, and therefore, by Theorem(5.30)(b)(ii), it is an integrator.
- (iii) (Integration by Parts) $R_t S_t - R_0 S_0 = \int_0^t R_- dS + \int_0^t S_- dR + [R, S]_t$ for all $t \in \mathbb{R}^+$.
- (iv) $\Delta[R, S] = \Delta R \Delta S$ (the equality being indistinguishability). So $[R, S]$ is continuous whenever R or S is continuous.
- (v) RS is an integrator.

Proof:

- (i) The only non trivial thing to check is that $[S]$ is increasing. Suppose that $s < t$ are in $\mathbb{R}^+$. Then (2) of Proposition(6.2) (with $R = S$) expresses $[S]_t - [S]_s$ as a limit in probability of non-negative random variables. Thus $[S]_s \leq [S]_t$ a.s. So on letting $E = \bigcup_{s, t \in \mathbb{Q}^+ : s < t} [[S]_s > [S]_t]$, we have that E is $\mathbb{P}$ -null. For $\omega \in \Omega \setminus E$, $t \mapsto [S]_t(\omega)$

is increasing, since $[S]$ is càdlàg. Thus by modifying $[S]$ on an evanescent set, we see that $[S]$ is increasing.

(ii) Since $[\cdot, \cdot]$ is bilinear, we see that $[R, S] = [\frac{1}{2}(R + S)] - [\frac{1}{2}(R - S)]$. Thus $[R, S]$ is the difference of two increasing process, and therefore has finite variation.

(iii) This just Definition(6.1) rewritten.

(iv) We have that $[R, S] = RS - R_0S_0 - R_- \cdot S - S_- \cdot R$ by definition. So Theorem(5.25) gives that

$$\begin{aligned}\Delta[R, S] &= \Delta(RS) - R_- \Delta S - S_- \Delta R \\ &= RS - R_- S_- - R_-(S - S_-) - S_-(R - R_-) \\ &= (R - R_-)(S - S_-) \\ &= \Delta R \Delta S,\end{aligned}$$

as required.

(v) We see that (iii) expresses RS as a sum of integrators (using (ii) and Theorem(4.21)(i)). Therefore RS an integrator (by Proposition(4.19)). $\square$

Of cause we as yet have no examples to show that $[R, S]$ need not be zero. We now give such an example.

(6.4)Proposition.

- (i) If B is a 1-dimensional Brownian motion, then $[B]_t = t$ for all t .
- (ii) If $B = (B^1, \dots, B^d)$ is a d -dimensional Brownian motion, then $[B^i, B^j]_t = \delta_{i,j}t$ for all i, j and t .

Proof:

- (i) First I claim that for an $N(0, 1)$ random variable X ,

$$(1) \quad \mathbb{E}(X^2) = 1 \quad \text{and} \quad \mathbb{E}(X^4) = 3.$$

Let $\phi(t) = \mathbb{E}(e^{itX})$. So $\phi(t) = e^{-t^2/2}$. Then

$$\begin{aligned}\phi'(t) &= -te^{-t^2/2}, & \phi''(t) &= (-1 + t^2)e^{-t^2/2}, \\ \phi'''(t) &= (3t - t^3)e^{-t^2/2}, & \phi^{iv}(t) &= (3 - 6t^2 + t^4)e^{-t^2/2}.\end{aligned}$$

Thus $\mathbb{E}(X^2) = -\phi''(0) = 1$, and $\mathbb{E}(X^4) = \phi^{iv}(0) = 3$. So (1) is proven.

Now suppose that $s < t$ are in $\mathbb{R}^+$, and let $X = (t - s)^{-1/2}(B_t - B_s)$. Then X is $N(0, 1)$ and $B_t - B_s = (t - s)^{1/2}X$. Therefore (1) gives that

$$(2) \quad \mathbb{E}((B_t - B_s)^2) = t - s \quad \text{and} \quad \mathbb{E}((B_t - B_s)^4) = 3(t - s)^2.$$

Now suppose that $t \in \mathbb{R}^+$ and that $0 = t_0 < \dots < t_r = t$. Let $\delta = \sup_k(t_{k+1} - t_k)$, $S = \sum_{k=0}^{r-1}(B_{t_{k+1}} - B_{t_k})^2$, and $\eta = \sum_{k=0}^{r-1}(t_{k+1} - t_k)^2$. Then by Proposition(6.2), $S \rightarrow [B]_t$ in probability as $\delta \downarrow 0$. Now $\mathbb{E}(S) = \sum_{k=0}^{r-1} \mathbb{E}[(B_{t_{k+1}} - B_{t_k})^2] = \sum_{k=0}^{r-1}(t_{k+1} - t_k) = t$, and

$$\begin{aligned}\mathbb{E}(S^2) &= \mathbb{E}\left(\sum_{k=0}^{r-1}(B_{t_{k+1}} - B_{t_k})^4 + \sum_{k \neq l} (B_{t_{k+1}} - B_{t_k})^2(B_{t_{l+1}} - B_{t_l})^2\right) \\ &= \sum_{k=0}^{r-1} 3(t_{k+1} - t_k)^2 + \sum_{k \neq l} \mathbb{E}[(B_{t_{k+1}} - B_{t_k})^2] \mathbb{E}[(B_{t_{l+1}} - B_{t_l})^2] \\ &= 3\eta + \sum_{k \neq l} (t_{k+1} - t_k)(t_{l+1} - t_l) \\ &= 3\eta + \sum_{k, l} (t_{k+1} - t_k)(t_{l+1} - t_l) - \eta \\ &= 2\eta + t^2.\end{aligned}$$

Thus $\mathbb{E}[(S - t)^2] = \mathbb{E}(S^2) - 2t\mathbb{E}(S) + t^2 = 2\eta \rightarrow 0$ as $\delta \downarrow 0$ (since $\eta \leq t\delta$). I.e. $S \rightarrow t$ in L^2 as $\delta \downarrow 0$, and therefore it follows that $[B]_t = t$.

(ii) We have proved this result in the case when $i = j$; so we consider the case when $i \neq j$. In this case (B^i, B^j) is a 2-dimensional Brownian motion. Thus

$$\begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix} \begin{pmatrix} B^i \\ B^j \end{pmatrix}$$

is also a 2-dimensional Brownian motion, since the 2×2 matrix is orthogonal. Thus $\frac{1}{\sqrt{2}}(B^i + B^j)$ and $\frac{1}{\sqrt{2}}(B^i - B^j)$ are 1-dimensional Brownian motions. Therefore (i) gives that $[B^i, B^j] = \frac{1}{2}[\frac{1}{\sqrt{2}}(B^i + B^j)] - \frac{1}{2}[\frac{1}{\sqrt{2}}(B^i - B^j)] = \frac{1}{2}t - \frac{1}{2}t = 0$ as required. $\square$

We can use this to prove a fact about Brownian motion stated in the previous chapter, namely that almost all paths of a Brownian motion fail to have finite variation.

(6.5) Proposition. *Suppose that B is a 1-dimensional Brownian motion. Then for almost all $\omega \in \Omega$, we have that for each $a < b$ in $\mathbb{R}^+$, the function $t \mapsto B_t(\omega)$ has infinite variation on the interval $[a, b]$. (The variation of a function $\alpha: [a, b] \rightarrow \mathbb{R}$ on $[a, b]$ is defined to be the supremum of all the sums $\sum_{k=0}^{r-1} |\alpha(t_{k+1}) - \alpha(t_k)|$, where $a = t_0 < \dots < t_r = b$.)*

Proof: For each $a < b$ in $\mathbb{R}^+$, let $E_{a,b}$ denote the set of all $\omega \in \Omega$ such that $t \mapsto B_t(\omega)$ has finite variation on $[a, b]$. We will show first that each $E_{a,b}$ is $\mathbb{P}$ -null. Let $V(\omega)$ denote the variation of $t \mapsto B_t(\omega)$ on $[a, b]$. Then on the set $E_{a,b}$, we have for $a = t_0 < \dots < t_r = b$ that

$$\begin{aligned} \sum_{k=0}^{r-1} (B_{t_{k+1}} - B_{t_k})^2 &\leq \sup_k |B_{t_{k+1}} - B_{t_k}| \sum_{k=0}^{r-1} |B_{t_{k+1}} - B_{t_k}| \\ &\leq \sup_k |B_{t_{k+1}} - B_{t_k}| V \\ &\rightarrow 0 \quad \text{as} \quad \sup_k (t_{k+1} - t_k) \rightarrow 0. \end{aligned}$$

But by (2) of Proposition(6.2), $\sum_{k=0}^{r-1} (B_{t_{k+1}} - B_{t_k})^2 \rightarrow [B]_b - [B]_a$ in probability as $\sup_k (t_{k+1} - t_k) \rightarrow 0$. Now $[B]_b - [B]_a = b - a \neq 0$ a.s., and so it follows that $E_{a,b}$ must be $\mathbb{P}$ -null. Let $E = \bigcup_{a,b \in \mathbb{Q}^+: a < b} E_{a,b}$. Then E is $\mathbb{P}$ -null, and the conclusion of the proposition is true for all $\omega \in \Omega \setminus E$. $\square$

Concerning processes of finite variation we have the following result.

(6.6)Proposition. Suppose that S is an integrator and that A is a finite-variation process. Then for each $t \in \mathbb{R}^+$,

$$[S, A]_t = \sum_{u \leq t} (\Delta S)_u (\Delta A)_u.$$

Thus if either S or A is continuous, then $[S, A] = 0$.

Proof: Suppose that $0 = t_0 < \dots < t_r = t$. Then $S_{t_{k+1}} A_{t_{k+1}} - S_{t_k} A_{t_k} = S_{t_{k+1}} (A_{t_{k+1}} - A_{t_k}) + A_{t_k} (S_{t_{k+1}} - S_{t_k})$. So by summing up for $k = 0, \dots, r-1$ this gives that $S_t A_t - S_0 A_0 = \sum_{k=0}^{r-1} S_{t_{k+1}} (A_{t_{k+1}} - A_{t_k}) + \sum_{k=0}^{r-1} A_{t_k} (S_{t_{k+1}} - S_{t_k})$. Let $\sup_k (t_{k+1} - t_k) \rightarrow 0$ through a sequence of values to give that

$$S_t A_t - S_0 A_0 = \int_0^t S_- dA + \int_0^t A_- dS,$$

where the first integral is a pathwise Stieltjes integral (use the dominated convergence theorem), and the second integral is a stochastic integral (use Corollary(4.29)). We have therefore from Definition(6.1) that

$$\begin{aligned} [S, A]_t &= S_t A_t - S_0 A_0 - \int_0^t S_- dA + \int_0^t A_- dS \\ &= \int_0^t (S - S_-) dA \\ &= \sum_{u \leq t} (\Delta S)_u (\Delta A)_u, \end{aligned}$$

as required. □

§6.3 THE MUTUAL VARIATION OF TWO STOCHASTIC INTEGRALS

Suppose that R and S are integrators, $X \in \Lambda^1(R)$ and $Y \in \Lambda^1(S)$. Then from the heuristic formula ' $(X dR)(Y dS) = (XY)(dR dS)$ ' we would expect that

$$(\dagger) \quad [X \bullet R, Y \bullet S] = (XY) \bullet [R, S].$$

This is indeed the case. The result was first given by Kunita and Watanabe in the case when R and S are L^2 -martingales, see [KW]. Also see IV.28 of Rogers & Williams [RW]. In this section we shall present a proof of (†) in the case when R and S are any integrators, $X \in \Lambda^1(R)$, and $Y \in \Lambda^1(S)$. We shall need the following result.

(6.7) Theorem.

(a) Suppose that $\alpha, \beta, \gamma: \mathbb{R}^+ \rightarrow \mathbb{R}$ are such that α and β are increasing and right continuous, and $|\gamma(t) - \gamma(s)| \leq (\alpha(t) - \alpha(s))^{1/2} (\beta(t) - \beta(s))^{1/2}$ for all $s \leq t$ in $\mathbb{R}^+$. Then the following are true.

- (i) γ is càdlàg with finite variation.
- (ii) $V_\gamma(t) - V_\gamma(s) \leq (\alpha(t) - \alpha(s))^{1/2} (\beta(t) - \beta(s))^{1/2}$ for all $s \leq t$ in $\mathbb{R}^+$.
- (iii) $\int_0^\infty |fg| dV_\gamma \leq (\int_0^\infty |f|^2 d\alpha)^{1/2} (\int_0^\infty |g|^2 d\beta)^{1/2}$ for all Borel functions $f, g: \mathbb{R}^{++} \rightarrow \mathbb{R}$.

(b) If R and S are integrators then the following hold.

- (i) $|[R, S]_t - [R, S]_s| \leq ([R]_t - [R]_s)^{1/2} ([S]_t - [S]_s)^{1/2}$ for all $s \leq t$ in $\mathbb{R}^+$.
- (ii) $\int_0^\infty |XY| dV_{[R, S]} \leq (\int_0^\infty |X|^2 d[R])^{1/2} (\int_0^\infty |Y|^2 d[S])^{1/2}$ for all previsible $X, Y: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}$.

Part (b)(ii) of this result in the case when R and S are L^2 -martingales is one of the Kunita-Watanabe inequalities (see IV.28 of Rogers & Williams [RW]).

Proof:

(a) Suppose that $s = t_0 < \dots < t_r = t$. Then using the Cauchy-Schwarz inequality we have that

$$\begin{aligned}
 (\dagger) \quad \sum_{k=0}^{r-1} |\gamma(t_{k+1}) - \gamma(t_k)| &\leq \sum_{k=0}^{r-1} (\alpha(t_{k+1}) - \alpha(t_k))^{1/2} (\beta(t_{k+1}) - \beta(t_k))^{1/2} \\
 &\leq \left(\sum_{k=0}^{r-1} (\alpha(t_{k+1}) - \alpha(t_k)) \right)^{1/2} \left(\sum_{k=0}^{r-1} (\beta(t_{k+1}) - \beta(t_k)) \right)^{1/2} \\
 &= (\alpha(t) - \alpha(s))^{1/2} (\beta(t) - \beta(s))^{1/2}.
 \end{aligned}$$

Put $s = 0$ and take the supremum to give that $V_\gamma(t) < \infty$. I.e. γ has finite variation. To show that γ is càdlàg it is enough to show it is right continuous (it is the difference of two increasing functions; therefore left and right limits exist). But if $t, t_0 \in \mathbb{R}^+$ with $t > t_0$, then $|\gamma(t) - \gamma(t_0)| \leq (\alpha(t) - \alpha(t_0))^{1/2} (\beta(t) - \beta(t_0))^{1/2} \rightarrow 0$ as $t \downarrow t_0$. This completes the proof of (i). For (ii) take general $s < t$ in $(\dagger)$ and take the supremum.

Now we consider (iii). Clearly without loss of generality we need only consider the case when $f, g \geq 0$. Let $\mathcal{A}$ denote the set of pairs (f, g) of non-negative Borel functions $f, g: \mathbb{R}^{++} \rightarrow \mathbb{R}^+$ for which (iii) holds. I claim that the following are true.

- (1) If $f, g: \mathbb{R}^{++} \rightarrow \mathbb{R}^+$ are Borel, and $(f_1, g_1), (f_2, g_2), \dots \in \mathcal{A}$ such that $f_n \uparrow f$ and $g_n \uparrow g$ pointwise, then $(f, g) \in \mathcal{A}$.
- (2) If $f, g: \mathbb{R}^{++} \rightarrow \mathbb{R}^+$ are Borel, $(f_1, g_1), (f_2, g_2), \dots \in \mathcal{A}$ such that $f_n \downarrow f$ and $g_n \downarrow g$ pointwise, $\int_0^\infty f_1 d\alpha < \infty$, and $\int_0^\infty g_1 d\beta < \infty$, then $(f, g) \in \mathcal{A}$.
- (3) If $A_1, \dots, A_r$ are disjoint Borel subsets of $\mathbb{R}^{++}$ such that for $k = 1, \dots, r$, $(\mathbf{1}_{A_k}, \mathbf{1}_{A_k}) \in \mathcal{A}$, and for some $a_1, \dots, a_r, b_1, \dots, b_r \in \mathbb{R}^+$, $f = \sum_{k=1}^r a_k \mathbf{1}_{A_k}$ and $g = \sum_{k=1}^r b_k \mathbf{1}_{A_k}$, then $(f, g) \in \mathcal{A}$.

(1) and (2) follow easily from the monotone convergence theorem and the dominated convergence theorem respectively. For (3) we have

$$\begin{aligned}
\int_0^\infty fg dV_\gamma &= \sum_{k=1}^r a_k b_k \int_0^\infty \mathbf{1}_{A_k} dV_\gamma \\
&\leq \sum_{k=1}^r a_k b_k \left(\int_0^\infty \mathbf{1}_{A_k} d\alpha \right)^{1/2} \left(\int_0^\infty \mathbf{1}_{A_k} d\beta \right)^{1/2} \\
&\quad \text{(since } (\mathbf{1}_{A_k}, \mathbf{1}_{A_k}) \in \mathcal{A} \text{)} \\
&\leq \left(\sum_{k=1}^r a_k^2 \int_0^\infty \mathbf{1}_{A_k} d\alpha \right)^{1/2} \left(\sum_{k=1}^r b_k^2 \int_0^\infty \mathbf{1}_{A_k} d\beta \right)^{1/2} \\
&\quad \text{(by the Cauchy-Schwarz inequality)} \\
&= \left(\int_0^\infty f^2 d\alpha \right)^{1/2} \left(\int_0^\infty g^2 d\beta \right)^{1/2},
\end{aligned}$$

as required.

Let $\mathcal{G} = \{A \subseteq \mathbb{R}^{++} \mid A \text{ is Borel and } (\mathbf{1}_A, \mathbf{1}_A) \in \mathcal{A}\},$

and $\mathcal{H} = \{A \subseteq \mathbb{R}^{++} \mid A \text{ is Borel and } A \cap (0, n] \in \mathcal{G} \text{ for all } n \in \mathbb{N}\}.$

We have that $\mathcal{G}$ contains all sets of the form $(s, t]$ ($s \leq t$ in $\mathbb{R}^+$) and therefore all disjoint unions of these by (3). These disjoint unions form a ring of subsets of $\mathbb{R}^{++}$, call it $\mathcal{R}$. It now follows that $\mathcal{R} \subseteq \mathcal{H}$. By (1) and (2), $\mathcal{H}$ is a monotone class (i.e. closed under taking countable increasing unions and countable decreasing intersections). Therefore $\mathcal{H}$ contains the monotone class generated by $\mathcal{R}$. But by the monotone class lemma this is just the σ -ring generated by $\mathcal{R}$, i.e. the Borel σ -field of $\mathbb{R}^{++}$. Thus we have shown that $\mathcal{H}$ contains all the Borel subsets of $\mathbb{R}^{++}$. Now if A is a Borel subset of $\mathbb{R}^{++}$, then $A \cap (0, n] \in \mathcal{G}$ for all n and $A \cap (0, n] \uparrow A$. Thus $A \in \mathcal{G}$ by (1). So $\mathcal{G}$ contains all the Borel subsets of $\mathbb{R}^{++}$. Therefore if $f, g: \mathbb{R}^{++} \rightarrow \mathbb{R}^+$ are simple Borel functions, then $(f, g) \in \mathcal{A}$ by (3). Now for general Borel functions $f, g: \mathbb{R}^{++} \rightarrow \mathbb{R}^+$, there exist sequences $(f_n)_n$ and $(g_n)_n$ of simple Borel functions from $\mathbb{R}^{++}$ to $\mathbb{R}^+$ such that $f_n \uparrow f$ and $g_n \uparrow g$ pointwise. Therefore $(f, g) \in \mathcal{A}$ by (1). This proves (iii) in full generality.

(b) It is enough to prove (i), since (ii) follows from (i) by applying (a)(iii) pathwise. To prove (i) fix some $s \leq t$ in $\mathbb{R}^+$, and for each càdlàg process X , let $DX = X_t - X_s$. Then we have to show that $(D[R, S])^2 \leq D[R]D[S]$. This we prove in the same way as the Cauchy-Schwarz inequality. We have for each $\lambda \in \mathbb{R}$ that $D[R + \lambda S] \geq 0$ a.s., and so expanding by bilinearity gives that $D[R] + 2\lambda D[R, S] + \lambda^2 D[S] \geq 0$ a.s. for each $\lambda \in \mathbb{R}$. Thus $\mathbb{P}[D[R] + 2\lambda D[R, S] + \lambda^2 D[S] \geq 0 \text{ for all } \lambda \in \mathbb{Q}] = 1$, and therefore by continuity in λ ,

$$(\dagger) \quad \mathbb{P}[D[R] + 2\lambda D[R, S] + \lambda^2 D[S] \geq 0 \text{ for all } \lambda \in \mathbb{R}] = 1.$$

We will have completed the proof if we show that $b^2 \leq ac$, whenever $a, b, c \in \mathbb{R}$ satisfy $a, c \geq 0$ and $a + 2b\lambda + c\lambda^2 \geq 0$ for all $\lambda \in \mathbb{R}$. Now if $c = 0$ then $b = 0$, and so $b^2 \leq ac$ is

trivial. If $c > 0$ then put $\lambda = -\frac{b}{c}$ to give that $a - \frac{b^2}{c} \geq 0$, which on rearranging gives that $b^2 \leq ac$. $\square$

We now can now state and prove the promised theorem.

(6.8) Theorem. Suppose that R and S are integrators, $X \in \Lambda^1(R)$ and $Y \in \Lambda^1(S)$. Then $XY \in \Lambda^1(V_{[R,S]}) \subseteq \Lambda^1([R, S])$ and $[X \cdot R, Y \cdot S] = (XY) \cdot [R, S]$.

Note.

1. Putting $R = S$ and $X = Y$ shows that if S is an integrator, and $X \in \Lambda^1(S)$, then $X^2 \in \Lambda^1([S])$. This by Theorem(5.30)(a)(ii) means that $\int_0^t X^2 d[S] < \infty$ a.s. for each $t \in \mathbb{R}^+$.

2. Putting $Y = 1$ gives that $[X \cdot R, S] = X \cdot [R, S]$ for all $X \in \Lambda^1(R)$.

Proof: Suppose throughout that R and S are integrators. We shall present the proof as a series of numbered steps.

(1) If $X_1, X_2, \dots, X, Y \in \Lambda^1(R)$ are such that $|X_n| \leq Y$ for all n , and $X_n \rightarrow X$ pointwise, then $[X_n \cdot R, S]_t \rightarrow [X \cdot R, S]_t$ in probability for each $t \in \mathbb{R}^+$.

Proof: By replacing X_n by $X_n - X$ and Y by $2Y$, we may assume that $X = 0$. Now by the definition of $[\cdot, \cdot]$ and Theorem(4.21), $[X_n \cdot R, S]_t = S_t \int_0^t X_n dR - \int_0^t (X_n \cdot R)_- dS - \int_0^t X_n S_- dR$. By the Dominated Convergence Theorem (4.17), $\Phi_R^0(\mathbf{1}_{(0,t] \times \Omega} X_n) \rightarrow 0$. Therefore by Theorem(5.23), $(X_n \cdot R)_t^* \rightarrow 0$ in probability. From Theorem(4.24)(ii), it now follows that $\Phi_S^0(\mathbf{1}_{(0,t] \times \Omega} (X_n \cdot R)_-) \rightarrow 0$. Thus $\int_0^t (X_n \cdot R)_- dS \rightarrow 0$ in probability. We have also that $\int_0^t X_n dR \rightarrow 0$ and $\int_0^t X_n S_- dR \rightarrow 0$ in probability, by the Dominated Convergence Theorem (4.17) using dominating processes Y and YS_- respectively. ($YS_- \in \Lambda^1(R)$ by Theorem(4.21)(iv), since by Theorem(4.24)(i), $S_- \in \Lambda^1(Y \cdot R)$.) It is now clear that $[X_n \cdot R, S] \rightarrow 0$ in probability as required.

Now let

$$C_{R,S} = \{ X \in \Lambda^1(R) \cap \Lambda^1(V_{[R,S]}) \mid [X \cdot R, S] = X \cdot [R, S] \}$$

Note that $1 \in \mathcal{C}_{R,S}$.

- (2) If $X_1, X_2, \dots \in \mathcal{C}_{R,S}$, $Y \in \Lambda^1(R) \cap \Lambda^1(V_{[R,S]})$, and $X: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}$ are such that $X_n \rightarrow X$ pointwise, and $|X_n| \leq Y$ for all n , then $X \in \mathcal{C}_{R,S}$.

Proof: We have that $[X_n \cdot R, S]_t = (X_n \cdot [R, S])_t$ for all $t \in \mathbb{R}^+$. So let $n \rightarrow \infty$, and use (1) on the left hand side and the Dominated Convergence Theorem (4.17) on the right hand side to give (2).

- (3) $\mathcal{E} \subseteq \mathcal{C}_{R,S}$.

Proof: Clearly by linearity we need only consider the case when $X = \mathbf{1}_A$, with A a previsible rectangle $(s, t] \times F$ ($0 \leq s < t < \infty$ and $F \in \mathcal{F}_s$). Now $(X \cdot R)_u = \mathbf{1}_F(R_{u \wedge t} - R_{u \wedge s})$, and thus it follows using Proposition(6.2) that $[X \cdot R, S]_u = \mathbf{1}_F([R, S]_{u \wedge t} - [R, S]_{u \wedge s}) = (X \cdot [R, S])_u$ as required.

- (4) $\mathbf{1}_A \in \mathcal{C}_{R,S}$ for all $A \in \mathcal{P}$.

Proof: Let

$$\mathcal{G} = \{ A \in \mathcal{P} \mid \mathbf{1}_A \in \mathcal{C}_{R,S} \},$$

$$\text{and } \mathcal{R} = \{ E \subseteq \mathbb{R}^{++} \times \Omega \mid \mathbf{1}_E \in \mathcal{E} \}.$$

Then $\mathcal{R}$ is a ring (it is the ring of subsets generated by the set of all previsible rectangles), and $\mathcal{R} \subseteq \mathcal{G}$ by (3). Now $\mathcal{G}$ is a monotone class by (2), and $\mathcal{R}$ generates $\mathcal{P}$ as a σ -ring. Thus by the monotone class lemma, $\mathcal{G} = \mathcal{P}$ as required.

- (5) $\mathcal{C}_{R,S} = \Lambda^1(R) \cap \Lambda^1(V_{[R,S]})$.

Proof: Suppose that X is an element of the right hand side. Then by writing $X = X^+ - X^-$ we may assume that $X \geq 0$. Take simple previsible $X_1, X_2, \dots: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}^+$ such that $X_n \uparrow X$. Then $X_n \in \mathcal{C}_{R,S}$ for all n , by (4) and linearity. Therefore $X \in \mathcal{C}_{R,S}$ by (2) with $Y = X$.

- (6) If $X \in \Lambda^1(R)$, then $X^2 \in \Lambda^1([R])$.

Proof: By considering $|X|$ we may suppose without loss of generality that $X \geq 0$. Put $X_n = X \wedge n$. Then $X_n \uparrow X$ pointwise. Now by (5), $X_n \in \mathcal{C}_{R,R}$ for all n , and

$X_n \in \mathcal{C}_{R, X_m \cdot R}$ for all n and m . Thus we see that $[X_n \cdot R, X_m \cdot R] = X_n \cdot [R, X_m \cdot R] = X_n \cdot (X_m \cdot [R]) = (X_n X_m) \cdot [R]$. So we have that

$$\int_0^t X_n X_m d[R] = [X_n \cdot R, X_m \cdot R]_t.$$

Let $n \rightarrow \infty$ and then $m \rightarrow \infty$ using the monotone convergence theorem pathwise on the left hand side and (1) on the right hand side (with $Y = X$) to give that

$$\int_0^t X^2 d[R] = [X \cdot R]_t < \infty \quad \text{a.s.}$$

(7) If $X \in \Lambda^1(R)$ and $Y \in \Lambda^1(S)$, then $XY \in \Lambda^1(V_{[R, S]})$.

Proof: By (6), $X^2 \in \Lambda^1([R])$ and $Y^2 \in \Lambda^1([S])$. Therefore by Theorem(6.7)(b)(ii), $XY \in \Lambda^1(V_{[R, S]})$ as required.

(8) If $X \in \Lambda^1(R)$ and $Y \in \Lambda^1(S)$, then $[X \cdot R, Y \cdot S] = (XY) \cdot [R, S]$.

Proof: By applying (7) with 1 as X , we find that $Y \in \Lambda^1(V_{[R, S]})$. But by applying (7) with 1 as Y , and $Y \cdot S$ as S , we find that $X \in \Lambda^1(V_{[R, Y \cdot S]})$. Thus by (5), it follows that $[R, Y \cdot S] = Y \cdot [R, S]$ and $[X \cdot R, Y \cdot S] = X \cdot [R, Y \cdot S]$. Combining these gives that $[X \cdot R, Y \cdot S] = X \cdot [R, Y \cdot S] = X \cdot (Y \cdot [R, S]) = (XY) \cdot [R, S]$ as required.

This completes the proof of the theorem. □

Concerning localization we have the following corollary.

(6.9)Corollary. *If R and S are integrators, and σ is a stopping time, then $[R^\sigma, S^\sigma] = [R, S]^\sigma$.*

Proof: Since $R^\sigma = \mathbf{1}_{(0, \sigma]} \cdot R$ and $S^\sigma = \mathbf{1}_{(0, \sigma]} \cdot S$ by Proposition(5.17), take $X = Y = \mathbf{1}_{(0, \sigma]}$ in Theorem(6.8). □

§6.4 THE ITÔ FORMULA

We shall state and prove this important result first for continuous integrators. We do this because the result is simpler to state and prove in this case.

(6.10) Theorem (The Itô Formula).

Suppose that $S = (S^1, \dots, S^d): \mathbb{R}^+ \times \Omega \rightarrow D$, where D is an open subset of $\mathbb{R}^d$, and $S^1, \dots, S^d$ are continuous integrators. Suppose that $f: D \rightarrow \mathbb{R}$ is a C^2 function. Then for $t \in \mathbb{R}^+$,

$$f(S_t) - f(S_0) = \sum_{i=1}^d \int_0^t \partial_i f(S) dS^i + \frac{1}{2} \sum_{i,j=1}^d \int_0^t \partial_i \partial_j f(S) d[S^i, S^j].$$

In particular we see that $f(S)$ is a continuous integrator.

In order to prove this we need a lemma.

(6.11) Lemma. Suppose that D is an open subset of $\mathbb{R}^d$, $N \in \mathbb{N} \cup \{\infty\}$, and $f: D \rightarrow \mathbb{C}$ is C^N . Then there exists a sequence of polynomial functions $p_1, p_2, \dots: \mathbb{R}^d \rightarrow \mathbb{C}$ such that $\partial^\alpha p_n \rightarrow \partial^\alpha f$ uniformly on all compact subsets of D for all $\alpha \in \mathbb{N}^d$ with $|\alpha| \leq N$. (Here we are using the notation of (1.6).)

Proof: It is enough to assume $N < \infty$, and show that given $K \subseteq D$ compact and $\varepsilon > 0$, there exists a polynomial $p: \mathbb{R}^d \rightarrow \mathbb{C}$ such that $|\partial^\alpha f - \partial^\alpha p| \leq \varepsilon$ when $|\alpha| \leq N$. (For if this is the case then for general N , take $K_1, K_2, \dots$ to be compact subsets of D such that $\text{int } K_n \uparrow D$. Then take $p_n: \mathbb{R}^d \rightarrow \mathbb{C}$ to be a polynomial such that $|\partial^\alpha f - \partial^\alpha p_n| \leq \frac{1}{n}$ on K_n when $|\alpha| \leq N \wedge n$.)

By rescaling we may assume without loss of generality that K is a subset of an open cube $U = \prod_{k=1}^d (a_k, a_k + 2\pi)$. Take a C^∞ function of compact support $\rho: \mathbb{R}^d \rightarrow [0, 1]$ such that $\rho = 1$ in a neighbourhood of K and $\text{supp } \rho \subseteq U \cap D$. Thus

$\rho f = f$ in a neighbourhood of K , and ρf is a C^N function on $\mathbb{R}^d$ (it is assumed to be zero outside D) with support a compact subset of $U \cap D$. Define $g: \mathbb{R}^d \rightarrow \mathbb{C}$ by

$$g(x) = \sum_{\xi \in \mathbb{Z}^d} (\rho f)(x + 2\pi\xi).$$

So g is C^N on $\mathbb{R}^d$. Let $\mathbb{T}^d = \mathbb{R}^d / 2\pi\mathbb{Z}^d$, and identify functions on $\mathbb{T}^d$ with suitably periodic functions on $\mathbb{R}^d$. Thus $g: \mathbb{T}^d \rightarrow \mathbb{C}$. Now to complete the proof we shall first approximate g by a C^∞ function h , then approximate h by a trigonometric polynomial T . Finally we shall approximate T by a polynomial p on K .

Take $\phi: \mathbb{R}^d \rightarrow \mathbb{R}^+$ to be a C^∞ function with compact support such that $\int_{\mathbb{R}^d} \phi = 1$, and define for $\delta > 0$, $\phi_\delta: \mathbb{T}^d \rightarrow \mathbb{R}^+$ by

$$\phi_\delta(x) = \frac{1}{\delta^d} \sum_{\xi \in \mathbb{Z}^d} \phi\left(\frac{x + 2\pi\xi}{\delta}\right).$$

Then $\phi_\delta: \mathbb{T}^d \rightarrow \mathbb{R}^+$ is C^∞ , and $\int_{\mathbb{T}^d} \phi_\delta = 1$. Define $g_\delta: \mathbb{T}^d \rightarrow \mathbb{C}$ by

$$g_\delta(x) = \int_{\mathbb{T}^d} \phi_\delta(x - y)g(y) dy.$$

By differentiation under the integral sign, it follows that g_δ is C^∞ . By a change of variables $g_\delta(x) = \int_{\mathbb{T}^d} \phi_\delta(y)g(x - y) dy$. So for $|\alpha| \leq N$, we have by differentiation under the integral sign that $\partial^\alpha g_\delta(x) = \int_{\mathbb{T}^d} \phi_\delta(y)\partial^\alpha g(x - y) dy$. Now since $\int_{\mathbb{T}^d} \phi_\delta = 1$, we deduce that

$$\partial^\alpha g_\delta(x) - \partial^\alpha g(x) = \int_{\mathbb{T}^d} \phi_\delta(y)(\partial^\alpha g(x - y) - \partial^\alpha g(x)) dy$$

Therefore

$$\begin{aligned} |\partial^\alpha g_\delta(x) - \partial^\alpha g(x)| &\leq \int_{\mathbb{T}^d} \phi_\delta(y)|\partial^\alpha g(x - y) - \partial^\alpha g(x)| dy \\ &\leq \sup_{y \in \text{supp } \phi_\delta} |\partial^\alpha g(x - y) - \partial^\alpha g(x)|. \end{aligned}$$

Now $\partial^\alpha g$ is uniformly continuous on $\mathbb{T}^d$, and as $\delta \downarrow 0$, $\text{supp } \phi_\delta$ is eventually contained in any given neighbourhood of 0 in $\mathbb{T}^d$. Thus $\sup |\partial^\alpha g_\delta - \partial^\alpha g| \rightarrow 0$ as $\delta \downarrow 0$. Take $\delta > 0$ such that $|\partial^\alpha g_\delta - \partial^\alpha g| \leq \frac{\varepsilon}{3}$ on all of $\mathbb{T}^d$ when $|\alpha| \leq N$, and put $h = g_\delta$.

Suppose that $u: \mathbb{T}^d \rightarrow \mathbb{C}$ is C^∞ . Let $(c_\xi)_{\xi \in \mathbb{Z}^d}$ be the Fourier coefficients of u . These are given by

$$c_\xi = \frac{1}{(2\pi)^d} \int_{\mathbb{T}^d} u(x) e^{-ix \cdot \xi} dx.$$

By integration by parts we have that

$$(\dagger) \quad (i\xi)^\alpha c_\xi = \frac{1}{(2\pi)^d} \int_{\mathbb{T}^d} \partial^\alpha u(x) e^{-ix \cdot \xi} dx \quad \text{for all } \alpha \in \mathbb{N}^d.$$

Thus $\sup_\xi |\xi^\alpha c_\xi| \leq \frac{1}{(2\pi)^d} \int_{\mathbb{T}^d} |\partial^\alpha u| < \infty$ for all $\alpha \in \mathbb{N}^d$. Therefore $\sum_\xi |c_\xi| < \infty$. So we see that $\sum_\xi c_\xi e^{ix \cdot \xi}$ converges uniformly with respect to x . But since $(x \mapsto e^{ix \cdot \xi})_\xi$ is a complete orthonormal basis for $L^2(\mathbb{T}^d)$ (with inner product $\langle v, w \rangle = \frac{1}{(2\pi)^d} \int_{\mathbb{T}^d} v \bar{w}$) this sum must converge to u . Thus we have shown that

$$u(x) = \sum_{\xi \in \mathbb{Z}^d} c_\xi e^{ix \cdot \xi} \quad \text{uniformly with respect to } x.$$

We have, by applying this to $\partial^\alpha u$ and using $(\dagger)$, that for each $\alpha \in \mathbb{N}^d$,

$$(\ddagger) \quad \partial^\alpha u(x) = \sum_{\xi \in \mathbb{Z}^d} (i\xi)^\alpha c_\xi e^{ix \cdot \xi} \quad \text{uniformly with respect to } x.$$

Now take $u = h$, and using $(\ddagger)$, take F to be a finite subset of $\mathbb{Z}^d$ such that $T: \mathbb{T}^d \rightarrow \mathbb{C}$, defined by $T(x) = \sum_{\xi \in F} c_\xi e^{ix \cdot \xi}$, has the property that $|\partial^\alpha h - \partial^\alpha T| \leq \frac{\varepsilon}{3}$ on $\mathbb{T}^d$ when $|\alpha| \leq N$.

To complete the proof consider T as a function on $\mathbb{R}^d$. From the power series expansion for e^x , we obtain that

$$T(x) = \sum_{\xi \in F} \left(c_\xi \sum_{k=0}^{\infty} \frac{1}{k!} (ix \cdot \xi)^k \right).$$

Let

$$q_n(x) = \sum_{\xi \in F} \left(c_\xi \sum_{k=0}^n \frac{1}{k!} (ix \cdot \xi)^k \right).$$

Then $\partial^\alpha q_n \rightarrow \partial^\alpha T$ uniformly on compact subsets of $\mathbb{R}^d$ for all α . Take $p = q_n$, choosing n so that $|\partial^\alpha T - \partial^\alpha q_n| \leq \frac{\varepsilon}{3}$ on K when $|\alpha| \leq N$. Now we have the result, since on K , $|\partial^\alpha f - \partial^\alpha p| \leq |\partial^\alpha g - \partial^\alpha h| + |\partial^\alpha h - \partial^\alpha T| + |\partial^\alpha T - \partial^\alpha p| \leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon$, when $|\alpha| \leq N$. (Note that $f = g$ in a neighbourhood of K . So $\partial^\alpha f = \partial^\alpha g$ on K .) $\square$

Proof of (6.10): Consider D and S as fixed, and let $\mathcal{A}$ denote the set of all C^2 functions $f: D \rightarrow \mathbb{R}$ such that for all $t \in \mathbb{R}^+$,

$$(1) \quad f(S_t) - f(S_0) = \sum_{i=1}^d \int_0^t \partial_i f(S) dS^i + \frac{1}{2} \sum_{i,j=1}^d \int_0^t \partial_i \partial_j f(S) d[S^i, S^j].$$

Then clearly $\mathcal{A}$ is a vector subspace of $C^2(D; \mathbb{R})$. I claim further that $fg \in \mathcal{A}$ whenever $f, g \in \mathcal{A}$. Suppose then that $f, g \in \mathcal{A}$, and let $F = f(S)$ and $G = g(S)$. Then

$$(2) \quad F_t - F_0 = \sum_{i=1}^d \int_0^t \partial_i f(S) dS^i + \frac{1}{2} \sum_{i,j=1}^d \int_0^t \partial_i \partial_j f(S) d[S^i, S^j],$$

$$(3) \quad G_t - G_0 = \sum_{i=1}^d \int_0^t \partial_i g(S) dS^i + \frac{1}{2} \sum_{i,j=1}^d \int_0^t \partial_i \partial_j g(S) d[S^i, S^j].$$

Thus we see that F and G are integrators, and that

$$(4) \quad \int_0^t F dG = \sum_{i=1}^d \int_0^t f(S) \partial_i g(S) dS^i + \frac{1}{2} \sum_{i,j=1}^d \int_0^t f(S) \partial_i \partial_j g(S) d[S^i, S^j],$$

$$(5) \quad \int_0^t G dF = \sum_{i=1}^d \int_0^t g(S) \partial_i f(S) dS^i + \frac{1}{2} \sum_{i,j=1}^d \int_0^t g(S) \partial_i \partial_j f(S) d[S^i, S^j],$$

$$(6) \quad [F, G]_t = \sum_{i,j=1}^d \int_0^t \partial_i f(S) \partial_j g(S) d[S^i, S^j].$$

((4) and (5) follow from (2) and (3) using Theorem(4.21). (6) follows from (2) and (3) using Theorem(6.8) together with Proposition(6.6).) Therefore the integration by parts formula Proposition(6.3)(iii) gives that

$$\begin{aligned}
F_t G_t - F_0 G_0 &= \int_0^t F dG + \int_0^t G dF + [F, G]_t \\
&= \sum_{i=1}^d \int_0^t (f(S) \partial_i g(S) + g(S) \partial_i f(S)) dS^i \\
&\quad + \frac{1}{2} \sum_{i,j=1}^d \int_0^t (f(S) \partial_i \partial_j g(S) + \partial_i f(S) \partial_j g(S) \\
&\quad \quad + \partial_j f(S) \partial_i g(S) + g(S) \partial_i \partial_j f(S)) d[S^i, S^j] \\
&= \sum_{i=1}^d \int_0^t \partial_i (fg)(S) dS^i + \frac{1}{2} \sum_{i,j=1}^d \int_0^t \partial_i \partial_j (fg)(S) d[S^i, S^j].
\end{aligned}$$

So we have that $fg \in \mathcal{A}$ as required.

Now the constant functions and the projections onto the coordinates are all elements of $\mathcal{A}$. So it follows that all (restrictions to D of) polynomial functions $p: \mathbb{R}^d \rightarrow \mathbb{R}$ are in $\mathcal{A}$. Now suppose that $f: D \rightarrow \mathbb{R}$ is any C^2 function. Then by Lemma(6.11), there exist polynomial functions $p_1, p_2, \dots: \mathbb{R}^d \rightarrow \mathbb{R}$ such that $\partial^\alpha p_n \rightarrow \partial^\alpha f$ uniformly on compact subsets of D when $|\alpha| \leq 2$. We have that

$$(7) \quad p_n(S_t) - p_n(S_0) = \sum_{i=1}^d \int_0^t \partial_i p_n(S) dS^i + \frac{1}{2} \sum_{i,j=1}^d \int_0^t \partial_i \partial_j p_n(S) d[S^i, S^j].$$

We intend to let $n \rightarrow \infty$ in (7). Consider the first term on the right hand side. Fix i , and let $U_{i,n} = \partial_i p_n(S) - \partial_i f(S)$. Then $U_{i,n}^* \rightarrow 0$ pointwise as $n \rightarrow \infty$ (since $\{S_u(\omega) \mid 0 \leq u \leq s\}$ is a compact subset of D for all $s \in \mathbb{R}^+$ and $\omega \in \Omega$). Therefore by Theorem(4.24)(ii), $\Phi_{S^i}^0(\mathbf{1}_{(0,t]} \times \Omega U_{i,n}) \rightarrow 0$. Thus $\int_0^t U_{i,n} dS^i \rightarrow 0$ in probability. I.e. $\int_0^t \partial_i p_n(S) dS^i \rightarrow \int_0^t \partial_i f(S) dS^i$ in probability as $n \rightarrow \infty$. The second term of the right hand side of (7) is similar (or even can be dealt with as pathwise Stieltjes integrals). So letting $n \rightarrow \infty$ in (7) gives (1). $\square$

Now to state and prove the general Itô formula (for discontinuous processes) we need another definition.

(6.12) Definition.

- (a) Suppose that $\alpha: \mathbb{R}^+ \rightarrow \mathbb{R}$ is càdlàg and has finite variation. Then α is said to *consist purely of jumps* if and only if $\alpha(t) = \sum_{s \leq t} \Delta \alpha(s)$ (where $\Delta \alpha(s) = \alpha(s) - \alpha(s-)$ as for processes) for all $t \in \mathbb{R}^+$.
- (b) A *jump process* is a finite-variation process whose paths consist purely of jumps.

□

Note.

- 1. In (a) the sum is a countable sum, and converges absolutely, since α has finite variation.
- 2. From (a), we must have that $\alpha(0) = 0$ if α consists purely of jumps.
- 3. If α consists purely of jumps, then clearly $\Delta \alpha$ completely determines α .
- 4. If $\alpha: \mathbb{R}^+ \rightarrow \mathbb{R}$ is càdlàg with finite variation, then on defining $\alpha_J: \mathbb{R}^+ \rightarrow \mathbb{R}$ by $\alpha_J(t) = \sum_{s \leq t} \Delta \alpha(s)$, and putting $\alpha_C = \alpha - \alpha_J$, we have that α_C is a continuous function of finite variation, α_J consists purely of jumps, and $\alpha = \alpha_C + \alpha_J$. Any such decomposition is unique, since given such a decomposition we must have $\Delta \alpha_J = \Delta \alpha$ (since $\Delta \alpha_C = 0$).

(6.13) Theorem (The Itô Formula).

Suppose that $S = (S^1, \dots, S^d): \mathbb{R}^+ \times \Omega \rightarrow D$, where D is an open subset of $\mathbb{R}^d$, $S^1, \dots, S^d$ are integrators, and the range of S_- is contained in D . Suppose that $f: D \rightarrow \mathbb{R}$ is a C^2 function. Then for $t \in \mathbb{R}^+$,

$$f(S_t) - f(S_0) = \sum_{i=1}^d \int_0^t \partial_i f(S_-) dS^i + \frac{1}{2} \sum_{i,j=1}^d \int_0^t \partial_i \partial_j f(S_-) d[S^i, S^j] + J_t,$$

where J is a jump process characterized by

$$\Delta J = \Delta f(S) - \sum_{i=1}^d \partial_i f(S_-) \Delta S^i - \frac{1}{2} \sum_{i,j=1}^d \partial_i \partial_j f(S_-) \Delta S^i \Delta S^j.$$

In particular we see that $f(S)$ is an integrator.

To prove this we shall need the following technical lemma.

(6.14) Lemma. Suppose that D is an open subset of $\mathbb{R}^d$, $\alpha = (\alpha_1, \dots, \alpha_d): \mathbb{R}^+ \rightarrow D$ is càdlàg, $\alpha(t-) \in D$ for all $t \in \mathbb{R}^{++}$, and $\sum_{s \leq t} |\Delta \alpha(s)|^2 < \infty$ for all $t \in \mathbb{R}^+$. Then if we fix $b \in \mathbb{R}^+$, and define for each C^2 function $f: D \rightarrow \mathbb{R}$,

$$J(f) = \sum_{t \leq b} \left| f(\alpha(t)) - f(\alpha(t-)) - \sum_{i=1}^d \partial_i f(\alpha(t-)) \Delta \alpha_i(t) - \frac{1}{2} \sum_{i,j=1}^d \partial_i \partial_j f(\alpha(t-)) \Delta \alpha_i(t) \Delta \alpha_j(t) \right|,$$

we have that J has the following properties.

- (i) $J(f) < \infty$ for all C^2 functions $f: D \rightarrow \mathbb{R}$.
- (ii) If $f_1, f_2, \dots: D \rightarrow \mathbb{R}$ are C^2 functions such that $f_n, \partial_i f_n, \partial_i \partial_j f_n \rightarrow 0$ uniformly on compact subsets of D for all i and j , then $J(f_n) \rightarrow 0$.

Proof: Let $K = \{ \alpha(t) \mid t \in [0, b] \} \cup \{ \alpha(t-) \mid t \in (0, b] \}$. Then K is a compact subset of D (see (1.4)). Let $\delta > 0$ be such that $C \subset D$, where $C = \{ x \in \mathbb{R}^d \mid \inf_{p \in K} |x - p| \leq \delta \}$. C is compact.

$$\text{Let } P = \{ t \in (0, b] \mid \alpha(t) \neq \alpha(t-) \},$$

$$\text{and } P_1 = \{ t \in P \mid |\Delta \alpha(t)| > \delta \},$$

$$\text{and } P_2 = \{ t \in P \mid |\Delta \alpha(t)| \leq \delta \}.$$

So P is the disjoint union of P_1 and P_2 . Now since $\sum_{t \leq b} |\Delta \alpha(t)|^2 < \infty$, we have that P is a countable set and P_1 is a finite set. Define for each C^2 function $f: D \rightarrow \mathbb{R}$, and $k = 1, 2$,

$$J_k(f) = \sum_{t \in P_k} \left| f(\alpha(t)) - f(\alpha(t-)) - \sum_{i=1}^d \partial_i f(\alpha(t-)) \Delta \alpha_i(t) - \frac{1}{2} \sum_{i,j=1}^d \partial_i \partial_j f(\alpha(t-)) \Delta \alpha_i(t) \Delta \alpha_j(t) \right|,$$

Then $J(f) = J_1(f) + J_2(f)$. We shall now obtain an estimate for $J_2(f)$. Let $m(f) = \sup\{|\partial_i \partial_j f(x)| \mid x \in C, \ 1 \leq i, j \leq d\}$. I claim that if $x \in K$ and $h \in \mathbb{R}^d$ such that $|h| \leq \delta$, then

$$(\dagger) \quad \left| f(x+h) - f(x) - \sum_{i=1}^d \partial_i f(x) h_i - \frac{1}{2} \sum_{i,j=1}^d \partial_i \partial_j f(x) h_i h_j \right| \leq d m(f) |h|^2.$$

To prove $(\dagger)$, let $g(t) = f(x + th)$. Then by Taylor's theorem, $g(1) = g(0) + g'(0) + \frac{1}{2} g''(\theta)$ for some $\theta \in (0, 1)$. Now for $0 \leq t \leq 1$, we have that $g''(t) = \sum_{i,j=1}^d \partial_i \partial_j f(x + th) h_i h_j$, and so

$$\begin{aligned} |g''(t)| &\leq \sum_{i,j=1}^d m(f) |h_i h_j| \quad (\text{since } x + th \in C) \\ &\leq \sum_{i,j=1}^d m(f) \left(\frac{1}{2} h_i^2 + \frac{1}{2} h_j^2 \right) \\ &\leq d m(f) |h|^2. \end{aligned}$$

Thus $|g(1) - g(0) - g'(0) - \frac{1}{2} g''(0)| \leq \frac{1}{2} |g''(0)| + \frac{1}{2} |g''(\theta)| \leq d m(f) |h|^2$. This is just $(\dagger)$.

Now for $t \in P_2$, take $x = \alpha(t-)$ and $h = \alpha(t) - \alpha(t-)$ in $(\dagger)$, and sum over all $t \in P_2$ to give that

$$(\dagger) \quad J_2(f) \leq d m(f) \sum_{t \in P_2} |\Delta \alpha(t)|^2.$$

Now the sum is finite by the hypothesis on α . So $J_2(f) < \infty$. But since P_1 is a finite set, $J_1(f) < \infty$. It follows that $J(f) = J_1(f) + J_2(f) < \infty$; thus proving (i). For (ii) we have that $m(f_n) \rightarrow 0$ as $n \rightarrow \infty$, and so $(\dagger)$ gives that $J_2(f_n) \rightarrow 0$. But clearly $J_1(f_n) \rightarrow 0$, since P_1 is a finite set. Thus $J(f_n) = J_1(f_n) + J_2(f_n) \rightarrow 0$ as required. $\square$

Proof of (6.13): Given càdlàg processes $X, Y: \mathbb{R}^+ \times \Omega \rightarrow \mathbb{R}$, write $X \equiv Y$ to mean that $X - Y$ is a jump process. We see that ‘ $\equiv$ ’ is an equivalence relation. Regard S and D as fixed, and let $\mathcal{A}$ denote the set of C^2 functions $f: D \rightarrow \mathbb{R}$ such that

$$(1) \quad f(S) - f(S_0) \equiv \sum_{i=1}^d \partial_i f(S_-) \cdot S^i + \frac{1}{2} \sum_{i,j=1}^d \partial_i \partial_j f(S_-) [S^i, S^j].$$

Then $\mathcal{A}$ is a vector subspace of $C^2(D; \mathbb{R})$. The proof of Theorem(6.10) can be adapted easily (replace certain equals signs by ‘ $\equiv$ ’) to show that $fg \in \mathcal{A}$ whenever $f, g \in \mathcal{A}$. Thus again all polynomial functions lie in $\mathcal{A}$.

Suppose that $f: D \rightarrow \mathbb{R}$ is any C^2 function. Take by Lemma(6.11) a sequence of polynomial functions $p_1, p_2, \dots: \mathbb{R}^d \rightarrow \mathbb{R}$ such that $\partial^\alpha p_n \rightarrow \partial^\alpha f$ uniformly on all compact subsets of D when $|\alpha| \leq 2$. Then since $p_n \in \mathcal{A}$ we have that

$$(2) \quad \begin{aligned} p_n(S_t) - p_n(S_0) &= \sum_{i=1}^d \int_0^t \partial_i p_n(S_-) dS^i \\ &\quad + \frac{1}{2} \sum_{i,j=1}^d \int_0^t \partial_i \partial_j p_n(S_-) d[S^i, S^j] + J_t^{(n)}, \end{aligned}$$

for all $t \in \mathbb{R}^+$, for some jump process $J^{(n)}$. Apply Δ to this equation using Theorem(5.25) and Proposition(6.3)(iv) to give that

$$\Delta p_n(S) = \sum_{i=1}^d \partial_i p_n(S_-) \Delta S^i + \frac{1}{2} \sum_{i,j=1}^d \partial_i \partial_j p_n(S_-) \Delta S^i \Delta S^j + \Delta J^{(n)}.$$

Thus

$$(3) \quad \Delta J^{(n)} = \Delta p_n(S) - \sum_{i=1}^d \partial_i p_n(S_-) \Delta S^i - \frac{1}{2} \sum_{i,j=1}^d \partial_i \partial_j p_n(S_-) \Delta S^i \Delta S^j.$$

Letting $n \rightarrow \infty$ in (2) will prove the result provided, there is a jump process J such that $J_t^{(n)} \rightarrow J_t$ in probability for all $t \in \mathbb{R}^+$, and such that

$$(4) \quad \Delta J = \Delta f(S) - \sum_{i=1}^d \partial_i f(S_-) \Delta S^i - \frac{1}{2} \sum_{i,j=1}^d \partial_i \partial_j f(S_-) \Delta S^i \Delta S^j.$$

Now $\sum_{u \leq t} (\Delta S_u^i)^2 = \sum_{u \leq t} \Delta[S^i]_u < \infty$ a.s. and so Lemma(6.14) applies to (a.a.) paths of S . Thus we can define $J: \mathbb{R}^+ \times \Omega \rightarrow \mathbb{R}$ by

$$(5) \quad J_t = \sum_{u \leq t} \left(\Delta f(S_u) - \sum_{i=1}^d \partial_i f(S_{u-}) \Delta S_u^i - \frac{1}{2} \sum_{i,j=1}^d \partial_i \partial_j f(S_{u-}) \Delta S_u^i \Delta S_u^j \right).$$

Then if we knew that J_t were $\mathcal{F}_t$ -measurable for all t , J would be a càdlàg adapted process, and hence a jump process satisfying (4). Now from Lemma(6.14)(ii) applied with $p_n - f$ as f_n and each path of S as α , we obtain that $|J_t^{(n)} - J_t| \rightarrow 0$ a.s. for all t . I.e. $J_t^{(n)} \rightarrow J_t$ a.s. for all t . Thus we see that J_t is $\mathcal{F}_t$ -measurable for all t . But also this shows that J is the required jump process. $\square$

(6.15) Example (Brownian Motion).

Suppose that $B = (B^1, \dots, B^d): \mathbb{R}^+ \times \Omega \rightarrow \mathbb{R}^d$ is a d -dimensional Brownian motion. Then we have seen (Proposition(6.4)(ii)) that $[B^i, B^j]_t = \delta_{i,j}t$ and so the Itô formula (Theorem(6.10)) gives that for every C^2 function $f: \mathbb{R}^d \rightarrow \mathbb{R}$

$$f(B_t) - f(B_0) = \sum_{i=1}^d \int_0^t \partial_i f(B) dB^i + \frac{1}{2} \int_0^t \Delta f(B_s) ds,$$

where here Δ denotes the Laplacian operator $\sum_{i=1}^d \partial_i^2$ and not the jumps (I apologize for this abuse of notation). $\square$

§6.5 THE DIFFERENTIAL NOTATION

It is a notational convenience sometimes to write $dX = H dS$ as shorthand for $X_t = X_0 + \int_0^t H dS$. In this short section we shall make this notation precise. We shall suppose that $\text{b}\mathcal{P}$ denotes the vector space of all bounded previsible processes $X: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}$, and so $\text{lb}\mathcal{P}$ denotes the vector space of all local bounded previsible processes $X: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}$. Let $\mathcal{I}$ denote the class of all integrators. We shall use the convention that indistinguishable processes are the same. Let $d\mathcal{I} = \mathcal{I}/\mathcal{C}$, where $\mathcal{C}$ is the subspace of $\mathcal{I}$ consisting of those processes with constant paths (i.e. the processes $S \in \mathcal{I}$ such that $S_t = S_0$ for all $t \in \mathbb{R}^+$). Let $d: \mathcal{I} \rightarrow d\mathcal{I}$ denote the quotient map (so d ‘forgets’ the starting point). Now make the following definition.

- (1) $\mathcal{A} = \text{lb}\mathcal{P} \oplus d\mathcal{I}$ as vector spaces.

Define a multiplication on $\mathcal{A}$ by the following formula.

- (2) $(X + dR)(Y + dS) = XY + d(X \cdot S + Y \cdot R + [R, S])$ for all $X, Y \in \text{lb}\mathcal{P}$ and $R, S \in \mathcal{I}$.

I.e. so that it extends the pointwise multiplication on $\text{lb}\mathcal{P}$, and the following properties hold.

- (3) $X dS = d(X \cdot S)$ for all $X \in \text{lb}\mathcal{P}$ and $S \in \mathcal{I}$.

- (4) $dR dS = d[R, S]$ for all $R, S \in \mathcal{I}$.

Now this makes $\mathcal{A}$ into a commutative algebra over $\mathbb{R}$. The only difficult thing to check is that the multiplication given by (2) is associative. But this amounts to proving the following properties.

- (5) $X(Y dS) = (XY) dS$ for all $X, Y \in \text{lb}\mathcal{P}$ and $S \in \mathcal{I}$.

- (6) $X(dR dS) = (X dR) dS$ for all $X \in \text{lb}\mathcal{P}$ and $R, S \in \mathcal{I}$.

- (7) $dR(dS dT) = (dR dS) dT$ for all $R, S, T \in \mathcal{I}$.

But (5) and (6) follow from Theorem(4.21)(v) and Theorem(6.8) (see Note 2 after the statement) respectively. For (7) we have to show that $[R, [S, T]] = [[R, S], T]$. But

since $[S, T]$ has finite variation, Proposition(6.6) gives that

$$\begin{aligned} [R, [S, T]]_t &= \sum_{u \leq t} \Delta R_u \Delta [S, T]_u \\ &= \sum_{u \leq t} \Delta R_u \Delta S_u \Delta T_u, \end{aligned}$$

which is clearly symmetric in R , S and T . The 1 of the algebra $\mathcal{A}$ is the previsible process 1.

It is clear that $\text{lb}\mathcal{P}$ is a subalgebra of $\mathcal{A}$, and that $d\mathcal{I}$ is an ideal of $\mathcal{A}$.

Now we can state the Itô formula as follows. Suppose we have S , f , D etc. as in Theorem(6.10). Then

$$df(S) = \sum_{i=1}^d \partial_i f(S) dS^i + \frac{1}{2} \sum_{i,j=1}^d \partial_i \partial_j f(S) dS^i dS^j.$$

We shall also write $dR_t = H_t dS_t$ to mean $dR = H dS$ when this is convenient.

So for example with B , f etc. as in Example(6.15) we have

$$df(B_t) = \sum_{i=1}^d \partial_i f(B_t) dB_t^i + \frac{1}{2} \Delta f(B_t) dt.$$

(We use dt as shorthand for dI_t , when $I_t = t$ for all t .)

CHAPTER 7: DISCRETE-TIME MARTINGALES

In order to study continuous time martingales (the subject of the next chapter) we shall require some results from the discrete time case. In this chapter we present these results. This approach to the proof that martingales are integrators is due to D. A. Edwards [Ed].

§7.1 CONDITIONAL EXPECTATION

Suppose that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space. Then suppose we have a σ -field $\mathcal{G}$ on Ω with $\mathcal{G} \subseteq \mathcal{F}$, and an L^1 real-valued random variable X . Then $A \mapsto \int_A X d\mathbb{P}: \mathcal{G} \rightarrow \mathbb{R}$ is a signed measure on the measurable space $(\Omega, \mathcal{G})$. This measure is absolutely continuous with respect to $\mathbb{P}|_{\mathcal{G}}$. Thus by the Radon-Nikodym theorem, it follows that there exists an L^1 $\mathcal{G}$ -measurable real-valued random variable Y such that $\int_A X d\mathbb{P} = \int_A Y d\mathbb{P}$ for all $A \in \mathcal{G}$. Moreover Y is unique up to equality $\mathbb{P}$ -a.e. This Y is called the conditional expectation of X with respect to $\mathcal{G}$. I.e. we make the following definition.

(7.1)Definition. Suppose that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, $\mathcal{G}$ is a σ -field on Ω with $\mathcal{G} \subseteq \mathcal{F}$ and X is an L^1 real-valued random variable. Then the *conditional expectation* of X with respect to $\mathcal{G}$ is the unique L^1 $\mathcal{G}$ -measurable real-valued random variable $\mathbb{E}(X|\mathcal{G})$ such that

$$(\dagger) \quad \int_A \mathbb{E}(X|\mathcal{G}) d\mathbb{P} = \int_A X d\mathbb{P} \quad \text{for all } A \in \mathcal{G}.$$

□

(7.2)Properties. Suppose that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space. Then we see that conditional expectation satisfies the following properties. (Where we suppose if not stated otherwise that $\mathcal{G}$ is a σ -field such that $\mathcal{G} \subseteq \mathcal{F}$.)

- (1) $E(\cdot|\mathcal{G})$ is linear.
- (2) $E(X|\mathcal{G}) \leq E(Y|\mathcal{G})$ whenever $X \leq Y$ in L^1 .
- (3) If $X \in L^1$ is $\mathcal{G}$ -measurable, then $E(X|\mathcal{G}) = X$.
- (4) $|E(X|\mathcal{G})| \leq E(|X||\mathcal{G})$ for all $X \in L^1$.
- (5) (The Tower Property) If $\mathcal{G}$ and $\mathcal{H}$ are σ -fields such that $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{F}$, then $E[E(X|\mathcal{G})|\mathcal{H}] = E(X|\mathcal{H})$ for all $X \in L^1$.
- (6) For each $X \in L^1$, $E(X|\{\emptyset, \Omega\})$ is constant and takes the value $E(X)$.
- (7) $E[E(X|\mathcal{G})] = E(X)$ for all $X \in L^1$.
- (8) If X is $\mathcal{G}$ -measurable, and $Y \in L^1$ such that also $XY \in L^1$, then $E(XY|\mathcal{G}) = XE(Y|\mathcal{G})$.
(Let μ and ν be the measures given by $d\mu = Y d\mathbb{P}$ and $d\nu = XY d\mathbb{P}$. Then $X \in L^1(\mu)$ and $d\nu = X d\mu$. Thus $E(XY|\mathcal{G}) d\mathbb{P}|_{\mathcal{G}} = d\nu|_{\mathcal{G}} = X d\mu|_{\mathcal{G}} = XE(Y|\mathcal{G}) d\mathbb{P}|_{\mathcal{G}}.$)
- (9) (Jensen's Inequality) If $\phi: \mathbb{R} \rightarrow \mathbb{R}$ is convex and $X \in L^1$ such that also $\phi \circ X \in L^1$, then $\phi \circ E(X|\mathcal{G}) \leq E(\phi \circ X|\mathcal{G})$.
(The proof is a trivial modification of the proof of Jensen's inequality for $E(\cdot)$.)
- (10) Suppose that $1 \leq p \leq \infty$. Then for each $X \in L^p$, $E(X|\mathcal{G}) \in L^p$ and $\|E(X|\mathcal{G})\|_p \leq \|X\|_p$.
(When $p < \infty$, apply Jensen's inequality (9) with $\phi(x) = |x|^p$.)
- (11) (The Monotone Convergence Theorem) If $X_1, X_2, \dots \in L^1$, $X \in L^1$, and $X_n \uparrow X$ a.s., then $E(X_n|\mathcal{G}) \uparrow E(X|\mathcal{G})$ a.s. □

§7.2 DISCRETE-TIME PROCESSES

(7.3)Definition. If $(\Omega, \mathcal{F}, \mathbb{P})$ is any probability space, then a *discrete-time filtration* on $(\Omega, \mathcal{F}, \mathbb{P})$ is a family $(\mathcal{F}_n)_{n \in \mathbb{N}}$ of σ -fields on Ω such that $\mathcal{F}_n \subseteq \mathcal{F}_m$ whenever $n \leq m$, and $\mathcal{F}_n \subseteq \mathcal{F}$ for all n . □

Throughout the rest of this chapter we shall assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space with a given discrete-time filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$ on it.

(7.4) Definition.

- (a) An (*adapted*) *process* is a family of random variables $X = (X_n)_{n \in \mathbb{N}}$ such that X_n is $\mathcal{F}_n$ -measurable for all n . As before the term ‘process’ will be taken to mean adapted process. We shall say that X is an L^p -*process* ($0 \leq p \leq \infty$) if and only if X is a process such that $X_n \in L^p$ for all n . A process X is *bounded* if and only if $\|X\|_\infty < \infty$, where $\|X\|_\infty = \sup_n \|X_n\|_\infty$.
- (b) A *stopping time* is a function $\sigma: \Omega \rightarrow \mathbb{N} \cup \{\infty\}$ such that $[\sigma \leq n] \in \mathcal{F}_n$ for all $n \in \mathbb{N}$. As in Chapter 5, a stopping time σ is said to be *finite* if and only if $\sigma < \infty$, and is said to be *bounded* if and only if $\sigma \leq N$ for some $N \in \mathbb{N}$.
- (c) If σ is a stopping time then the σ -field $\mathcal{F}_\sigma$ is defined by

$$\mathcal{F}_\sigma = \{A \in \mathcal{F} \mid A \cap [\sigma \leq n] \in \mathcal{F}_n \text{ for all } n \in \mathbb{N}\}.$$

- (d) If σ is a finite stopping time and X is a process, then the random variable X_σ is defined by $X_\sigma(\omega) = X_{\sigma(\omega)}(\omega)$. X_σ is $\mathcal{F}_\sigma$ -measurable.
- (e) A *martingale* (resp. *submartingale*, *supermartingale*) is an L^1 -process X such that $X_n = E(X_m | \mathcal{F}_n)$ (resp. $X_n \leq E(X_m | \mathcal{F}_n)$, $X_n \geq E(X_m | \mathcal{F}_n)$) whenever $n \leq m$. An L^p -*martingale* ($1 \leq p \leq \infty$) is an L^p -process that is also a martingale (and similarly for submartingales and supermartingales).
- (f) A *previsible process* is a family of random variables $H = (H_n)_{n \in \mathbb{N}}$ such that H_n is $\mathcal{F}_{n-1}$ -measurable for all $n \in \mathbb{N}$ (where $\mathcal{F}_{-1}$ is taken to be $\mathcal{F}_0$). □

Note.

1. We have analogues of some of the results given in §5.1. For example we have the following results.

- (i) If σ and τ are stopping times, then so are $\sigma \wedge \tau$ and $\sigma \vee \tau$.
- (ii) If $n \in \mathbb{N}$ and $\tau(\omega) = n$ for all $\omega \in \Omega$, then τ is a stopping time and $\mathcal{F}_\tau = \mathcal{F}_n$.
- (iii) If σ and τ are stopping times such that $\sigma \leq \tau$, then $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$.

2. We have clearly that an L^1 -process X is a martingale (resp. submartingale, supermartingale) if and only if $\mathbb{E}(X_m - X_n | \mathcal{F}_n) = 0$ (resp. $\geq 0, \leq 0$) whenever $n \leq m$.

§7.3 THE MARTINGALE TRANSFORM AND MUTUAL VARIATION

For discrete-time processes we have the following analogues of the stochastic integral and mutual variation.

(7.5) Definition.

- (a) If X is a process and H is a previsible process, then we define the process $H * X$, called the *martingale transform* of X with respect to H , by $(H * X)_n = H_0 X_0 + \sum_{k=1}^n H_k (X_k - X_{k-1})$. It is easy to see that this is adapted.
- (b) If X and Y are processes, then the *mutual variation process* $[X, Y]$ is given by $[X, Y]_n = X_0 Y_0 + \sum_{k=0}^{n-1} (X_{k+1} - X_k)(Y_{k+1} - Y_k)$. We shall call $[X, X]$ the *quadratic variation* of X and denote it by $[X]$. $\square$

We have that $H * X$ and $[X, Y]$ are linear in both variables. Also the following proposition (the proof of which is a trivial piece of algebra and will be left to the reader) is the discrete-time analogue of Theorem(4.21) and Theorem(6.8).

(7.6) Proposition.

- (i) If X is a process and H and K are previsible processes, then $H * (K * X) = (HK) * X$.
- (ii) If X and Y are processes, and H and K are previsible processes, then $[H * X, K * Y] = (HK) * [X, Y]$. Putting in particular $K = 1$ gives that $[H * X, Y] = H * [X, Y]$. $\square$

Now we shall consider the connections between the martingale transform, mutual variation and martingales.

(7.7) Proposition. Suppose that $1 \leq p, q \leq \infty$ are such that $\frac{1}{p} + \frac{1}{q} \leq 1$, and that H is an L^p -previsible process. Then taking r such that $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$ (and so $1 \leq r \leq \infty$), the following results hold.

- (i) If M is an L^q -martingale, then $H * M$ is an L^r -martingale.
- (ii) If $H \geq 0$ and X is a L^q -submartingale (resp. L^q -supermartingale), then $H * X$ is a L^r -submartingale (resp. L^r -supermartingale).

Proof: Note that from the inequality $\|UV\|_r \leq \|U\|_p \|V\|_q$ (see Proposition(1.9)(a)), it follows that $H * X$ is an L^r -process whenever X is an L^q -process.

(i) We have that $(H * M)_{n+1} - (H * M)_n = H_{n+1}(M_{n+1} - M_n)$. Therefore it follows that $\mathbb{E}[(H * M)_{n+1} - (H * M)_n | \mathcal{F}_n] = H_{n+1} \mathbb{E}(M_{n+1} - M_n | \mathcal{F}_n) = 0$ (since H_{n+1} is $\mathcal{F}_n$ -measurable).

(ii) We have that $\mathbb{E}[(H * X)_{n+1} - (H * X)_n | \mathcal{F}_n] = H_{n+1} \mathbb{E}(X_{n+1} - X_n | \mathcal{F}_n) \geq 0$ (resp. ≤ 0). □

(7.8) Proposition. Suppose that $1 \leq p, q \leq \infty$ with $\frac{1}{p} + \frac{1}{q} \leq 1$, M is an L^p -martingale, and that N is an L^q -martingale. Then taking r such that $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$, we have that $MN - [M, N]$ is an L^r -martingale.

Proof: Let $X = MN - [M, N]$. Then $X_{n+1} - X_n = (M_{n+1}N_{n+1} - M_nN_n) - (M_{n+1} - M_n)(N_{n+1} - N_n) = M_n(N_{n+1} - N_n) + N_n(M_{n+1} - M_n)$. Thus on defining previsible processes H and K by $H_0 = K_0 = 0$, $H_n = M_{n-1}$ for $n > 0$, and $K_n = N_{n-1}$ for $n > 0$, we see that $X = H * N + K * M$. Thus by Proposition(7.7)(i), X is an L^r -martingale as required. □

§7.4 SOME GENERAL RESULTS

In this section we shall present some well-known results, which we shall need later.

(7.9) Theorem (Optional Stopping).

Suppose that X is a martingale (resp. submartingale, supermartingale), and σ and τ are bounded stopping times such that $\sigma \leq \tau$. Then $X_\sigma, X_\tau \in L^1$ and $X_\sigma = \mathbb{E}(X_\tau | \mathcal{F}_\sigma)$ (resp. $X_\sigma \leq \mathbb{E}(X_\tau | \mathcal{F}_\sigma)$, $X_\sigma \geq \mathbb{E}(X_\tau | \mathcal{F}_\sigma)$).

Proof: We consider only the case when X is a martingale; the other cases being similar. Take $N \in \mathbb{N}$ such that $0 \leq \sigma \leq \tau \leq N$. We have that $|X_\sigma|, |X_\tau| \leq \sum_{n=0}^N |X_n|$, and so $X_\sigma, X_\tau \in L^1$. Suppose that $A \in \mathcal{F}_\sigma$, and define a previsible process H by $H_n = \mathbf{1}_{A \cap [\sigma < n \leq \tau]}$ ($A \cap [\sigma < n \leq \tau] = (A \cap [\sigma \leq n-1]) \setminus [\tau \leq n-1] \in \mathcal{F}_{n-1}$ for all n , so H really is previsible). We see therefore by Proposition(7.7)(i) (with $p = \infty$ and $q = r = 1$) that $H * X$ is a martingale. Thus $0 = \mathbb{E}[(H * X)_0] = \mathbb{E}[(H * X)_N] = \mathbb{E}[\mathbf{1}_A(X_\tau - X_\sigma)]$. I.e. $\int_A X_\sigma d\mathbb{P} = \int_A X_\tau d\mathbb{P}$ for all $A \in \mathcal{F}_\sigma$. So $X_\sigma = \mathbb{E}(X_\tau | \mathcal{F}_\sigma)$ as required. $\square$

In a similar fashion we have the following result.

(7.10) Proposition. Suppose that X is a martingale (resp. submartingale, supermartingale), and σ is any stopping time. Then $(X_{\sigma \wedge n})_n$ is also a martingale (resp. submartingale, supermartingale). Here we mean with respect to the given filtration and **not** $(\mathcal{F}_{\sigma \wedge n})_n$.

Proof: Let $H_n = \mathbf{1}_{[\sigma \geq n]}$. Then $[\sigma \geq n] = \Omega \setminus [\sigma \leq n-1] \in \mathcal{F}_{n-1}$ and so $H = (H_n)_n$ is a non-negative previsible process. Now $X_{\sigma \wedge n} = (H * X)_n$. Thus the result now follows by using Proposition(7.7). $\square$

(7.11) The Doob-Meyer Decomposition.

Suppose that X is an L^1 -process. Then there exist unique M and A such that $X = M + A$, with M a martingale, and A an L^1 -previsible process with $A_0 = 0$. Given this decomposition, which is called the Doob-Meyer decomposition of X , A is increasing (resp. decreasing) if and only if X is a submartingale (resp. supermartingale).

Proof: If M and A satisfy the requirements of the Doob-Meyer decomposition of X , then by applying $\mathbb{E}(\cdot|\mathcal{F}_n)$ to the equation $(M_{n+1} - M_n) + (A_{n+1} - A_n) = X_{n+1} - X_n$, we obtain that

$$(\dagger) \quad A_{n+1} - A_n = \mathbb{E}(X_{n+1} - X_n|\mathcal{F}_n) \quad \text{for all } n \in \mathbb{N}.$$

Thus summing up gives that

$$(\ddagger) \quad A_n = \sum_{k=0}^{n-1} \mathbb{E}(X_{k+1} - X_k|\mathcal{F}_k).$$

So the uniqueness of M and A has been demonstrated.

For the existence of M and A , define A by $(\ddagger)$, and put $M = X - A$. From $(\ddagger)$, it is clear that A is previsible with $A_0 = 0$. Also from $(\ddagger)$ follows $(\dagger)$. From $(\dagger)$ we deduce that $\mathbb{E}(M_{n+1} - M_n|\mathcal{F}_n) = 0$, and hence that M is a martingale. The last statement of the result follows from $(\dagger)$. $\square$

When considering a supermartingale X , we see from (7.11), that $X = M - A$ for some martingale M and some increasing previsible process A with $A_0 = 0$. We shall also refer to $X = M - A$ as the Doob-Meyer decomposition of X .

§7.5 SOME ESTIMATES FOR MARTINGALE TRANSFORMS

Our aim in this section is to prove inequalities of the form $\|(H*M)_n\|_p \leq C_p \|H\|_\infty \|M_n\|_p$ for H a previsible process, M an L^p -martingale and $n \in \mathbb{N}$. The first to be considered is the case when $p = 2$.

(7.12) Proposition. *Suppose that M is an L^2 -martingale and that H is a bounded previsible process. Then $H*M$ is an L^2 -martingale and*

$$\|(H*M)_n\|_2 \leq \|H\|_\infty \|M_n\|_2 \quad \text{for all } n \in \mathbb{N}.$$

Proof: It follows from Proposition(7.7)(i), that $H*M$ is an L^2 -martingale. Now by Proposition(7.8), we have that $(H*M)^2 - [H*M]$ is a martingale which is zero at time zero. So $\mathbb{E}[(H*M)_n^2] = \mathbb{E}([H*M]_n)$. Similarly $\mathbb{E}(M_n^2) = \mathbb{E}([M]_n)$. But $[H*M] = H^2*[M] \leq \|H\|_\infty^2[M]$, by Proposition(7.6)(ii), and the fact that $[M]$ is increasing. Thus it follows that $\mathbb{E}[(H*M)_n^2] \leq \|H\|_\infty^2 \mathbb{E}(M_n^2)$. Take square roots to give the result. $\square$

Next we consider the case when $p = 1$. In this case we only get a weak version of the inequality (see Proposition(7.14)(ii)). In order to prove this result we shall need a preliminary result.

(7.13)Proposition. *Suppose that Z is a supermartingale, $0 < c < \infty$ such that $0 \leq Z \leq c$, and that H is a bounded previsible process. Then*

$$\mathbb{E}[(H*Z)_n^2] \leq 8c\|H\|_\infty^2 \mathbb{E}(Z_0) \quad \text{for all } n \in \mathbb{N}.$$

Proof: Fix n . We may assume that $H_k = Z_k = 0$ for all $k > n$, since on putting $H_k = Z_k = 0$ for $k > n$, the value of $(H*Z)_n$ is not changed, $\|H\|_\infty$ is not made bigger, and Z remains a supermartingale.

Let $Z = M - A$ be the Doob-Meyer decomposition of Z . Then by construction M and A are L^∞ -processes. We have that $M = Z + A$ and so $M^2 = Z^2 + 2ZA + A^2$. Thus

$$\begin{aligned} (1) \quad \mathbb{E}(M_{k+1}^2 - M_k^2 | \mathcal{F}_k) &= \mathbb{E}(Z_{k+1}^2 - Z_k^2 | \mathcal{F}_k) + 2\mathbb{E}(Z_{k+1}A_{k+1} - Z_kA_k | \mathcal{F}_k) \\ &\quad + A_{k+1}^2 - A_k^2. \end{aligned}$$

(Recall that A is previsible.) Now $\mathbb{E}(Z_{k+1} | \mathcal{F}_k) = Z_k + \mathbb{E}(Z_{k+1} - Z_k | \mathcal{F}_k) = Z_k - (A_{k+1} - A_k)$. Thus $\mathbb{E}(Z_{k+1}A_{k+1} - Z_kA_k | \mathcal{F}_k) = \mathbb{E}(Z_{k+1} | \mathcal{F}_k)A_{k+1} - Z_kA_k = Z_k(A_{k+1} - A_k) -$

$A_{k+1}(A_{k+1} - A_k)$. Substituting this in (1) gives that

$$\begin{aligned}
 (2) \quad \mathbb{E}(M_{k+1}^2 - M_k^2 | \mathcal{F}_k) &= \mathbb{E}(Z_{k+1}^2 - Z_k^2 | \mathcal{F}_k) + 2Z_k(A_{k+1} - A_k) \\
 &\quad - (A_{k+1} - A_k)^2 \\
 &\leq \mathbb{E}(Z_{k+1}^2 - Z_k^2 | \mathcal{F}_k) + 2Z_k(A_{k+1} - A_k).
 \end{aligned}$$

Now take expectation, sum up for $k = 0, \dots, n$, add $\mathbb{E}(M_0^2) = \mathbb{E}(Z_0^2)$, and use the fact that $Z_{n+1} = 0$ to give that

$$(3) \quad \mathbb{E}(M_{n+1}^2) \leq 2\mathbb{E}\left[\sum_{k=0}^n Z_k(A_{k+1} - A_k)\right].$$

Since A is increasing and $0 \leq Z \leq c$, we see that the right hand side of (3) is bounded above by $2c\mathbb{E}(A_{n+1})$. But $A_{n+1} = M_{n+1}$, and so $\mathbb{E}(A_{n+1}) = \mathbb{E}(M_{n+1}) = \mathbb{E}(M_0) = \mathbb{E}(Z_0)$. Thus we have that

$$(4) \quad \mathbb{E}(A_{n+1}^2) = \mathbb{E}(M_{n+1}^2) \leq 2c\mathbb{E}(Z_0).$$

From $H_{n+1} = 0$, it follows that $(H*Z)_n = (H*Z)_{n+1} = (H*M)_{n+1} + (H*A)_{n+1}$.

Thus

$$\begin{aligned}
 \|(H*Z)_n\|_2 &\leq \|(H*M)_{n+1}\|_2 + \|(H*A)_{n+1}\|_2 \\
 &\leq \|H\|_\infty \|M_{n+1}\|_2 + \|H\|_\infty \|A_{n+1}\|_2 \\
 &\quad \text{(first term by Proposition(7.12),} \\
 &\quad \text{the second since } |(H*A)_{n+1}| \leq \|H\|_\infty A_{n+1}) \\
 &\leq 2\|H\|_\infty [2c\mathbb{E}(Z_0)]^{1/2} \quad \text{(by (4)).}
 \end{aligned}$$

Squaring now gives the result. □

(7.14) Proposition.

- (i) If X is a non-negative supermartingale, and H is a bounded previsible process, then for each $n \in \mathbb{N}$, and each $\lambda > 0$,

$$\lambda \mathbb{P}[|(H * X)_n| > \lambda] \leq 6 \|H\|_\infty \mathbb{E}(X_0).$$

- (ii) If M is a martingale, and H is a bounded previsible process, then for each $n \in \mathbb{N}$, and each $\lambda > 0$,

$$\lambda \mathbb{P}[|(H * M)_n| > \lambda] \leq 12 \|H\|_\infty \|M_n\|_1.$$

- (iii) If X is any L^1 -process, and H is a bounded previsible process, then for each $n \in \mathbb{N}$, and each $\lambda > 0$,

$$\lambda \mathbb{P}[|(H * X)_n| > \lambda] \leq 12 \|H\|_\infty \mathbb{E} \left[\sum_{k=0}^{n-1} |\mathbb{E}(X_k - X_{k+1} | \mathcal{F}_k)| + |X_n| \right].$$

Note. The expectation in the inequality of (iii) is an expression similar to one that occurs in the definition of quasimartingales (see Definition(8.8)(d) and (e)).

Proof:

(i) Fix $n \in \mathbb{N}$ and $\lambda > 0$. Suppose that $0 < c < \infty$. Define a stopping time σ by $\sigma = \inf\{k \mid X_k > c\}$, and a process Z by $Z_k = \mathbf{1}_{[\sigma > k]} X_k$. Now $Z_k = X_{\sigma \wedge k} - \mathbf{1}_{[\sigma \leq k]} X_\sigma$. But $(X_{\sigma \wedge k})_k$ is a supermartingale by Proposition(7.10), and $(\mathbf{1}_{[\sigma \leq k]} X_\sigma)_k$ is an increasing process; thus Z is a supermartingale. By construction $0 \leq Z \leq c$, and so Proposition(7.13) applies to Z . We obtain therefore (using also that $Z_0 \leq X_0$) that

$$(1) \quad \mathbb{E}[(H * Z)_n^2] \leq 8c \|H\|_\infty^2 \mathbb{E}(Z_0) \leq 8c \|H\|_\infty^2 \mathbb{E}(X_0).$$

Now on the set $[\sigma > n]$ we have that $(H * X)_n = (H * Z)_n$. Therefore

$$(2) \quad [| (H * X)_n | > \lambda] \subseteq [\sigma \leq n] \cup [| (H * Z)_n | > \lambda].$$

Now $[\sigma \leq n] = [X_{\sigma \wedge n} > c]$, and so $\mathbf{1}_{[\sigma \leq n]} \leq \frac{1}{c} X_{\sigma \wedge n}$. Thus taking the expectation of both sides gives that $\mathbb{P}[\sigma \leq n] \leq \frac{1}{c} \mathbb{E}(X_{\sigma \wedge n})$. Now by Optional Stopping (Theorem(7.9)), $\mathbb{E}(X_{\sigma \wedge n}) \leq \mathbb{E}(X_0)$. Thus

$$(3) \quad \mathbb{P}[\sigma \leq n] \leq \frac{1}{c} \mathbb{E}(X_0).$$

Now $\mathbf{1}_{[|(H*Z)_n| > \lambda]} \leq \frac{1}{\lambda^2} (H*Z)_n^2$. Taking the expectation of both sides and using (1) gives that

$$(4) \quad \mathbb{P}[|(H*Z)_n| > \lambda] \leq \frac{8c}{\lambda^2} \|H\|_\infty^2 \mathbb{E}(X_0).$$

Formulae (2), (3) and (4) together give that

$$(5) \quad \mathbb{P}[|(H*X)_n| > \lambda] \leq \left(\frac{1}{c} + \frac{8c}{\lambda^2} \|H\|_\infty^2 \right) \mathbb{E}(X_0).$$

Now since Z and σ do not appear in (5), it is true for all c such that $0 < c < \infty$. If $\|H\|_\infty > 0$ then put $c = \lambda/(4\|H\|_\infty)$ to give the result. When $\|H\|_\infty = 0$, the result is trivial.

(ii) This follows immediately from part (iii).

(iii) Fix $n \in \mathbb{N}$, and define processes Y and Z by

$$Y_k = \mathbb{E} \left[\sum_{j=k}^{n-1} \mathbb{E}(X_j - X_{j+1} | \mathcal{F}_j)^+ + X_n^+ \mid \mathcal{F}_k \right],$$

and

$$Z_k = \mathbb{E} \left[\sum_{j=k}^{n-1} \mathbb{E}(X_j - X_{j+1} | \mathcal{F}_j)^- + X_n^- \mid \mathcal{F}_k \right],$$

when $k < n$, and $Y_k = X_n^+$, $Z_k = X_n^-$ when $k \geq n$. Now $Y_k \geq \mathbb{E}(Y_{k+1} | \mathcal{F}_k)$ and $Z_k \geq \mathbb{E}(Z_{k+1} | \mathcal{F}_k)$. So Y and Z are non-negative supermartingales. Thus (i) applies

to Y and Z . Now $X_k = Y_k - Z_k$ for $k \leq n$. So $(H*X)_n = (H*Y)_n - (H*Z)_n$. Thus for $\lambda > 0$ we have that

$$\begin{aligned} \lambda \mathbb{P}[|(H*X)_n| > \lambda] &\leq \lambda \mathbb{P}[|(H*Y)_n| > \frac{\lambda}{2}] + \lambda \mathbb{P}[|(H*Z)_n| > \frac{\lambda}{2}] \\ &\leq 12\|H\|_\infty \mathbb{E}(Y_0 + Z_0) \quad (\text{by (i)}). \end{aligned}$$

But

$$\mathbb{E}(Y_0 + Z_0) = \mathbb{E}\left[\sum_{k=0}^{n-1} |\mathbb{E}(X_k - X_{k+1} | \mathcal{F}_k)| + |X_n|\right],$$

and so the result now follows. $\square$

Now to obtain an inequality for the case $1 < p < \infty$, we shall use interpolation for the range $1 < p < 2$. Then to deal with the range $2 < p < \infty$ we shall use the following duality result.

(7.15) Proposition. *Suppose that $1 \leq p, q, r \leq \infty$ are such that $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} \leq 1$, M is an L^p -martingale, N is an L^q -martingale and that H is an L^r -previsible process. Then for each $n \in \mathbb{N}$,*

$$\mathbb{E}[(H*M)_n N_n] = \mathbb{E}[M_n (H*N)_n].$$

Proof: By Proposition(7.7)(i) and Proposition(7.8), $(H*M)N - [H*M, N]$ is a martingale. It is zero at time zero. Thus $\mathbb{E}[(H*M)_n N_n] = \mathbb{E}([H*M, N]_n)$. But by Proposition(7.6)(ii), $[H*M, N] = H*[M, N]$, and so we obtain that $\mathbb{E}[(H*M)_n N_n] = \mathbb{E}[(H*[M, N])_n]$. Now similarly $\mathbb{E}[M_n (H*N)_n] = \mathbb{E}[(H*[M, N])_n]$. Thus the result now follows. $\square$

(7.16) Proposition. *Suppose that $1 < p < \infty$. Then there exists a constant C_p , depending only on p , such that $0 \leq C_p < \infty$, and for each L^p -martingale M , and each bounded previsible process H ,*

$$\|(H*M)_n\|_p \leq C_p \|H\|_\infty \|M_n\|_p \quad \text{for all } n \in \mathbb{N}.$$

Proof: We have already proved this result for the case when $p = 2$ with $C_2 = 1$ (Proposition(7.12)). We shall consider now the case when $1 < p < 2$. We shall deduce this case by interpolation (the Marcinkiewicz interpolation theorem see Sadosky [Sa] Theorem 4.1 of Chapter 4). In order to give a self contained account we shall prove the result from first principles.

Fix n . Without loss of generality we can assume that $\|H\|_\infty = 1$. For each $\lambda > 0$ define martingales $X^{(\lambda)}$ and $Y^{(\lambda)}$ by

$$(1) \quad X_k^{(\lambda)} = E(\mathbf{1}_{[|M_n| > \lambda]} M_n | \mathcal{F}_k),$$

$$(2) \quad Y_k^{(\lambda)} = E(\mathbf{1}_{[|M_n| \leq \lambda]} M_n | \mathcal{F}_k).$$

So $X_k^{(\lambda)} + Y_k^{(\lambda)} = M_k$ when $k \leq n$, and $Y^{(\lambda)}$ is bounded. From Proposition(7.14)(ii) and Proposition(7.12) we obtain that

$$(3) \quad \frac{\lambda}{2} \mathbb{P}[|(H * X^{(\lambda)})_n| > \frac{\lambda}{2}] \leq 12 \|X_n^{(\lambda)}\|_1,$$

$$(4) \quad \frac{\lambda^2}{4} \mathbb{P}[|(H * Y^{(\lambda)})_n| > \frac{\lambda}{2}] \leq \|Y_n^{(\lambda)}\|_2^2.$$

(The inequality (4) being deduced from Proposition(7.12) using $\alpha^2 \mathbb{P}[|U| > \alpha] \leq E(U^2)$.) Now since $(H * M)_n = (H * X^{(\lambda)})_n + (H * Y^{(\lambda)})_n$, we have that $[|(H * M)_n| > \lambda] \subseteq [|(H * X^{(\lambda)})_n| > \frac{\lambda}{2}] \cup [|(H * Y^{(\lambda)})_n| > \frac{\lambda}{2}]$. Therefore from (3) and (4) we obtain that

$$(5) \quad \mathbb{P}[|(H * M)_n| > \lambda] \leq \frac{24}{\lambda} \|X_n^{(\lambda)}\|_1 + \frac{4}{\lambda^2} \|Y_n^{(\lambda)}\|_2^2.$$

Therefore

$$(6) \quad \begin{aligned} \|(H * M)_n\|_p^p &= p \int_0^\infty \lambda^{p-1} \mathbb{P}[|(H * M)_n| > \lambda] d\lambda \\ &\quad (\text{by Proposition(1.11)(ii)}) \\ &\leq 24p \int_0^\infty \lambda^{p-2} \|X_n^{(\lambda)}\|_1 d\lambda + 4p \int_0^\infty \lambda^{p-3} \|Y_n^{(\lambda)}\|_2^2 d\lambda. \end{aligned}$$

Now

$$\begin{aligned}
(7) \quad \int_0^\infty \lambda^{p-2} \|X_n^{(\lambda)}\|_1 d\lambda &= \int_0^\infty \int_\Omega \lambda^{p-2} \mathbf{1}_{[|M_n| > \lambda]}(\omega) |M_n(\omega)| \mathbb{P}(d\omega) d\lambda \\
&= \int_\Omega |M_n(\omega)| \int_0^{|M_n(\omega)|} \lambda^{p-2} d\lambda \mathbb{P}(d\omega) \\
&\quad \text{(by interchanging the order of integration)} \\
&= \frac{1}{p-1} \|M_n\|_p^p.
\end{aligned}$$

And similarly we have that

$$\begin{aligned}
(8) \quad \int_0^\infty \lambda^{p-3} \|Y_n^{(\lambda)}\|_2^2 d\lambda &= \int_0^\infty \int_\Omega \lambda^{p-3} \mathbf{1}_{[|M_n| \leq \lambda]}(\omega) |M_n(\omega)|^2 \mathbb{P}(d\omega) d\lambda \\
&= \int_\Omega \int_{|M_n(\omega)|}^\infty |M_n(\omega)|^2 \lambda^{p-3} d\lambda \mathbb{P}(d\omega) \\
&= \frac{1}{2-p} \|M_n\|_p^p.
\end{aligned}$$

Therefore (7) together with (8) and (9) gives that

$$(9) \quad \|(H * M)_n\|_p^p \leq \left(\frac{24p}{p-1} + \frac{4p}{2-p} \right) \|M_n\|_p^p,$$

Thus we have proved the result when $1 < p < 2$ with

$$(10) \quad C_p = \left(\frac{24p}{p-1} + \frac{4p}{2-p} \right)^{1/p} \quad (1 < p < 2).$$

Now we consider the case when $2 < p < \infty$. Let q be such that $\frac{1}{p} + \frac{1}{q} = 1$; so $1 < q < 2$. If N is an L^q -martingale, then

$$\begin{aligned}
|\mathbb{E}[(H * M)_n N_n]| &= |\mathbb{E}[M_n (H * N)_n]| && \text{(by Proposition(7.15))} \\
&\leq \|M_n\|_p \|(H * N)_n\|_q && \text{(by Hölder's inequality)} \\
&\leq C_q \|H\|_\infty \|M_n\|_p \|N_n\|_q && \text{(by the previous case).}
\end{aligned}$$

Now N_n can be any L^q $\mathcal{F}_n$ -measurable random variable desired (let $N_k = \mathbb{E}(U|\mathcal{F}_k)$, where U is the desired random variable). Thus $\|(H * M)_n\|_p \leq C_q \|H\|_\infty \|M_n\|_p$ as required ($C_p = C_q$). $\square$

CHAPTER 8: MARTINGALES AND THE STOCHASTIC INTEGRAL

In this chapter we shall consider the connections between our theory of stochastic integration and martingales. Among the results proven are that all local martingales are integrators, and that $H \cdot M$ is a local martingale, when M is a local martingale, and H is a locally bounded previsible process. Our approach to the theory will also be proved to be equivalent to the Strasbourg approach (see Chapter 1).

§8.1 UNIFORM INTEGRABILITY

In order to obtain convergence in L^p in certain situations, we shall need the concept of uniform integrability and uniform L^p . In this section we present some results which we shall need.

Throughout this section we shall assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is any probability space.

(8.1) Definition. If $\mathcal{X} \subset L^1$, we say that $\mathcal{X}$ is *uniformly integrable* if and only if

$$\sup_{X \in \mathcal{X}} \int_{|X| > N} |X| d\mathbb{P} \rightarrow 0 \quad \text{as } N \rightarrow \infty.$$

If $0 < p < \infty$ and $\mathcal{X} \subset L^p$, we say that $\mathcal{X}$ is *uniformly L^p* if and only if $\{|X|^p \mid X \in \mathcal{X}\}$ is uniformly integrable. Note that the terms ‘uniformly L^1 ’ and ‘uniformly integrable’ are equivalent. $\square$

We see at once that a subset of a uniformly integrable subset of L^1 is also uniformly integrable. Finite subsets of L^1 are automatically uniformly integrable. A finite union of uniformly integrable subsets of L^1 is also uniformly integrable. If $\mathcal{X}, \mathcal{Y} \subseteq L^1$, $\mathcal{X}$ is uniformly integrable, and for each $Y \in \mathcal{Y}$, there is an $X \in \mathcal{X}$ such that $|X| \geq |Y|$, then $\mathcal{Y}$ is also uniformly integrable. Similar facts are true for uniformly L^p subsets of L^p .

Here is an alternative characterization of uniform integrability.

(8.2)Proposition. Suppose that $\mathcal{X} \subseteq L^1$. Then $\mathcal{X}$ is uniformly integrable if and only if the following properties are satisfied.

- (1) $\sup_{X \in \mathcal{X}} \|X\|_1 < \infty$.
- (2) For each $\varepsilon > 0$, there exists a $\delta > 0$ such that $\int_A |X| d\mathbb{P} < \varepsilon$ for all $X \in \mathcal{X}$, and all $A \in \mathcal{F}$ such that $\mathbb{P}(A) < \delta$.

Proof: Suppose that $\mathcal{X}$ is uniformly integrable. For $X \in \mathcal{X}$ and $N \in \mathbb{N}$, write $|X| = Z_X^{(N)} + N \wedge |X|$. Then $Z_X \geq 0$ and $\int_{\Omega} Z_X^{(N)} d\mathbb{P} \leq \int_{[|X| > N]} |X| d\mathbb{P}$. So $\sup_{X \in \mathcal{X}} \int_{\Omega} Z_X^{(N)} d\mathbb{P} \rightarrow 0$ as $N \rightarrow \infty$. Now suppose that $\varepsilon > 0$ is given, and take N such that $\sup_{X \in \mathcal{X}} \int_{\Omega} Z_X^{(N)} d\mathbb{P} < \frac{\varepsilon}{2}$. Then for each $A \in \mathcal{F}$,

$$\begin{aligned} \sup_{X \in \mathcal{X}} \int_A |X| d\mathbb{P} &\leq \sup_{X \in \mathcal{X}} \int_{\Omega} Z_X^{(N)} d\mathbb{P} + \sup_{X \in \mathcal{X}} \int_A (N \wedge |X|) d\mathbb{P} \\ &< \frac{\varepsilon}{2} + N\mathbb{P}(A). \end{aligned}$$

Putting $A = \Omega$ gives (1). Taking $\delta = \varepsilon/(2N)$ gives (2).

Now suppose conversely that $\mathcal{X}$ satisfies (1) and (2) above. Then let $C = \sup_{X \in \mathcal{X}} \|X\|_1$; so $C < \infty$. We have for each $X \in \mathcal{X}$, that $\mathbb{P}[|X| > N] \leq \frac{1}{N} \|X\|_1 \leq \frac{C}{N}$. Now suppose that $\varepsilon > 0$ is given. Let $\delta > 0$ be that given by (2), and then take N_0 such that $\frac{C}{N_0} < \delta$. For $N \geq N_0$ and $X \in \mathcal{X}$, we have that $\mathbb{P}[|X| > N] \leq \frac{C}{N} < \delta$. Therefore by (2), $\int_{[|X| > N]} |X| d\mathbb{P} < \varepsilon$. Hence $\sup_{X \in \mathcal{X}} \int_{[|X| > N]} |X| d\mathbb{P} \leq \varepsilon$ for all $N \geq N_0$. I.e. $\sup_{X \in \mathcal{X}} \int_{[|X| > N]} |X| d\mathbb{P} \rightarrow 0$ as $N \rightarrow \infty$. So $\mathcal{X}$ is uniformly integrable as required. $\square$

This characterization of uniformly integrable makes it easier to prove certain results about uniform integrability. For instance the following proposition.

(8.3)Proposition. If $0 < p < \infty$, $\mathcal{X}, \mathcal{Y} \subseteq L^p$ are uniformly L^p , and $\alpha, \beta \in \mathbb{R}$, then the set $\alpha\mathcal{X} + \beta\mathcal{Y} = \{\alpha X + \beta Y \mid X \in \mathcal{X}, Y \in \mathcal{Y}\}$ is uniformly L^p .

PROOF: When $p = 1$, this follows easily using Proposition(8.2). For general p , use the case when $p = 1$, and that $|\alpha X + \beta Y|^p \leq 2^p |\alpha|^p |X|^p + 2^p |\beta|^p |Y|^p$, which follows

from the inequality $(s+t)^p \leq 2^p(s^p + t^p)$ that holds for all $s, t \in \mathbb{R}^+$. To prove this inequality, note that $s+t \leq 2(s \vee t)$. Therefore $(s+t)^p \leq 2^p(s^p \vee t^p) \leq 2^p(s^p + t^p)$. $\square$

(8.4)Theorem. Suppose that $0 < p < \infty$, $X_1, X_2, \dots \in L^p$ and that X is a real-valued random variable. Then the following are equivalent.

- (1) $(X_n)_n$ is uniformly L^p and $X_n \rightarrow X$ in probability.
- (2) $X \in L^p$ and $X_n \rightarrow X$ in L^p .

Proof: Suppose that (1) holds. We shall deduce (2). A subsequence $(X_{n_k})_k$ of $(X_n)_n$ converges a.s. to X . Thus by Fatou's Lemma,

$$\begin{aligned} \int_{\Omega} |X|^p d\mathbb{P} &= \int_{\Omega} \liminf_{k \rightarrow \infty} |X_{n_k}|^p d\mathbb{P} \\ &\leq \liminf_{k \rightarrow \infty} \int_{\Omega} |X_{n_k}|^p d\mathbb{P} \\ &\leq \sup_n \|X_n\|_p^p < \infty, \end{aligned}$$

by (1) of Proposition(8.2). Let $Y_n = |X_n - X|^p$. So $(Y_n)_n$ is uniformly integrable (since $(X_n - X)_n$ is uniformly L^p by Proposition(8.3)). Now suppose that $\varepsilon > 0$ is given. Take N such that $\sup_n \int_{[Y_n > N]} Y_n d\mathbb{P} < \frac{\varepsilon}{2}$. Since $Y_n \rightarrow 0$ in probability, take n_0 such that $\int_{\Omega} (N \wedge Y_n) d\mathbb{P} < \frac{\varepsilon}{2}$ for all $n \geq n_0$. Then for each $n \geq n_0$,

$$\begin{aligned} \int_{\Omega} Y_n d\mathbb{P} &= \int_{[Y_n > N]} Y_n d\mathbb{P} + \int_{[Y_n \leq N]} Y_n d\mathbb{P} \\ &< \frac{\varepsilon}{2} + \int_{\Omega} (N \wedge Y_n) d\mathbb{P} \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

Thus $\int_{\Omega} Y_n d\mathbb{P} \rightarrow 0$ as required.

Now suppose conversely that (2) holds. We shall deduce (1). Clearly $X_n \rightarrow X$ in probability. Let $Y_n = |X_n - X|^p$. We shall show that $(Y_n)_n$ is uniformly integrable. But since $\int_{[Y_n > N]} Y_n d\mathbb{P} \leq \|Y_n\|_1$, we have that for each N and n_0 ,

$$\sup_n \int_{[Y_n > N]} Y_n d\mathbb{P} \leq \sup_{1 \leq n \leq n_0} \int_{[Y_n > N]} Y_n d\mathbb{P} \wedge \sup_{n > n_0} \|Y_n\|_1.$$

Now suppose that $\varepsilon > 0$ is given. Take n_0 such that the second term on the right hand side is $< \varepsilon$. Then take N_0 such that the first term on the right hand side is $< \varepsilon$ for all $N \geq N_0$ (since $\{Y_1, \dots, Y_{n_0}\}$ is uniformly integrable). Thus $\sup_n \int_{[Y_n > N]} Y_n d\mathbb{P} < \varepsilon$ for all $N \geq N_0$. We see therefore that $(Y_n)_n$ is uniformly integrable. So $(X_n - X)_n$ is uniformly L^p , and hence $(X_n)_n$ is also uniformly L^p by Proposition(8.3). $\square$

(8.5)Proposition. Suppose that $0 < p < q \leq \infty$ and that $\mathcal{X}$ is a bounded subset of L^q . Then as a subset of L^p , $\mathcal{X}$ is uniformly L^p .

Proof: The case when $q = \infty$ is easy; so suppose that $q < \infty$. By replacing $\mathcal{X}$ by $\{|X|^p | X \in \mathcal{X}\}$, and q by q/p we may also assume without loss of generality that $p = 1$. So $1 = p < q < \infty$. Let r be such that $1 < r < \infty$ and $\frac{1}{q} + \frac{1}{r} = 1$. Also let $C = \sup_{X \in \mathcal{X}} \|X\|_q$. So $C < \infty$. Then for $X \in \mathcal{X}$,

$$\begin{aligned} \int_{[|X| > N]} |X| d\mathbb{P} &= \int_{\Omega} \mathbf{1}_{[|X| > N]} |X| d\mathbb{P} \\ &\leq \|\mathbf{1}_{[|X| > N]}\|_r \|X\|_q \quad (\text{by Hölder's inequality}) \\ &\leq C(\mathbb{P}[|X| > N])^{1/r}. \end{aligned}$$

But $\mathbb{P}[|X| > N] \leq (\frac{1}{N} \|X\|_q)^q \leq (\frac{C}{N})^q$, and so $\sup_{X \in \mathcal{X}} \int_{[|X| > N]} |X| d\mathbb{P} \leq C(\frac{C}{N})^{q/r} \rightarrow 0$ as $N \rightarrow \infty$. $\square$

§8.2 UNIFORM INTEGRABILITY AND CONDITIONAL EXPECTATION

We shall present here a couple of results which will be used in the theory of martingales. These results show that certain families defined using conditional expectation are uniformly integrable.

(8.6)Proposition. Suppose that $1 \leq p < \infty$ and that $\mathcal{X} \subseteq L^p$ is uniformly L^p . Then $(\mathbb{E}(X|\mathcal{G}))_{X \in \mathcal{X}, \mathcal{G} \subseteq \mathcal{F}}$ is uniformly L^p . ($\mathcal{G}$ is assumed to denote a σ -field on Ω .)

Note. Taking $\mathcal{X}$ to consist of one element shows that $(E(X|\mathcal{G}))_{\mathcal{G} \subseteq \mathcal{F}}$ is uniformly L^p for all $X \in L^p$.

Proof: Since by Jensen's inequality (see (9) of (7.2)), $|E(X|\mathcal{G})|^p \leq E(|X|^p|\mathcal{G})$, we may, by replacing $\mathcal{X}$ by $\{|X|^p | X \in \mathcal{X}\}$, assume that $p = 1$. Suppose that $\varepsilon > 0$ is given, and take (using Proposition(8.2)) $\delta > 0$ such that $\int_A |X| d\mathbb{P} < \varepsilon$ for all $X \in \mathcal{X}$, and all $A \in \mathcal{F}$ such that $\mathbb{P}(A) < \delta$. Let $C = \sup_{X \in \mathcal{X}} \|X\|_1$; so $C < \infty$. Take N_0 such that $\frac{C}{N_0} < \delta$. Now suppose that $N \geq N_0$, $X \in \mathcal{X}$ and that $\mathcal{G} \subseteq \mathcal{F}$. Let $Y = E(X|\mathcal{G})$. Then $\{|Y| > N\} \in \mathcal{G}$, and $\mathbb{P}(|Y| > N) \leq \frac{1}{N} \|Y\|_1 \leq \frac{1}{N} \|X\|_1 < \delta$. Thus

$$\begin{aligned} \int_{\{|Y| > N\}} |Y| d\mathbb{P} &\leq \int_{\{|Y| > N\}} E(|X| |\mathcal{G}) d\mathbb{P} \\ &= \int_{\{|Y| > N\}} |X| d\mathbb{P} \\ &< \varepsilon. \end{aligned}$$

So $\sup_{X \in \mathcal{X}, \mathcal{G} \subseteq \mathcal{F}} \int_{\{|E(X|\mathcal{G})| > N\}} |E(X|\mathcal{G})| d\mathbb{P} \leq \varepsilon$ for all $N \geq N_0$. Thus proving that $(E(X|\mathcal{G}))_{X \in \mathcal{X}, \mathcal{G} \subseteq \mathcal{F}}$ is uniformly integrable. $\square$

We shall also require the following technical lemma.

(8.7) Lemma. Suppose that $\mathcal{G}_0, \mathcal{G}_1, \dots$ are σ -fields on Ω such that $\mathcal{F} \supseteq \mathcal{G}_0 \supseteq \mathcal{G}_1 \supseteq \dots$, and that $X_0, X_1, \dots \in L^1$ satisfy the following properties.

- (1) For each n , X_n is $\mathcal{G}_n$ -measurable.
- (2) $\sum_{n=1}^{\infty} E[|E(X_n - X_{n-1}|\mathcal{G}_n)|] < \infty$.

Then $(X_n)_{n=0}^{\infty}$ is uniformly integrable.

Note. (2) is satisfied when $E(X_n|\mathcal{G}_m) \leq X_m$ for $n \leq m$, and $\lim_{n \rightarrow \infty} E(X_n) < \infty$.

Proof: Define Y_n for $n = 0, 1, \dots$ by

$$Y_n = E \left[|X_0| + \sum_{k=1}^n |E(X_k - X_{k-1}|\mathcal{G}_k)| \mid \mathcal{G}_n \right].$$

Then since

$$X_n = E \left[X_0 + \sum_{k=1}^n E(X_k - X_{k-1} | \mathcal{G}_k) \mid \mathcal{G}_n \right],$$

we see that $|X_n| \leq Y_n$. Therefore it is enough to prove that $(Y_n)_{n=0}^\infty$ is uniformly integrable.

Now we have that $Y_n \geq 0$, $E(Y_n | \mathcal{G}_{n+1}) \leq Y_{n+1}$, and that $\lim_{n \rightarrow \infty} E(Y_n) < \infty$. Let $\alpha = \lim_{n \rightarrow \infty} E(Y_n)$. So $0 \leq \alpha < \infty$. Given $\varepsilon > 0$, take n_0 such that $\alpha - \frac{\varepsilon}{2} < E(Y_{n_0}) \leq \alpha$. By applying Proposition(8.2) to the uniformly integrable set $\{Y_{n_0}\}$, take $\delta > 0$ such that $\int_A |Y_{n_0}| d\mathbb{P} < \frac{\varepsilon}{2}$ whenever $A \in \mathcal{F}$ with $\mathbb{P}(A) < \delta$. Let N_0 be such that $\frac{\alpha}{N_0} < \delta$, and $\int_{[Y_n > N]} Y_n d\mathbb{P} < \varepsilon$ whenever $N \geq N_0$ and $n < n_0$. Then for $N \geq N_0$ and $n \geq n_0$,

$$\begin{aligned} \int_{[Y_n > N]} Y_n d\mathbb{P} &= E(Y_n) - \int_{[Y_n \leq N]} Y_n d\mathbb{P} \\ &\leq E(Y_n) - \int_{[Y_n \leq N]} Y_{n_0} d\mathbb{P} \\ &\quad (\text{since } E(Y_{n_0} | \mathcal{G}_n) \leq Y_n) \\ &= E(Y_n) - E(Y_{n_0}) + \int_{[Y_n > N]} Y_{n_0} d\mathbb{P} \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon, \end{aligned}$$

since $\mathbb{P}[Y_n > N] \leq \frac{1}{N} E(Y_n) \leq \frac{\alpha}{N_0} < \delta$. Thus $\sup_n \int_{[Y_n > N]} Y_n d\mathbb{P} \leq \varepsilon$ for all $N \geq N_0$ as required. $\square$

§8.3 CONTINUOUS-TIME MARTINGALES

We shall suppose throughout the rest of this chapter that $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t \in \mathbb{R}^+})$ is a standard setup. We shall now define martingales, submartingales and supermartingales. In this case of continuous time we shall require them to be càdlàg processes. The prefix ‘pre’ will be used when this requirement is dropped.

(8.8) Definition.

- (a) A *premartingale* (resp. *presubmartingale*, *presupermartingale*) is a vector-valued adapted function $X: \mathbb{R}^+ \rightarrow L^1$ such that $X_s = E(X_t | \mathcal{F}_s)$ (resp. $X_s \leq E(X_t | \mathcal{F}_s)$, $X_s \geq E(X_t | \mathcal{F}_s)$) whenever $s, t \in \mathbb{R}^+$ are such that $s \leq t$.
- (b) A *martingale* (resp. *submartingale*, *supermartingale*) is a càdlàg process X such that $(X_t)_t$ is a premartingale (resp. presubmartingale, presupermartingale).
- (c) Suppose that $1 \leq p \leq \infty$. A martingale (resp. submartingale, supermartingale) X is said to be an L^p -martingale (resp. L^p -submartingale, L^p -supermartingale) if and only if $X_t \in L^p$ for all $t \in \mathbb{R}^+$ (i.e. X is an L^p -process). (Similarly for premartingales etc.)
- (d) If $X: \mathbb{R}^+ \rightarrow L^1$ is adapted, then we define

$$\text{Var}(X) = \sup_{0=t_0 < \dots < t_r < \infty} E \left[\sum_{k=0}^{r-1} |E(X_{t_k} - X_{t_{k+1}} | \mathcal{F}_{t_k})| + |X_{t_r}| \right],$$

and for $b \in \mathbb{R}^+$,

$$\text{Var}_b(X) = \sup_{0=t_0 < \dots < t_r = b} E \left[\sum_{k=0}^{r-1} |E(X_{t_k} - X_{t_{k+1}} | \mathcal{F}_{t_k})| + |X_{t_r}| \right],$$

(the supremum being taken over all partitions $0 = t_0 < \dots < t_r = b$ of $[0, b]$).

- (e) A *quasimartingale* is a càdlàg process X such that $\text{Var}(X) < \infty$. □

Note.

1. It is easy to see that if X is a premartingale, presubmartingale or a presupermartingale, then $\text{Var}_b(X) < \infty$ for all $b \in \mathbb{R}^+$.
2. It is also easy to prove that a non-negative supermartingale X is a quasimartingale and that $\text{Var}(X) = E(X_0)$.
3. The class of all quasimartingales forms a vector space and Var is a norm on it.

4. If $X: \mathbb{R}^+ \rightarrow L^1$ is adapted, then $\text{Var}_s(X) \leq \text{Var}_t(X) \leq \text{Var}(X)$ for all $s \leq t$ in $\mathbb{R}^+$.

(8.9)Notation. We shall use $\|\cdot\|_\infty$ to denote the supremum norm on $\mathcal{E}$. □

The following theorem can now be proved using the results of Chapter 7. This method of proof is due to D. A. Edwards (see [Ed]), though he did not obtain the results (iv) and (v) below.

(8.10)Theorem.

(i) If Z is a presupermartingale such that $0 \leq Z \leq c$ ($c \in \mathbb{R}^+$), and $b \in \mathbb{R}^+$, then

$$\mathbb{E}[(H \cdot Z)_b^2] \leq 8c \|H\|_\infty \mathbb{E}(Z_0) \quad \text{for all } H \in \mathcal{E}.$$

(ii) If X is a non-negative presupermartingale and $b \in \mathbb{R}^+$, then

$$\lambda \mathbb{P}[|(H \cdot X)_b| > \lambda] \leq 6 \|H\|_\infty \mathbb{E}(X_0) \quad \text{for all } H \in \mathcal{E} \text{ and } \lambda > 0.$$

(iii) If M is a premartingale and $b \in \mathbb{R}^+$, then

$$\lambda \mathbb{P}[|(H \cdot M)_b| > \lambda] \leq 12 \|H\|_\infty \|M_b\|_1 \quad \text{for all } H \in \mathcal{E} \text{ and } \lambda > 0.$$

(iv) Suppose that $X: \mathbb{R}^+ \rightarrow L^1$ is an adapted vector-valued function and that $b \in \mathbb{R}^+$ is such that $\text{Var}_b(X) < \infty$. Then

$$\lambda \mathbb{P}[|(H \cdot X)_b| > \lambda] \leq 12 \|H\|_\infty \text{Var}_b(X) \quad \text{for all } H \in \mathcal{E} \text{ and } \lambda > 0.$$

(v) For each p such that $1 < p < \infty$, there is a constant C_p depending only on p such that $0 \leq C_p < \infty$, and such that for each L^p -premartingale M and $b \in \mathbb{R}^+$,

$$\|(H \cdot M)_b\|_p \leq C_p \|H\|_\infty \|M_b\|_p \quad \text{for all } H \in \mathcal{E}.$$

Proof: We shall indicate only how to deduce (i) from Proposition(7.13), since the other parts are deduced in the same way from Proposition(7.14) and Proposition(7.16). So suppose that Z , c and b are as in the statement of (i), and that $H \in \mathcal{E}$. Without loss of generality we can assume that $[H \neq 0] \subseteq (0, b] \times \Omega$. So $H_t(\omega) = \sum_{k=0}^{r-1} C_k(\omega) \mathbf{1}_{(t_k, t_{k+1}]}(t)$, where $0 = t_0 < \dots < t_r = b$, and for $k = 0, \dots, r-1$, C_k is a simple $\mathcal{F}_{t_k}$ -measurable random variable. Now $(H \cdot Z)_b = \sum_{k=0}^{r-1} C_k(Z_{t_{k+1}} - Z_{t_k})$. By putting $t_k = b$ and $C_k = 0$ for $k \geq r$, and $C_{-1} = 0$, and then applying Proposition(7.13) with $(\mathcal{F}_{t_k})_k$ as the discrete-time filtration, $(Z_{t_k})_k$ as Z , and $(C_{k-1})_k$ as H , we obtain that $\mathbb{E}[(H \cdot Z)_b^2] \leq 8c\|H\|_\infty \mathbb{E}(Z_0)$ as required. $\square$

(8.11)Theorem. *Suppose that X is an L^1 -càdlàg process such that $\text{Var}_b(X) < \infty$ for all $b \in \mathbb{R}^+$. Then X is an integrator.*

Proof: By Theorem(8.10)(iv), $\int_0^b \dots dX: \mathcal{E} \rightarrow L^0$ is continuous. Thus by Theorem(4.30), it follows that X must be an integrator. $\square$

The hypothesis of Theorem(8.11) is satisfied by all martingales, submartingales, supermartingales and quasimartingales. Thus we have immediately the following result.

(8.12)Theorem. *All martingales, submartingales, supermartingales and quasimartingales are integrators.* $\square$

Note. All that is stated here is that the processes are L^0 -integrators. It is **not** true in general that all martingales are L^1 -integrators (see Bichteler [Bi2] 3.7 and 7.2).

The assumption in Theorem(8.11) that X is càdlàg is a strong one. The following result enables one to prove that certain adapted $X: \mathbb{R}^+ \rightarrow L^1$ are integrators, and hence deduce the existence of càdlàg versions of X .

(8.13)Proposition. Suppose that $X: \mathbb{R}^+ \rightarrow L^1$ is an adapted vector-valued function such that $\text{Var}_b(X) < \infty$ for all $b \in \mathbb{R}^+$. Then for each $t \in \mathbb{R}^+$, the limit $\lim_{u \downarrow t} X_u$ exists in L^1 . If this limit is X_t for all $t \in \mathbb{R}^+$, then X is an integrator, and so by Theorem(5.10) there exists a càdlàg process Y such that $Y_t = X_t$ a.s. for all $t \in \mathbb{R}^+$.

Proof: Fix $t \in \mathbb{R}^+$. Then we see by Theorem(4.31) (since $\int_0^b \dots dX: \mathcal{E} \rightarrow L^0$ is continuous) that $\lim_{u \downarrow t} X_u$ exists in L^0 . Let Z_t denote this limit. We have to show that $Z_t \in L^1$ and that $X_u \rightarrow Z_t$ in L^1 -norm. Take $u_n > t$ such that $u_n \downarrow t$. Then by Lemma(8.7) applied with $\mathcal{G}_n = \mathcal{F}_{u_n}$, we obtain that $(X_{u_n})_n$ is uniformly integrable. But $X_{u_n} \rightarrow Z_t$ in probability. So by Theorem(8.4), $Z_t \in L^1$ and $X_{u_n} \rightarrow Z_t$ in L^1 -norm. But L^1 is a metric space; so it follows that $X_u \rightarrow Z_t$ in L^1 -norm. Now if $Z_t = X_t$ a.s. for all $t \in \mathbb{R}^+$, then it follows that X is an integrator by Theorem(4.30). $\square$

We can now use Proposition(8.13) to prove the existence of càdlàg versions of premartingales, presubmartingales and presupermartingales.

(8.14)Theorem.

- (i) If M is a premartingale, then M is an integrator, and hence there exists a martingale N such that $N_t = M_t$ a.s. for all $t \in \mathbb{R}^+$.
- (ii) If X is a presubmartingale such that the increasing function $t \mapsto \mathbb{E}(X_t)$ is right continuous, then X is an integrator, and hence there exists a submartingale Y such that $Y_t = X_t$ a.s. for all $t \in \mathbb{R}^+$.

Proof:

- (i) M satisfies the hypothesis of Proposition(8.13), and so for each $t \in \mathbb{R}^+$, the limit $\lim_{s \downarrow t} M_s$ exists in L^1 . Denote this limit by Z_t . Now since M is a premartingale, $M_t = \mathbb{E}(M_s | \mathcal{F}_t)$ for all $s > t$. Let $s \downarrow t$ to give that $M_t = \mathbb{E}(Z_t | \mathcal{F}_t) = Z_t$ (since

the convergence is in L^1). Proposition(8.13) now shows that M is an integrator, and hence has a càdlàg version.

(ii) Again X satisfies the hypothesis of Proposition(8.13). For $t \in \mathbb{R}^+$, let Z_t denote the limit in L^1 , $\lim_{s \downarrow t} X_s$. We have that $X_t \leq E(X_s | \mathcal{F}_t)$ for all $s > t$. Let $s \downarrow t$ to deduce that $X_t \leq Z_t$. But $E(Z_t) = \lim_{s \downarrow t} E(X_s) = E(X_t)$, and so $X_t = Z_t$ a.s. Thus Proposition(8.13) again proves that X is an integrator, and hence has a càdlàg version. $\square$

We have commented on the fact that L^1 -martingales are not necessarily L^1 -integrators. But when $1 < p < \infty$, we do have that all L^p -martingales are L^p -integrators, as the following result states.

(8.15)Theorem. *Suppose that $1 < p < \infty$. Then all L^p -martingales are L^p -integrators. Furthermore if M is an L^p -martingale and $H \in \Lambda^1(M, L^p)$, then $H \cdot M$ is also an L^p -martingale.*

Proof: Suppose that M is an L^p -martingale. Then by Theorem(8.10)(v), we see that the mapping $\int_0^b \dots dM: \mathcal{E} \rightarrow L^p$ is continuous. Thus by Theorem(4.30), it follows that M is an L^p -integrator as required. Now suppose that $H \in \Lambda^1(M, L^p)$. We know that $H \cdot M$ is an L^p -process. Thus it remains to see that it is a martingale. But since M is an L^1 -integrator and $H \in \Lambda^1(M, L^1)$, this will follow from (i) of the following lemma. $\square$

(8.16)Lemma.

- (i) *If M is a martingale and an L^1 -integrator, and $H \in \Lambda^1(M, L^1)$, then $H \cdot M$ is a martingale and an L^1 -integrator.*
- (ii) *If X is a submartingale and an L^1 -integrator, and $H \in \Lambda^1(M, L^1)$ with $H \geq 0$, then $H \cdot X$ is a submartingale and an L^1 -integrator.*

Proof:

(i) We already know that $H \cdot M$ is an L^1 -integrator by Theorem(4.21)(i). We have to show that $H \cdot M$ is a martingale. Firstly consider the case when $H \in \mathcal{E}$. To show that $(H \cdot M)_u = \mathbb{E}[(H \cdot M)_v | \mathcal{F}_u]$ when $u \leq v$ in $\mathbb{R}^+$, write $H_t(\omega) = \sum_{k=0}^{r-1} C_k(\omega) \mathbf{1}_{(t_k, t_{k+1}]}(t)$, where $0 = t_0 < \dots < t_r < \infty$, for $k = 0, \dots, r-1$, C_k is a simple $\mathcal{F}_{t_k}$ -measurable random variable, and u and v occur in the set $\{t_0, \dots, t_r\}$. Now apply Proposition(7.7)(i) with $(M_{t_k})_k$ as M ($t_k = t_r$ if $k \geq r$), $(C_{k-1})_k$ as H ($C_k = 0$ if k is not one of $0, \dots, r-1$), and $(\mathcal{F}_{t_k})_k$ as the filtration.

Now consider general H . Take $H_1, H_2, \dots \in \mathcal{E}$ such that $\Phi_M^1(\mathbf{1}_{(0,t] \times \Omega}(H_n - H)) \rightarrow 0$ for all $t \in \mathbb{R}^+$. Then $(H_n \cdot M)_t \rightarrow (H \cdot M)_t$ in L^1 -norm for all t . Thus if $u \leq v$ in $\mathbb{R}^+$, then $(H_n \cdot M)_u = \mathbb{E}[(H_n \cdot M)_v | \mathcal{F}_u]$ for all n . Let $n \rightarrow \infty$ to deduce that $(H \cdot M)_u = \mathbb{E}[(H \cdot M)_v | \mathcal{F}_u]$.

(ii) The proof here is the same as for (i), except when approximating H by $H_1, H_2, \dots \in \mathcal{E}$, take $H_n \geq 0$ (by replacing H_n by H_n^+ if necessary). $\square$

§8.4 SOME GENERAL RESULTS

We shall present here some results some of which are the continuous-time analogues of those appearing in §7.4. The first is an analogue of Theorem(7.9). All these results are well known, but are included for completeness.

(8.17)Theorem (Optional Stopping).

Suppose that X is a martingale (resp. submartingale, supermartingale), and that σ and τ are bounded stopping times such that $\sigma \leq \tau$. Then $X_\sigma, X_\tau \in L^1$ and $X_\sigma = \mathbb{E}(X_\tau | \mathcal{F}_\sigma)$ (resp. $X_\sigma \leq \mathbb{E}(X_\tau | \mathcal{F}_\sigma)$, $X_\sigma \geq \mathbb{E}(X_\tau | \mathcal{F}_\sigma)$).

Proof: If σ and τ are simple stopping times, then this result follows directly from Theorem(7.9). For general σ and τ take finite simple stopping times $\sigma_1, \sigma_2, \dots$ and $\tau_1, \tau_2, \dots$ such that $\sigma_n \downarrow \sigma$, $\tau_n \downarrow \tau$ and $\sigma_n \leq \tau_n$ (see the proof of Proposition(5.6)(f)).

Now I claim that $X_\sigma \in L^1$ and $X_{\sigma_n} \rightarrow X_\sigma$ in L^1 -norm. It is enough, by Proposition(8.4), to prove that $(X_{\sigma_n})_n$ is uniformly integrable. In the martingale case this follows from Proposition(8.6), since $X_{\sigma_n} = E(X_{\sigma_1} | \mathcal{F}_{\sigma_n})$ by the simple case. In the submartingale case we have that $E(X_{\sigma_n}) \geq E(X_0)$ so in this case use Lemma(8.7) with $\mathcal{G}_n = \mathcal{F}_{\sigma_n}$ (see the note afterwards). Similarly $X_\tau \in L^1$ and $X_{\tau_n} \rightarrow X_\tau$ in L^1 -norm.

Suppose that X is a martingale. If $A \in \mathcal{F}_\sigma$ then $A \in \mathcal{F}_{\sigma_n}$ for all n . Thus from the simple case, $\int_A X_{\sigma_n} d\mathbb{P} = \int_A X_{\tau_n} d\mathbb{P}$ for all n . Let $n \rightarrow \infty$ to give that $\int_A X_\sigma d\mathbb{P} = \int_A X_\tau d\mathbb{P}$. Thus $X_\sigma = E(X_\tau | \mathcal{F}_\sigma)$ as required. The submartingale and supermartingale cases follow in the same way. $\square$

We have the following analogue of Theorem(7.10). (Recall from §5.3 that $X_t^\sigma = X_{\sigma \wedge t} - X_0$.)

(8.18)Theorem. *If X is a martingale (resp. submartingale, supermartingale), and σ is a stopping time, then X^σ is a martingale (resp. submartingale, supermartingale).*

Proof: We could deduce it from Proposition(7.10) using an argument similar to the deduction of Theorem(8.17) from Theorem(7.9). Alternatively we can deduce it from Theorem(8.17) as follows. We know that $X_{\sigma \wedge t} \in L^1$ by Theorem(8.17). We consider the case when X is a martingale (the other cases being similar). Suppose that α and β are bounded stopping times such that $\alpha \leq \beta$. Then $E(X_\beta^\sigma - X_\alpha^\sigma) = E(X_{\sigma \wedge \beta} - X_{\sigma \wedge \alpha}) = 0$, since $X_{\sigma \wedge \alpha} = E(X_{\sigma \wedge \beta} | \mathcal{F}_{\sigma \wedge \alpha})$ by Theorem(8.17). The result now follows from the following lemma. $\square$

(8.19)Lemma. *Suppose that $X: \mathbb{R}^+ \rightarrow L^1$ is a vector-valued adapted function such that, $E(X_\tau - X_\sigma) = 0$ (resp. $\geq 0, \leq 0$), whenever σ and τ are finite simple stopping times such that $\sigma \leq \tau$. Then X is a premartingale (resp. presubmartingale, presupermartingale).*

Note. In order to prove that X is a premartingale, it is enough to show that $E(X_\sigma - X_0) = 0$ for all finite simple stopping times σ . (Since subtraction gives the hypothesis of the lemma.)

Proof: We consider only the case ' $= 0$ ' (the others are similar). Suppose that $s \leq t$ are in $\mathbb{R}^+$ and that $F \in \mathcal{F}_s$. Let σ be the finite simple stopping that is equal to s on F , and equal to t outside F . Let τ be the constant stopping taking the value t . Then by hypothesis, $0 = E(X_\tau - X_\sigma) = \int_F (X_t - X_s) d\mathbb{P}$. Thus we see that X is a premartingale as required. $\square$

Recall that if Z is a non-negative supermartingale, then Z is a quasimartingale and $\text{Var}(Z) = E(Z_0)$ (see Note 2 after Definition(8.8)). Thus we see that if X is a càdlàg process and $X = Y - Z$, where Y and Z are non-negative supermartingales, then X is a quasimartingale and $\text{Var}(X) \leq E(Y_0) + E(Z_0)$. We shall need the following converse of this result in order to prove the theorem of Bichteler, Dellacherie, Mokobodzki and Meyer concerning the characterization of integrators. The result is due to Rao and Stricker. The proof given is essentially that given in Rogers and Williams [RW].

(8.20) Theorem(Rao, Stricker).

Suppose that X is a quasimartingale. Then there exist unique non-negative supermartingales Y and Z such that

$$X = Y - Z \quad \text{and} \quad \text{Var}(X) = E(Y_0) + E(Z_0).$$

We have further that if $X = U - V$, where U and V are non-negative supermartingales, then $U - Y = V - Z$ is a non-negative supermartingale.

In order to prove this we shall use the following lemma which is a generalization of the monotone convergence theorem.

(8.21) Lemma. Suppose that $(\Omega, \mathcal{F}, \mu)$ is any measure space, and that $(f_\alpha)_{\alpha \in I}$ is a net in $L^1(\mu)$ such that the following properties hold.

(1) $(f_\alpha)_\alpha$ is increasing; i.e. $f_\alpha \leq f_\beta$ when $\alpha \leq \beta$.

(2) $\sup_\alpha \int_\Omega f_\alpha d\mu < \infty$.

Then the net $(f_\alpha)_\alpha$ is convergent in $L^1(\mu)$.

Proof: Let $C = \sup_\alpha \int_\Omega f_\alpha d\mu$. Now take $\alpha_1 \leq \alpha_2 \leq \dots$ such that $\int_\Omega f_{\alpha_n} d\mu \rightarrow C$.

By the monotone convergence theorem there exists $f \in L^1(\mu)$ such that $f_{\alpha_n} \uparrow f$ a.e.

We have that $\int_\Omega f_{\alpha_n} d\mu \rightarrow \int_\Omega f d\mu$, and so $f_{\alpha_n} \rightarrow f$ in $L^1(\mu)$.

Now I claim that $f_\beta \leq f$ a.e. for all $\beta \in I$. For suppose that $\beta \in I$. Then take $\gamma_1 \in I$ such that $\gamma_1 \geq \beta$ and $\gamma_1 \geq \alpha_1$. Given $\gamma_{n-1} \in I$, take $\gamma_n \in I$ such that $\gamma_n \geq \gamma_{n-1}$ and $\gamma_n \geq \alpha_n$. Let $g = \lim_{n \rightarrow \infty} f_{\gamma_n}$. Then $f_\beta \leq g$ a.e. and $f \leq g$ a.e. Now $\int_\Omega g d\mu \leq C = \int_\Omega f d\mu$; so $g = f$ a.e. Hence $f_\beta \leq f$ a.e. as required.

Suppose that $\varepsilon > 0$ is given. Then take n such that $\|f_{\alpha_n} - f\|_1 < \varepsilon$. Thus for $\beta \geq \alpha_n$ we have that $f_{\alpha_n} \leq f_\beta \leq f$, and so $\|f_\beta - f\|_1 < \varepsilon$. Hence $f_\beta \rightarrow f$ in $L^1(\mu)$ as $\beta \rightarrow \infty$. $\square$

Proof of (8.20): For $t \in \mathbb{R}^+$, let F_t denote the set of finite subsets ξ of $[t, \infty)$ such that $t \in \xi$. For $\xi \in F_t$ define

$$Q_t^{(\pm)}(\xi) = \mathbb{E} \left[\sum_{k=0}^{r-1} \mathbb{E}(X_{t_k} - X_{t_{k+1}} | \mathcal{F}_{t_k})^\pm + X_{t_r}^\pm \mid \mathcal{F}_t \right],$$

where $\xi = \{t_0, \dots, t_r\}$ with $t = t_0 < \dots < t_r$. We shall present the proof as a series of numbered steps.

(1) $Q_t^{(\pm)}(\xi) \leq Q_t^{(\pm)}(\eta)$, when $\xi, \eta \in F_t$ with $\xi \subseteq \eta$.

Proof: It is enough to consider the case when $\eta = \xi \cup \{u\}$ with $u \notin \xi$. Suppose that $\xi = \{t_0, \dots, t_r\}$ with $t = t_0 < \dots < t_r$. If $t_k < u < t_{k+1}$ then

$$\begin{aligned} \mathbb{E}(X_{t_k} - X_{t_{k+1}} | \mathcal{F}_{t_k})^\pm &= (\mathbb{E}(X_{t_k} - X_u | \mathcal{F}_{t_k}) + \mathbb{E}[\mathbb{E}(X_u - X_{t_{k+1}} | \mathcal{F}_u) | \mathcal{F}_{t_k}])^\pm \\ &\leq \mathbb{E}(X_{t_k} - X_u | \mathcal{F}_{t_k})^\pm + \mathbb{E}[\mathbb{E}(X_u - X_{t_{k+1}} | \mathcal{F}_u)^\pm | \mathcal{F}_{t_k}]. \end{aligned}$$

So applying $\mathbb{E}[\cdot|\mathcal{F}_t]$ and adding the other terms gives that $Q_t^{(\pm)}(\xi) \leq Q_t^{(\pm)}(\eta)$. The case when $u > t_r$ is similar.

(2) The net $(Q_t^{(\pm)}(\xi))_{\xi \in F_t}$ (where F_t has the partial order ' $\subseteq$ ') is convergent in L^1 .

Proof: We have that $\sup_{\xi} \mathbb{E}[Q_t^{(\pm)}(\xi)] \leq \text{Var}(X) < \infty$. Therefore use Lemma(8.21) together with (1).

Now let $P_t^{(\pm)} = \lim_{\xi \rightarrow \infty} Q_t^{(\pm)}(\xi)$.

(3) $P^{(\pm)}$ is a non-negative presupermartingale.

Proof: Suppose that $s < t$ in $\mathbb{R}^+$. Then for $\xi \in F_s$ such that $t \in \xi$,

$$Q_s^{(\pm)}(\xi) \geq \mathbb{E}[Q_t^{(\pm)}(\xi \cap [t, \infty)) | \mathcal{F}_s].$$

Let $\xi \rightarrow \infty$ through those $\xi \in F_s$ such that $t \in \xi$, to give that $P_s^{(\pm)} \geq \mathbb{E}(P_t^{(\pm)} | \mathcal{F}_s)$ as required.

(4) $X = P^{(+)} - P^{(-)}$ and $\text{Var}(X) = \mathbb{E}(P_0^{(+)}) + \mathbb{E}(P_0^{(-)})$.

Proof: We have that $X_t = Q_t^{(+)}(\xi) - Q_t^{(-)}(\xi)$ for all $\xi \in F_t$. Let $\xi \rightarrow \infty$ to give the first result. From this we obtain that $\text{Var}(X) \leq \mathbb{E}(P_0^{(+)}) + \mathbb{E}(P_0^{(-)})$. For the reverse inequality, note that $\mathbb{E}[Q_0^{(+)}(\xi)] + \mathbb{E}[Q_0^{(-)}(\xi)] \leq \text{Var}(X)$ for all $\xi \in F_0$. Now let $\xi \rightarrow \infty$.

(5) If $X = U - V$, where U and V are non-negative presupermartingales, then

$U - P^{(+)} = V - P^{(-)}$ is a non-negative presupermartingale.

Proof: For $s < t$ in $\mathbb{R}^+$, $\mathbb{E}(X_s - X_t | \mathcal{F}_s)^+ \leq \mathbb{E}(U_s - U_t | \mathcal{F}_s)$. So for $t \in \mathbb{R}^+$, $Q_t^{(+)}(\xi) \leq U_t$ for all $\xi \in F_t$. Let $\xi \rightarrow \infty$ to give that $P_t^{(+)} \leq U_t$ a.s. Thus $U - P^{(+)}$ is non-negative.

Now suppose that $s < t$ in $\mathbb{R}^+$. Then for each $\xi \in F_s$ such that $t \in \xi$ we have that

$$\mathbb{E}[Q_s^{(+)}(\xi) - Q_t^{(+)}(\xi \cap [t, \infty)) | \mathcal{F}_s] \leq \mathbb{E}(U_s - U_t | \mathcal{F}_s).$$

Thus letting $\xi \rightarrow \infty$ subject to $t \in \xi$ gives that $\mathbb{E}(P_s^{(+)} - P_t^{(+)} | \mathcal{F}_s) \leq \mathbb{E}(U_s - U_t | \mathcal{F}_s)$.

It now follows that $U - P^{(+)}$ is a presupermartingale.

(6) If in the situation of (5) above $E(U_0) + E(V_0) = \text{Var}(X)$, then $U = P^{(+)}$ and $V = P^{(-)}$.

Proof: Let S denote the non-negative presupermartingale $U - P^{(+)} = V - P^{(-)}$. Then $\text{Var}(X) = E(U_0) + E(V_0) = E(P_0^{(+)}) + E(P_0^{(-)}) + 2E(S_0) = \text{Var}(X) + 2E(S_0)$. Thus $E(S_0) = 0$. But for each $t \in \mathbb{R}^+$, $E(S_t) \leq E(S_0)$ (since S is a presupermartingale), and so it follows that $S_t = 0$ as required.

(7) For each $t \in \mathbb{R}^+$, $P_s^{(\pm)} \rightarrow P_t^{(\pm)}$ in L^1 as $s \downarrow t$ and so by Proposition(8.13), $P^{(+)}$ and $P^{(-)}$ have càdlàg versions Y and Z respectively.

Proof: By Proposition(8.13), the limit $\lim_{s \downarrow t} P_s^{(\pm)}$ exists in L^1 . Denote this limit by $R_t^{(\pm)}$. Now $R^{(\pm)}$ is a non-negative presupermartingale; since if $s < t$ are in $\mathbb{R}^+$, and $A \in \mathcal{F}_s$, then for all $\delta > 0$, $\int_A P_{s+\delta}^{(\pm)} d\mathbb{P} \geq \int_A P_{t+\delta}^{(\pm)} d\mathbb{P}$; so letting $\delta \downarrow 0$ gives that $\int_A R_s^{(\pm)} d\mathbb{P} \geq \int_A R_t^{(\pm)} d\mathbb{P}$. Now since $P^{(\pm)}$ is a presupermartingale, it follows that $R^{(\pm)} \leq P^{(\pm)}$ (since $R_t^{(\pm)} = \lim_{s \downarrow t} E(P_s^{(\pm)} | \mathcal{F}_t)$ in L^1). Now since X is càdlàg we have from (4), that $X_t = \lim_{s \downarrow t} X_s = R_t^{(+)} - R_t^{(-)}$. Thus by (5), $P^{(\pm)} \leq R^{(\pm)}$, and so $R^{(\pm)} = P^{(\pm)}$ as required.

Now we see from (1)—(7) that the theorem holds. □

(8.22)Corollary. *The class of all quasimartingales is stable under stopping (see §5.3).*

Proof: The class of all non-negative supermartingales is stable under stopping by Theorem(8.18), and so therefore is the class of their differences. But this is the class of quasimartingales by Theorem(8.20). □

§8.5 LOCAL MARTINGALES

Recall the work of §5.3. Theorem(8.18) above proves that the classes of all martingales, submartingales and supermartingales are stable under stopping. Thus we can talk about *local martingales*, *local submartingales* and *local supermartingales*. By

Corollary(8.22) we can talk about *local quasimartingales*. Now if X is an L^p -martingale ($1 \leq p \leq \infty$) and σ is a stopping time, then the martingale X^σ is an L^p -martingale, since $X_{\sigma \wedge t} = \mathbb{E}(X_t | \mathcal{F}_{\sigma \wedge t}) \in L^p$ by the Optional Stopping Theorem(8.17). Thus we can also talk about *local L^p -martingales*. The class of L^p -integrators ($0 \leq p < \infty$) is also stable under stopping (since $S^\sigma = \mathbf{1}_{(0, \sigma]} \cdot S$ whenever S is an integrator and σ is a stopping time). Hence we can also talk about *local L^p -integrators*. All local L^p -integrators are integrators by Theorem(5.18). By Theorem(8.12) and Theorem(5.18) we have the following theorem.

(8.23)Theorem. *All local martingales, local submartingales, local supermartingales, and local quasimartingales are integrators.* $\square$

In fact more is true. We have the following result.

(8.24)Theorem. *All local martingales, local submartingales, local supermartingales, and local quasimartingales are local L^1 -integrators.*

In order to prove this we need the following lemma. This can be found in Edwards [Ed] (Theorem 5).

(8.25)Lemma. *If X is a non-negative supermartingale, then X is a local L^1 -integrator.*

Proof: Suppose that $0 < c < \infty$, and let σ be the début of $[X > c]$. Let A be the increasing non-negative càdlàg process given by $A_t = \mathbf{1}_{[\sigma \leq t]} X_\sigma$. (For each Borel subset B of $\mathbb{R} \setminus \{0\}$, $[A_t \in B] = [\sigma \leq t] \cap [X_\sigma \in B] \in \mathcal{F}_t$. Thus A is adapted.) Let $Z = \mathbf{1}_{[0, \sigma)} X$. So $X^\sigma = Z + A - X_0$, and it follows that $Z = X^\sigma - A + X_0$ is a supermartingale (since X^σ is a supermartingale, and A is increasing). Moreover $0 \leq Z \leq c$, and so by Theorem(8.10)(i), we have that for each $b \in \mathbb{R}^+$,

$$\mathbb{E}[(H \cdot Z)_b^2] \leq 8c \|H\|_\infty \mathbb{E}(Z_0) \quad \text{for all } H \in \mathcal{E}.$$

Thus $\int_0^b \dots dZ: \mathcal{E} \rightarrow L^2$ is continuous. Hence by Theorem(4.30), Z is an L^2 -integrator, and therefore certainly an L^1 -integrator. Now since A is an increasing L^1 càdlàg process, we have that A is an L^1 -integrator by Theorem(5.30)(a)(i). Thus $X^\sigma = Z + A - X_0$ is an L^1 -integrator.

We now let c run through the set $\mathbb{N}_{>0}$. Let σ_n be the début of $[X > n]$. Then $\sigma_n \uparrow \infty$ and X^{σ_n} is an L^1 -integrator. So $(\sigma_n)_n$ is a reducing sequence of X into the class of all L^1 -integrators, and X is therefore a local L^1 -integrator as required. $\square$

Proof of (8.24): By Theorem(5.16), it is enough to prove that all martingales, submartingales, supermartingales and quasimartingales are local L^1 -integrators. But since all such processes are local quasimartingales (take the sequence of constant stopping times $0, 1, 2, \dots$ as the reducing sequence), it is enough to consider quasimartingales. But by Theorem(8.20) any quasimartingale is the difference of two non-negative supermartingales, and so it is a local L^1 -integrator by Lemma(8.25). $\square$

Note. Suppose that we are given a submartingale X , and suppose that for some $b \in \mathbb{R}^+$, we take Y and Z to be càdlàg processes given by $Y_t = E(X_b^+ | \mathcal{F}_t)$ and $Z_t = Y_t - X_t$ for $t \leq b$, and $Y_t = Y_b$, $Z_t = Z_b$ for $t > b$. Then we have that $X_{b \wedge t} = Y_t - Z_t$ for all t and that Y and Z are non-negative supermartingales. In this way we see that X is a local difference of two non-negative supermartingales. Thus we can avoid the use of the Rao-Stricker Theorem when proving that all local martingales, local submartingales and local supermartingales (but not local quasimartingales) are local L^1 -integrators.

We can now use Theorem(8.24) to prove the following important result, which is a continuous-time analogue of Proposition(7.7).

(8.26) Theorem.

(i) If M is a local martingale, and H is a locally bounded previsible process, then

$H \cdot M$ is a local martingale.

- (ii) If X is a local submartingale (resp. local supermartingale), and H is a non-negative locally bounded previsible process, then $H \cdot X$ is a local submartingale (resp. local supermartingale).

Proof:

(i) By Theorem(8.24) there are stopping times $\sigma_1, \sigma_2, \dots$ such that $\sigma_n \uparrow \infty$, and such that for each n , M^{σ_n} is a martingale and an L^1 -integrator, and H is bounded on $(0, \sigma_n]$. Now since $(H \cdot M)^{\sigma_n} = (\mathbf{1}_{(0, \sigma_n]} H) \cdot (M^{\sigma_n})$ (use the formula ' $S^\sigma = \mathbf{1}_{(0, \sigma]} \cdot S$ ') and $\mathbf{1}_{(0, \sigma_n]} H \in \Lambda^1(M^{\sigma_n}, L^1)$, it follows from Lemma(8.16)(i) that $(H \cdot M)^{\sigma_n}$ is a martingale. So $H \cdot M$ is a local martingale as required.

(ii) Similar (use Lemma(8.16)(ii)). □

§8.6 MUTUAL VARIATION OF MARTINGALES AND LOCAL MARTINGALES

When M and N are local martingales we can give the following characterization of $[M, N]$. In this case, this is often given as definition of $[M, N]$. See for example Theorem IV.26.4 of Rogers & Williams [RW] together with the comment after its statement.

(8.27)Theorem. Suppose that M and N are local martingales. Then $[M, N]$ is the unique (up to indistinguishability) càdlàg process that satisfies the following properties.

- (1) $[M, N]$ has finite variation and $[M, N]_0 = 0$.
- (2) $MN - [M, N]$ is a local martingale.
- (3) $\Delta[M, N] = \Delta M \Delta N$.

In order to prove this it is necessary to prove a couple of lemmas first.

(8.28) Lemma. If A is a integrable-variation process with $A_0 = 0$, then for each $b \in \mathbb{R}^+$, and each bounded martingale Z ,

$$\mathbb{E}(Z_b A_b) = \mathbb{E} \left[\int_0^b Z dA \right],$$

where the integral is a pathwise Stieltjes integral.

Proof: Let $0 = t_0 < \dots < t_r = b$. Then

$$\begin{aligned} \mathbb{E}(Z_b A_b) &= \mathbb{E} \left[\sum_{k=1}^r Z_b (A_{t_k} - A_{t_{k-1}}) \right] \\ &= \mathbb{E} \left[\sum_{k=1}^r \mathbb{E}[Z_b (A_{t_k} - A_{t_{k-1}}) | \mathcal{F}_{t_k}] \right] \\ &= \mathbb{E} \left[\sum_{k=1}^r Z_{t_k} (A_{t_k} - A_{t_{k-1}}) \right] \quad (\text{since } Z_{t_k} = \mathbb{E}(Z_b | \mathcal{F}_{t_k})) \\ &= \mathbb{E} \left[\int_0^b U dA \right], \end{aligned}$$

where $U_t(\omega) = \sum_{k=1}^r Z_{t_k}(\omega) \mathbf{1}_{(t_{k-1}, t_k]}(t)$.

Now let $\sup_k (t_k - t_{k-1}) \rightarrow 0$ through a sequence of partitions of $[0, b]$, and use the dominated convergence theorem twice on the right hand side, to give the required result. $\square$

(8.29) Lemma. If M is a continuous local martingale with finite variation, and $M_0 = 0$, then $M = 0$.

Proof: Let $(\sigma_n)_n$ be a reducing sequence of M into the class of all martingales. By replacing σ_n by $\sigma_n \wedge \tau_n$, where τ_n is the debut of $[V_M > n]$ (V_M is the variation process of M ; see §5.5), we may assume that $V_{M^{\sigma_n}}$ is bounded for all n (since $V_{M^{\sigma_n}} = V_M^{\sigma_n}$). Thus it is enough to prove the following statement (just apply it to M^{σ_n}).

(†) If M is a martingale with bounded variation (i.e. V_M is bounded on $\mathbb{R}^+ \times \Omega$) and $M_0 = 0$, then $M = 0$.

To prove this, suppose that $b \in \mathbb{R}^+$. Then by taking $Z = A = M$ in Lemma(8.28), we obtain that $\mathbb{E}(M_b^2) = \mathbb{E}(\int_0^b M dM)$. But M is continuous, and therefore previsible (when restricted to $\mathbb{R}^{++} \times \Omega$). So $\int_0^b M dM = (M \cdot M)_b$. Now by Lemma(8.16)(ii) (M is a L^1 -integrator, since it has integrable variation), $M \cdot M$ is a martingale. Thus $\mathbb{E}[(M \cdot M)_b] = 0$, and so $\mathbb{E}(M_b^2) = 0$. This is for all $b \in \mathbb{R}^+$; so $M = 0$. $\square$

Proof of (8.27): We know already (Proposition(6.3)(ii) and (iv)) that $[M, N]$ satisfies (1) and (3). Now $MN - [M, N] = M_0 N_0 + M_- \cdot N + N_- \cdot M$ (by Proposition(6.3)(iii)). So $MN - [M, N]$ is a local martingale, since $M_- \cdot N$ and $N_- \cdot M$ are local martingales by Theorem(8.26)(i).

Now suppose that A is any càdlàg process satisfying (1), (2) and (3). I.e. the following hold.

- (i) A has finite variation and $A_0 = 0$.
- (ii) $MN - A$ is a local martingale.
- (iii) $\Delta A = \Delta M \Delta N$.

Let $Z = [M, N] - A$. Then from (1), (2), (3), (i), (ii) and (iii), we see that Z is a local martingale, has finite variation, $Z_0 = 0$ and $\Delta Z = 0$. The last property says that Z is continuous. So Lemma(8.29) applies to Z and gives that $Z = 0$. So $A = [M, N]$, and the required uniqueness is proved. $\square$

§8.7 L^2 -MARTINGALES AND THE STRASBOURG APPROACH

In this section we shall present the long awaited proof that the approach to the Itô calculus given in this thesis is consistent with the Strasbourg approach. In order to do this we shall use the following version of Theorem(8.27) for L^2 -martingales. (See Theorem IV.26.2 of Rogers and Williams [RW]. He proves it by a different method.)

(8.30)Theorem. Suppose that M and N are L^2 -martingales. Then $[M, N]$ is the unique càdlàg process such that the following hold.

- (1) $[M, N]$ has integrable variation and $[M, N]_0 = 0$.
- (2) $MN - [M, N]$ is a martingale.
- (3) $\Delta[M, N] = \Delta M \Delta N$.

Proof: The uniqueness follows immediately from Theorem(8.27). So we only have to show that $[M, N]$ satisfies (1), (2) and (3) above. By polarization we need only consider the case where $M = N$ (i.e. use the formula $[M, N] = [\frac{1}{2}(M + N)] - [\frac{1}{2}(M - N)]$). First I claim that M^2 is a submartingale. Certainly it is an L^1 -process. If $s \leq t$ in $\mathbb{R}^+$, then $M_t^2 - M_s^2 = (M_t - M_s)^2 + 2M_s(M_t - M_s)$. So by applying $\mathbb{E}[\cdot|\mathcal{F}_s]$, and noting that $\mathbb{E}[M_s(M_t - M_s)|\mathcal{F}_s] = M_s\mathbb{E}[(M_t - M_s)|\mathcal{F}_s] = 0$, we obtain the following statement which is enough to prove that M^2 is a submartingale.

- (a) $\mathbb{E}(M_t^2 - M_s^2|\mathcal{F}_s) = \mathbb{E}[(M_t - M_s)^2|\mathcal{F}_s] \geq 0$, for each $s \leq t$ in $\mathbb{R}^+$.

So if σ is a stopping time bounded above by $b \in \mathbb{R}^+$, then by the Optional Stopping Theorem(8.17) $0 \leq M_\sigma^2 \leq \mathbb{E}(M_b^2|\mathcal{F}_\sigma)$. Thus using Proposition(8.6) we obtain that

- (b) The set $\{M_\sigma^2 \mid \sigma \leq b \text{ is a stopping time}\}$ is uniformly integrable for each $b \in \mathbb{R}^+$.

Now let $X = M^2 - [M]$. Then by Theorem(8.27), X is a local martingale. Let $(\sigma_n)_n$ be a reducing sequence of X into the class of all martingales. For each bounded stopping time α , $\mathbb{E}(X_\alpha^{\sigma_n}) = 0$ by the Optional Stopping Theorem(8.17). I.e. $\mathbb{E}([M]_{\sigma_n \wedge \alpha}) = \mathbb{E}(M_{\sigma_n \wedge \alpha}^2 - M_0^2)$. But $(M_{\sigma_n \wedge \alpha}^2)_n$ is uniformly integrable by (b), and $([M]_{\sigma_n \wedge \alpha})_n$ is increasing. So letting $n \rightarrow \infty$ gives that

- (c) $\mathbb{E}([M]_\alpha) = \mathbb{E}(M_\alpha^2 - M_0^2)$ for all bounded stopping times α .

Taking α to be the constant stopping time t gives that $\mathbb{E}([M]_t) < \infty$, and so $[M]_t$ is an L^1 -process. Thus (1) is verified. We may rewrite (c) as $\mathbb{E}(X_\alpha - X_0) = 0$.

So by Lemma(8.19), X is a martingale. I.e. (2) is proven. We already know (3) (Theorem(8.27) or Proposition(6.3)(iv)). $\square$

Now for each L^2 -martingale M , we associate a measure μ_M on $(\mathbb{R}^{++} \times \Omega, \mathcal{P})$ by defining $\mu_M E = \mathbb{E}(\int_0^\infty \mathbf{1}_E d[M])$. So for each non-negative previsible process $H: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}^+$,

$$(†) \quad \int_{\mathbb{R}^{++} \times \Omega} H d\mu_M = \mathbb{E} \left(\int_0^\infty H d[M] \right)$$

We can now prove the following theorem.

(8.31)Theorem. Suppose that $M: \mathbb{R}^+ \times \Omega \rightarrow \mathbb{R}$ is an L^2 -martingale. Then the following hold.

- (i) M is an L^2 -integrator, and $H \cdot M$ is an L^2 -martingale for all $H \in \Lambda^1(M, L^2)$.
- (ii) $\mathcal{L}^1(M, L^2) = \mathcal{L}^2(\mathbb{R}^{++} \times \Omega, \mathcal{P}, \mu_M)$.
- (iii) $\mathbb{E}(\int H dM \int K dM) = \int_{\mathbb{R}^{++} \times \Omega} H K d\mu_M$ for all $H, K \in \mathcal{L}^1(M, L^2)$.
- (iv) $\Phi_M^2(H) = (\int_{\mathbb{R}^{++} \times \Omega} H^2 d\mu_M)^{1/2}$ for all previsible $H: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}$.

Proof: We already know that (i) holds (see Theorem(8.15)). Next we shall prove (iii). Suppose that $H, K \in \mathcal{L}^1(M, L^2)$. Then let $X = (H \cdot M)(K \cdot M) - [H \cdot M, K \cdot M]$. By Theorem(8.30), X is a martingale. Thus $\mathbb{E}(X_t) = 0$. But by Theorem(6.8), $X = (H \cdot M)(K \cdot M) - (HK) \cdot [M]$. So $\mathbb{E}(X_t) = 0$ gives that

$$\mathbb{E} \left(\int_0^t H dM \int_0^t K dM \right) = \mathbb{E} \left(\int_0^t H K d[M] \right) = \int_{(0,t] \times \Omega} H K d\mu_M.$$

Let $t \rightarrow \infty$ to give (iii). Now we see that μ_M satisfies the hypothesis of Theorem(5.34) and so the rest of the result now follows from that theorem. $\square$

We have the following corollary.

(8.32)Corollary. *If M is an L^2 -martingale, then μ_M is the unique positive measure on $(\mathbb{R}^{++} \times \Omega, \mathcal{P})$ such that*

$$(\dagger) \quad \mu_M((s, t] \times F) = \mathbb{E}[\mathbf{1}_F(M_t - M_s)^2] = \mathbb{E}[\mathbf{1}_F(M_t^2 - M_s^2)],$$

whenever $s \leq t$ in $\mathbb{R}^+$ and $F \in \mathcal{F}_s$.

Proof: Putting $H = K = \mathbf{1}_{(s, t] \times F}$ in (iii) of Theorem(8.31) gives the first equality of $(\dagger)$. The second equality follows from (a) in the proof of Theorem(8.30). For uniqueness note that $(\dagger)$ determines μ_M on the set of previsible rectangles. This is a semi-ring that generates $\mathcal{P}$ as a σ -field. So the usual monotone class argument proves the result. $\square$

So from Theorem(8.31) we see that for each L^2 -martingale M , we have that $\mathcal{L}^1(M, L^2) = \mathcal{L}^2(\mu_M)$, and that the stochastic integral $\int \dots dM: \mathcal{L}^1(M, L^2) \rightarrow L^2$ is the unique linear isometry from $\mathcal{L}^2(\mu_M)$ to L^2 extending the mapping $\int \dots dM: \mathcal{E} \rightarrow L^2$ that is defined in Definition(4.6). This is precisely how (the first version of) the stochastic integral is defined in the Strasbourg approach.

The next step in the Strasbourg approach (after defining $H \cdot M$ as in Definition(4.13)(b)) is to use localization to define integration with respect to local L^2 -martingales (only $H \cdot M$ is defined). This is clearly consistent with our integral by the work of §5.3. Integrals with respect to finite-variation processes are defined pathwise in the Strasbourg approach. By Theorem(5.30) this is consistent with our integral. Then the integrals are taken with respect to processes that are a sum of a local L^2 -martingale and a finite-variation process. We thus arrive at the following conclusion.

(8.33)Conclusion. *Our stochastic integral is defined and is equal to the Strasbourg integral, whenever the latter is defined.* $\square$

When we have a **continuous** local martingale M , we can determine $\Lambda^1(M)$, and see that it is the same set as that given by the Strasbourg approach. (A continuous

local martingale M is a local bounded martingale, and hence a local L^2 -martingale. For if $(\sigma_n)_n$ is a reducing sequence of M into the class of all martingales, let τ_n be the début of $[|M - M_0| > n]$. Then $(\sigma_n \wedge \tau_n)_n$ is a reducing sequence of M into the class of all bounded martingales.)

(8.34)Theorem. *Suppose that M is a continuous local martingale. Then $\Lambda^1(M)$ is the set of all previsible processes H such that $\int_0^t H^2 d[M] < \infty$ a.s. for all $t \in \mathbb{R}^+$.*

Proof: If $H \in \Lambda^1(M)$, then by Theorem(6.8), $H^2 \in \Lambda^1([M])$, and so $\int_0^t H^2 d[M] < \infty$ a.s. for all $t \in \mathbb{R}^+$. Now suppose conversely that H is a previsible process such that $\int_0^t H^2 d[M] < \infty$ a.s. for all $t \in \mathbb{R}^+$. Then let $(\sigma_n)_n$ be a reducing sequence of M into the class of all L^2 -martingales. Replacing σ_n with the minimum of σ_n and the début of $[H^2 \cdot [M] > n]$, we may assume that $(H^2 \cdot [M])^{\sigma_n}$ is bounded. Thus for each n , $\int_0^\infty (\mathbf{1}_{(0, \sigma_n \wedge t]} H)^2 d[M^{\sigma_n}] = (H^2 \cdot [M])_t^{\sigma_n}$ is a bounded random variable, and so has finite expectation. Now apply Theorem(8.31)(ii) to give that $\mathbf{1}_{(0, \sigma_n \wedge t]} H \in \mathcal{L}^1(M^{\sigma_n}, L^2) \subseteq \mathcal{L}^1(M^{\sigma_n})$. Thus $\mathbf{1}_{(0, \sigma_n]} H \in \Lambda^1(M^{\sigma_n})$. But $M^{\sigma_n} = \mathbf{1}_{(0, \sigma_n]} \cdot M$ and so by Theorem(4.21)(iv), $\mathbf{1}_{(0, \sigma_n]} H = \mathbf{1}_{(0, \sigma_n]}^2 H \in \Lambda^1(M)$. Now use Theorem(5.21) to give that $H \in \Lambda^1(M)$ as required. $\square$

CHAPTER 9: THE CHARACTERIZATION OF INTEGRATORS AS SEMIMARTINGALES

The purpose of this chapter is to present a proof of the theorem of Bichteler, Dellacherie, Mokobodzki and Meyer that characterizes integrators as semimartingales. In doing this we shall also find a characterization of local L^1 -integrators, and prove the existence and uniqueness of various versions of the Doob-Meyer decomposition of certain types of process.

As usual we shall assume unless otherwise stated that $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t \in \mathbb{R}^+})$ is a standard setup.

§9.1 DEFINITIONS

(9.1)Definition. A *semimartingale* is a càdlàg process S such that $S = M + A$ for some local martingale M , and some finite-variation process A . $\square$

We have immediately, from Theorem(8.23) and Theorem(5.30), the following theorem (which is clearly one half of the characterization theorem).

(9.2)Theorem. All *semimartingales* are *integrators*. $\square$

§9.2 SOME CHARACTERIZATIONS OF LOCAL L^1 -INTEGRATORS AND THE DOOB-MEYER DECOMPOSITION

In this section we shall derive a Doob-Meyer decomposition result for local L^1 -integrators. In order to get uniqueness we shall need the following concept.

(9.3)Definition. An integrable-variation process A is said to be *natural* if and only if $[Z, A]$ is a martingale for all bounded martingales Z . (Note that by Proposition(6.6), $[Z, A]_t = \sum_{u \leq t} \Delta Z_u \Delta A_u$.) $\square$

Note.

1. Since $[Z, A]^\sigma = [Z', A^\sigma]$ (by Proposition(6.6)), the class of all natural integrable-variation processes is closed under stopping. Thus we may talk about *local natural integrable-variation processes*.

2. It has been shown (firstly by C. Doléans) that an integrable-variation process is natural if and only if it previsible (when restricted to $\mathbb{R}^{++} \times \Omega$). We shall not need this result, and content ourselves with referring the reader to Pellaumail [Pe] or Theorem 9.12 of Métivier & Pellaumail [MP] for more information. Also a development of the Doob-Meyer decomposition in terms of previsible processes is given in Section 6 of Chapter VI of Rogers & Williams [RW].

Our aim here is to get a decomposition of each local L^1 -integrator as the sum of a local martingale and a local natural integrable-variation process. In order to do this we shall require the following tool.

(9.4)Definition.

- (a) For each $b \in \mathbb{R}^+$, let $\mathcal{P}_b$ denote the restriction of the previsible σ -field $\mathcal{P}$ to $(0, b] \times \Omega$ (i.e. the σ -field of all previsible subsets of $(0, b] \times \Omega$).
- (b) For each L^1 -integrator S , define $\lambda_S: \bigcup_{b \in \mathbb{R}^+} \mathcal{P}_b \rightarrow \mathbb{R}$ by $\lambda_S E = \mathbb{E}(\int \mathbf{1}_E dS)$ ($\mathbf{1}_E \in \mathcal{L}^1(S, L^1)$, since $0 \leq \mathbf{1}_E \leq \mathbf{1}_{(0, b] \times \Omega}$). $\square$

(9.5)Proposition. Suppose that S is an L^1 -integrator.

- (i) For each $b \in \mathbb{R}^+$, λ_S is a signed measure on $((0, b] \times \Omega, \mathcal{P}_b)$.
- (ii) For each previsible rectangle $R = (s, t] \times F$ ($s \leq t$ in $\mathbb{R}^+$ and $F \in \mathcal{F}_s$), $\lambda_S R = \int_F (S_t - S_s) d\mathbb{P}$.

- (iii) S is a martingale if and only if $\lambda_S = 0$.
(iv) S is a submartingale if and only if $\lambda_S \geq 0$.

Proof:

(i) Clearly λ_S is finitely additive. Now if $E_1, E_2, \dots \in \mathcal{P}_b$, and $E_n \uparrow E$, then $\mathbf{1}_{E_n} \uparrow \mathbf{1}_E$. So by the Monotone Convergence Theorem (4.16), we see that $\int \mathbf{1}_{E_n} dS \rightarrow \int \mathbf{1}_E dS$ in L^1 -norm. Take expectations to give that $\lambda_S E_n \rightarrow \lambda_S E$. These facts are enough to show that λ_S is a measure.

(ii) We have that $\int \mathbf{1}_R dS = \mathbf{1}_F(S_t - S_s)$. Now take expectations to obtain the result.

(iii) Suppose that $\lambda_S = 0$. Then for each $s \leq t$ in $\mathbb{R}^+$ and $F \in \mathcal{F}_s$, we have by (ii) that $\int_F (S_t - S_s) d\mathbb{P} = 0$. So S must be a martingale.

Now suppose conversely that S is a martingale and that $E \in \mathcal{P}_b$ for some $b \in \mathbb{R}^+$. Then by Proposition(8.16)(i), $\mathbf{1}_E \cdot S$ is a martingale, and so $E[(\mathbf{1}_E \cdot S)_b] = 0$. But this just says that $\lambda_S E = 0$.

(iv) This similar to (iii). □

(9.6)Lemma. Suppose that $\lambda: \bigcup_b \mathcal{P}_b \rightarrow \mathbb{R}$ is a signed measure on each $((0, b] \times \Omega, \mathcal{P})$ for each $b \in \mathbb{R}^+$ and that all evanescent sets are λ -null. Then there exists a unique natural integrable-variation process A with $A_0 = 0$ such that $\lambda = \lambda_A$. If further $\lambda \geq 0$, then A is increasing.

Proof: First suppose that A is as stated in the conclusion, and that Z is a bounded martingale. Then since $[Z, A]$ is a martingale, $E([Z, A]_t) = 0$. But by Proposition(6.6), $[Z, A]_t = \sum_{u \leq t} \Delta Z_u \Delta A_u = \int_0^t (Z - Z_-) dA$. So $E(\int_0^t Z dA) = E(\int_0^t Z_- dA)$. But the left hand side is just $E(Z_t A_t)$ by Lemma(8.28). The right hand side however is $\int_{(0, t] \times \Omega} Z_- d\lambda$ (since $\lambda = \lambda_A$). Thus we have proved that

$$(1) \quad E(Z_t A_t) = \int_{(0, t] \times \Omega} Z_- d\lambda \quad \text{for all bounded martingales } Z.$$

Now Z_t can be any bounded $\mathcal{F}_t$ -measurable random variable desired (let Z be the bounded martingale such that $Z_t = \mathbb{E}(U|\mathcal{F}_t)$ for all t , where U is the desired random variable, by using Theorem(8.14)(i)). So A_t is uniquely determined by (1).

Now we consider the existence of A . Write $\lambda = \lambda^+ - \lambda^-$ (the Jordan decomposition of signed measures). Then it is enough to prove existence for $\lambda^\pm$. So we may assume that $\lambda \geq 0$.

For each $X \in L^\infty$, let $Z(X)$ denote the unique (up to indistinguishability) bounded martingale such that $Z_t(X) = \mathbb{E}(X|\mathcal{F}_t)$. For each $t \in \mathbb{R}^+$, define a measure ν_t on $(\Omega, \mathcal{F})$ by

$$(2) \quad \nu_t(F) = \int_{(0,t] \times \Omega} Z_-(\mathbf{1}_F) d\lambda.$$

(This is well defined as a function on $\mathcal{F}$, since evanescent sets are λ -null. It is clear that ν_t is finitely additive. If $F_1, F_2, \dots \in \mathcal{F}$ and $F_n \uparrow F$, then $Z_-(\mathbf{1}_{F_n}) \uparrow Z_-(\mathbf{1}_F)$ outside an evanescent set. Thus the monotone convergence theorem gives that $\nu_t F_n \uparrow \nu_t F$, and so ν_t is a measure as stated.) Now all $\mathbb{P}$ -null sets are ν_t -null, and so by the Radon-Nikodym theorem there exists $A_t \in L^1$ such that $d\nu_t = A_t d\mathbb{P}$.

From (2) we see that if $s \leq t$, then $\nu_s \leq \nu_t$, and so $A_s \leq A_t$ a.s. Also $\nu_0 = 0$, and so $A_0 = 0$ a.s. I claim that $A: \mathbb{R}^+ \rightarrow L^1$ is adapted. In order to prove this we shall need to know that

$$(3) \quad \mathbb{E}(XA_t) = \int_{(0,t] \times \Omega} Z_-(X) d\lambda \quad \text{for all } X \in L^\infty.$$

To prove (3), note that it is clear when $X = \mathbf{1}_F$ ($F \in \mathcal{F}$) by (2). Hence by linearity (3), holds for all simple $X \in L^\infty$. For a general X , derive (3) by uniform approximation of X by simple elements of L^∞ .

Now suppose that $X \in L^\infty$. Then $\mathbb{E}[X\mathbb{E}(A_t|\mathcal{F}_t)|\mathcal{F}_t] = \mathbb{E}(X|\mathcal{F}_t)\mathbb{E}(A_t|\mathcal{F}_t) = \mathbb{E}[\mathbb{E}(X|\mathcal{F}_t)A_t|\mathcal{F}_t]$. So taking expectations gives that $\mathbb{E}[X\mathbb{E}(A_t|\mathcal{F}_t)] = \mathbb{E}[\mathbb{E}(X|\mathcal{F}_t)A_t]$.

But since $Z_-(\mathbb{E}(X|\mathcal{F}_t)) = Z_-(X)$ on the set $(0, t] \times \Omega$, (3) gives that $\mathbb{E}[\mathbb{E}(X|\mathcal{F}_t)A_t] = \mathbb{E}(XA_t)$, and so $\mathbb{E}[X\mathbb{E}(A_t|\mathcal{F}_t)] = \mathbb{E}(XA_t)$. But this is for all $X \in L^\infty$; so $\mathbb{E}(A_t|\mathcal{F}_t) = A_t$. Therefore A_t is $\mathcal{F}_t$ -measurable as required.

Now putting $X = 1$ in (3) gives that $\mathbb{E}(A_t) = \lambda((0, t] \times \Omega)$, and so $t \mapsto \mathbb{E}(A_t)$ is right continuous. Thus by Theorem(8.14)(ii), A has a càdlàg version, and so we may assume that A is a càdlàg process. Now by defining a set E as in the proof of Proposition(6.3)(i), we can prove that A is an increasing process (we already know that $A_s \leq A_t$ a.s. whenever $s \leq t$). We have that $A_0 = 0$, so it remains to check that $\lambda = \lambda_A$, and that $[Z, A]$ is a martingale for all bounded martingales Z .

Suppose that $s < t$ are in $\mathbb{R}^+$ and that $F \in \mathcal{F}_s$. Then since $Z_-(\mathbf{1}_F) = \mathbf{1}_{\mathbb{R}^+ \times F}$ on the set $(s, t] \times \Omega$, it follows that

$$\begin{aligned} \mathbb{E}[\mathbf{1}_F(A_t - A_s)] &= \int_{(s, t] \times \Omega} Z_-(\mathbf{1}_F) d\lambda \quad (\text{from (3) by subtraction}) \\ &= \lambda((s, t] \times F). \end{aligned}$$

But the left hand side is just $\lambda_A((s, t] \times F)$. So λ and λ_A agree on the collection of previsible rectangles. Now by a monotone class argument $\lambda = \lambda_A$.

Putting $X = Z_t$ in (3) gives that (1) above holds for this A . By reversing the argument used there to deduce (1), we see that $\mathbb{E}([Z, A]_t) = 0$ for each bounded martingale Z . Now suppose that Z is a bounded martingale and that σ is a bounded stopping time. Then from the formula $[Z, A] = \sum_{u \leq t} \Delta Z_u \Delta A_u$, we clearly have that $[Z, A]^\sigma = [Z^\sigma, A]$. Thus $\mathbb{E}([Z, A]_t^\sigma) = 0$, and so taking t to be an upper bound for σ , this says that $\mathbb{E}([Z, A]_\sigma) = 0$. Thus by Lemma(8.19), $[Z, A]$ is a martingale as required. $\square$

Here is the first decomposition result.

(9.7)Theorem (The Doob-Meyer Decomposition).

Suppose that S is an L^1 -integrator. Then there exist unique càdlàg processes M and A such the following properties hold.

- (i) $S = M + A$.
- (ii) M is a martingale and an L^1 -integrator.
- (iii) A is a natural integrable-variation process with $A_0 = 0$.

Proof: Suppose that A is any natural integrable-variation process, and let $M = S - A$. Then M and A are L^1 -integrators (see Theorem(5.30)), and $\lambda_M = \lambda_S - \lambda_A$. By Proposition(9.5)(iii), M is a martingale if and only if $\lambda_M = 0$. But this is so if and only if $\lambda_A = \lambda_S$. So by Lemma(9.6), A is uniquely determined by (i), (ii) and (iii). For the existence of the decomposition just take, using Lemma(9.6), A to be such that $\lambda_A = \lambda_S$. $\square$

We can now give the local version of this result.

(9.8)Theorem (The Doob-Meyer Decomposition).

Suppose that S is a local L^1 -integrator. Then there exist unique càdlàg processes M and A such the following properties hold.

- (i) $S = M + A$.
- (ii) M is a local martingale.
- (iii) A is a local natural integrable-variation process with $A_0 = 0$.

Proof: Let $(\sigma_n)_n$ be a reducing sequence of S into the class of all L^1 -integrators. Let $S^{\sigma_n} = M^{(n)} + A^{(n)}$ be the decomposition of S^{σ_n} that is given by Theorem(9.7). Now suppose that $n, m \in \mathbb{N}$ are such that $n \leq m$. Then $S^{\sigma_n} = (M^{(m)})^{\sigma_n} + (A^{(m)})^{\sigma_n}$ is also a decomposition of S^{σ_n} that satisfies the properties of Theorem(9.7). Thus by uniqueness in that theorem, we have that $M^{(n)} = (M^{(m)})^{\sigma_n}$ and $A^{(n)} = (A^{(m)})^{\sigma_n}$. Therefore there are càdlàg processes M and A such that $M^{\sigma_n} = M^{(n)}$ and $A^{\sigma_n} = A^{(n)}$ for all n . These M and A are easily seen to satisfy (i), (ii) and (iii).

Now we consider the uniqueness of the decomposition. Suppose that $S = \tilde{M} + \tilde{A}$ is another decomposition of S satisfying (ii) and (iii). Then take $(\tau_n)_n$ to be a

sequence of stopping times that is a reducing sequence of M and $\tilde{M}$ into the class of all martingales, and a reducing sequence of A and $\tilde{A}$ into the class of all natural integrable-variation processes. Then $S^{\tau_n} = M^{\tau_n} + A^{\tau_n} = \tilde{M}^{\tau_n} + \tilde{A}^{\tau_n}$. So by the uniqueness in Theorem(9.7), $M^{\tau_n} = \tilde{M}^{\tau_n}$ and $A^{\tau_n} = \tilde{A}^{\tau_n}$. Thus $M = \tilde{M}$ and $A = \tilde{A}$ as required. $\square$

This now allows us to give the following characterization of local L^1 -integrators.

(9.9)Theorem. *Suppose that S is a càdlàg process. Then the following are equivalent.*

- (1) S is a local L^1 -integrator.
- (2) S is a local quasimartingale.
- (3) S is the sum of a local martingale and a local integrable-variation process.

In particular, we have that all local L^1 -integrators are semimartingales.

Proof: We have from Theorem(9.8), that (1) implies (3). From Theorem(8.24), we have that (2) implies (1). Thus to complete the proof we need to show that (3) implies (2). But if X is the sum of a martingale and an integrable-variation process, then it is easy to see that $\text{Var}_b(X) < \infty$ for all $b \in \mathbb{R}^+$. That (3) implies (2) is now clear. $\square$

Now concerning the Doob-Meyer decomposition of submartingales we have the following result. (Recall that all local submartingales are local L^1 -integrators.)

(9.10)Theorem.

- (i) *If X is a local submartingale, and $X = M + A$ is its Doob-Meyer decomposition (as given by Theorem(9.8)), then A is an increasing process.*
- (ii) *If X is a submartingale that is bounded below on $\mathbb{R}^+ \times \Omega$, and $X = M + A$ is its Doob-Meyer decomposition, then M is a martingale and A is an increasing L^1 -process.*

Proof:

(i) First I claim that if X is a submartingale and an L^1 -integrator, and $X = M + A$ is its Doob-Meyer decomposition as given in Theorem(9.7), then A is increasing. To prove this claim, note that $\lambda_A = \lambda_X \geq 0$ (see Theorem(9.5)(iv) and the proof of Theorem(9.7)). Thus A is increasing by Lemma(9.6). The required result now follows from this claim by localization (as in the proof of Theorem(9.8)).

(ii) By part (i), A is increasing. By adding a constant to X , we may assume that $X \geq 0$. Now the proof proceeds as the proof of Theorem(8.30) (with X , A and M here playing the roles of M^2 , $[M]$ and X of Theorem(8.30)). $\square$

The hypothesis in Theorem(9.10)(ii) that X is bounded below can clearly be weakened to that the set $\{X_\sigma \mid \sigma \leq b \text{ is a stopping time}\}$ is uniformly integrable for all $b \in \mathbb{R}^+$. Such processes are said to be of class (DL).

§9.3 CHANGE OF PROBABILITY

In order to prove the theorem of Bichteler et al., we shall have to consider the effect of changing the probability measure on $(\Omega, \mathcal{F})$.

(9.11)Definition. Two σ -finite measures μ and ν on a measurable space $(\Omega, \mathcal{F})$ are said to be *equivalent* if and only if they have the same null sets (i.e. for $E \in \mathcal{F}$, $\mu(E) = 0$ if and only if $\nu(E) = 0$). In this case, we have by the Radon-Nikodym theorem, that $d\nu = h d\mu$ and $d\mu = k d\nu$ for some $h \in L^1(\mu)$ and $k \in L^1(\nu)$. We must have that $hk = 1$ and $h, k > 0$ a.e. $\square$

Suppose that $\tilde{\mathbb{P}}$ is a probability on $(\Omega, \mathcal{F})$ equivalent to $\mathbb{P}$. Then $(\Omega, \mathcal{F}, \tilde{\mathbb{P}}, (\mathcal{F}_t)_t)$ is also a standard setup. The spaces $L^0(\mathbb{P})$ and $L^0(\tilde{\mathbb{P}})$ are the same, and have the same topology. Thus we see that the class of all integrators is unchanged by replacing $\mathbb{P}$ by $\tilde{\mathbb{P}}$. Also $\mathcal{L}^1(S)$, $\Lambda^1(S)$, $\int X dS$, $X \cdot S$ and $[R, S]$ remain unchanged. But the corresponding L^p objects (for $p > 0$) are changed (unless for example the densities in

both directions are bounded). The classes of all martingales and all local martingales are clearly also affected; but the class of semimartingales is not. We shall now prove this result. The proof illustrates something the power of our theory.

(9.12)Theorem. *Suppose that $\tilde{\mathbb{P}}$ is a probability on $(\Omega, \mathcal{F})$ equivalent to $\mathbb{P}$. Then the class of all $\tilde{\mathbb{P}}$ -semimartingales is the same as the class of all $\mathbb{P}$ -semimartingales.*

Proof: We shall write $\mathbb{E}$ for expectation with respect to $\mathbb{P}$, and $\tilde{\mathbb{E}}$ for expectation with respect to $\tilde{\mathbb{P}}$. Let $z \in L^1(\mathbb{P})$ and $w \in L^1(\tilde{\mathbb{P}})$ be such that $d\tilde{\mathbb{P}} = z d\mathbb{P}$ and $d\mathbb{P} = w d\tilde{\mathbb{P}}$. Now let Z be the $\mathbb{P}$ -martingale such that $Z_t = \mathbb{E}(z|\mathcal{F}_t)$ for all t , and W be the $\tilde{\mathbb{P}}$ -martingale such that $W_t = \tilde{\mathbb{E}}(w|\mathcal{F}_t)$ for all t (use Theorem(8.14)(i)). Then by the definition of conditional expectation, we see that $d\tilde{\mathbb{P}}|_{\mathcal{F}_t} = Z_t d\mathbb{P}|_{\mathcal{F}_t}$ and $d\mathbb{P}|_{\mathcal{F}_t} = W_t d\tilde{\mathbb{P}}|_{\mathcal{F}_t}$. Thus $ZW = 1$. Now Z and W are integrators by Theorem(8.12). Let $X = Z_- \cdot W$.

Now suppose that S is a $\mathbb{P}$ -semimartingale. Then $S = S_0 + M + A$, where M is a local $\mathbb{P}$ -martingale with $M_0 = 0$, and A is a finite-variation process with $A_0 = 0$. Now define

$$(1) \quad N = M + [M, X],$$

I claim that N is a local $\tilde{\mathbb{P}}$ -martingale. To prove this we shall use the differential notation that was discussed in §6.5. We have that

$$dN = (1 + dX) dM = (1 + Z_- dW) dM.$$

Therefore

$$\begin{aligned} d(NZ) &= N_- dZ + Z_- dN + dN dZ \quad (\text{by Proposition(6.3)(i)}) \\ &= N_- dZ + (Z_- + dZ)(1 + Z_- dW) dM \\ &= N_- dZ + (Z_- + Z_-^2 dW + dZ + Z_- dZ dW) dM \end{aligned}$$

$$\begin{aligned}
&= N_- dZ + Z_- (1 + Z_- dW + W_- dZ + dZ dW) dM \\
&= N_- dZ + Z_- (1 + d(ZW)) dM \\
&= N_- dZ + Z_- dM \quad (\text{since } ZW = 1).
\end{aligned}$$

I.e. $NZ = N_- \cdot Z + Z_- \cdot M$. But Z and M are local $\mathbb{P}$ -martingales. Therefore NZ is a local $\mathbb{P}$ -martingale by Theorem(8.26)(i). Now suppose that σ is a stopping time such that $(NZ)^\sigma$ is a $\mathbb{P}$ -martingale. Then for each bounded stopping time α , $\tilde{\mathbb{E}}(N_{\sigma \wedge \alpha}) = \mathbb{E}(N_{\sigma \wedge \alpha} Z_{\sigma \wedge \alpha})$, since $d\tilde{\mathbb{P}}|_{\mathcal{F}_{\sigma \wedge \alpha}} = Z_{\sigma \wedge \alpha} d\mathbb{P}|_{\mathcal{F}_{\sigma \wedge \alpha}}$ by the optional stopping theorem(8.17). I.e. $\tilde{\mathbb{E}}(N_\alpha^\sigma) = \mathbb{E}[(NZ)_\alpha^\sigma]$, and so $\tilde{\mathbb{E}}(N_\alpha^\sigma) = 0$. It now follows from Lemma(8.19) that N^σ is a $\tilde{\mathbb{P}}$ -martingale. Therefore N is a local $\tilde{\mathbb{P}}$ -martingale as required.

Now we have that $S = S_0 + N + B$, where

$$N = M + [M, X] \quad \text{and} \quad B = A - [M, X].$$

N is a local $\tilde{\mathbb{P}}$ -martingale, and B is a finite-variation process. Thus S is a $\tilde{\mathbb{P}}$ -semimartingale as required. By symmetry we also have that all $\tilde{\mathbb{P}}$ -semimartingales are $\mathbb{P}$ -semimartingales. $\square$

§9.4 THE PROOF OF THE CHARACTERIZATION THEOREM

The idea of the proof is to show that given an integrator S , we can change the probability to make S into a local quasimartingale. Then by the work of §9.2, S is a semimartingale with respect to this new probability. Finally Theorem(9.12) shows that S is a semimartingale with respect to the original probability. The method we shall use is based on that in Dellacherie & Meyer [DM] (see Theorem VIII.80). We shall require the following ‘minimax’ theorem.

(9.13) Theorem.

Suppose that K is a compact convex subset of a real topological vector space E , $\mathcal{A}$ is a non-empty convex family of continuous real-valued affine functions on K , and that $c \in \mathbb{R}$ is such that for each $f \in \mathcal{A}$, there exists an $x \in K$ such that $f(x) \leq c$. Then there exists an $x_0 \in K$ such that $f(x_0) \leq c$ for all $f \in \mathcal{A}$.

Note. By a real-valued affine function on K , we shall mean a function of the form $x \mapsto Tx + \alpha$, where $T: E \rightarrow \mathbb{R}$ is linear, and $\alpha \in \mathbb{R}$.

Proof: First we shall prove the following statement.

(†) If $f_1, \dots, f_n \in \mathcal{A}$, then there exists an $x \in K$ such that $f_k(x) \leq c$ for $k = 1, \dots, n$.

To prove this, define $\theta: K \rightarrow \mathbb{R}^n$ by $\theta(x) = (f_1(x), \dots, f_n(x))$, and let $M = (-\infty, c]^n$. We have to show that $\theta(K) \cap M \neq \emptyset$. Suppose, for a contradiction, that $\theta(K) \cap M = \emptyset$. Then since $\theta(K)$ is compact, M is closed, and both are convex, it follows that there is a linear map $\gamma: \mathbb{R}^n \rightarrow \mathbb{R}$ such that

$$(\ddagger) \quad \sup \gamma(M) < \inf \gamma \theta(K).$$

(See for example Theorem 3.4(b) of Rudin [Ru].) Now write $\gamma(t) = a_1 t_1 + \dots + a_n t_n$. Then since $\sup \gamma(M) < \infty$, we must have that $a_1, \dots, a_n \geq 0$. Let $a = a_1 + \dots + a_n$. Then $a > 0$, since $\gamma \neq 0$. By replacing γ by $\frac{1}{a}\gamma$, we may assume that $a_1 + \dots + a_n = 1$. Let $f = a_1 f_1 + \dots + a_n f_n$. Then $f \in \mathcal{A}$ (since $\mathcal{A}$ is convex). So there exists an $x \in K$ such that $f(x) \leq c$. But $f(x) = \gamma \theta(x) > \gamma(c, \dots, c) = c$, by (‡), since $(c, \dots, c) \in M$. This is a contradiction, and so (†) is proved.

Now for each $f \in \mathcal{A}$, let $K_f = \{x \in K \mid f(x) \leq c\}$. Then each K_f is a closed subset of K , and (†) says that the family $(K_f)_{f \in \mathcal{A}}$ has the finite intersection property. Thus $\bigcap_{f \in \mathcal{A}} K_f \neq \emptyset$, since K is compact. Let x_0 be an element of this intersection. □

(9.14)**Lemma.** Suppose that $(\Omega, \mathcal{F}, \mathbb{P})$ is any probability space, $\mathcal{X}$ is a non-empty bounded convex subset of $L^0(\mathbb{P})$, and also that $\mathcal{X} \subseteq L^1(\mathbb{P})$. Then there exists a probability $\tilde{\mathbb{P}}$ on $(\Omega, \mathcal{F})$ equivalent to $\mathbb{P}$, with bounded density with respect to $\mathbb{P}$, and such that $\sup_{X \in \mathcal{X}} \tilde{\mathbb{E}}(X) < \infty$. (Where again $\mathbb{E}$ and $\tilde{\mathbb{E}}$ denote expectations with respect to $\mathbb{P}$ and $\tilde{\mathbb{P}}$ respectively.)

Proof: First we shall prove the following statement.

(†) For each $\varepsilon > 0$, there exists a random variable V such that $0 \leq V \leq 1$, $\mathbb{E}(V) \geq 1 - \varepsilon$, and $\sup_{X \in \mathcal{X}} \mathbb{E}(XV) < \infty$.

To prove (†), let $K_\varepsilon = \{V \in L^\infty(\mathbb{P}) \mid 0 \leq V \leq 1, \mathbb{E}(V) \geq 1 - \varepsilon\}$. Then K_ε is a weak* compact convex subset of $L^\infty(\mathbb{P})$ ($= L^1(\mathbb{P})^*$). Now for each $c > 0$, and each $X \in \mathcal{X}$, $\mathbb{P}[|X| > c] = \mathbb{P}[|c^{-1}X| > 1] \leq \mathbb{E}(1 \wedge |c^{-1}X|) = \rho_0(c^{-1}X)$. But since $\mathcal{X}$ is bounded in $L^0(\mathbb{P})$, $\sup_{X \in \mathcal{X}} \rho_0(c^{-1}X) \rightarrow 0$ as $c \rightarrow \infty$. Now take $c > 0$ such that $\sup_{X \in \mathcal{X}} \rho_0(c^{-1}X) \leq \varepsilon$. Then $\mathbb{P}[|X| > c] \leq \varepsilon$ for all $X \in \mathcal{X}$. For $X \in \mathcal{X}$, let $V_X = \mathbf{1}_{[|X| \leq c]}$. Then $V_X \in K_\varepsilon$, and $\mathbb{E}(XV_X) \leq c$. Now apply Theorem(9.13) with $(U \mapsto \mathbb{E}(XU))_{X \in \mathcal{X}}$ as $\mathcal{A}$, $L^\infty(\mathbb{P})$ with the weak* topology as E , and K_ε as K , to give that there exist a $V \in K_\varepsilon$ such that $\mathbb{E}(XV) \leq c$ for all $X \in \mathcal{X}$. Thus (†) is proven.

Now using (†), take $V_0, V_1, \dots$ to be random variables such that for each n , $0 \leq V_n \leq 1$, $\mathbb{E}(V_n) \geq 1 - 2^{-n}$ and $\sup_{X \in \mathcal{X}} \mathbb{E}(XV_n) < \infty$. Let $c_n \in [0, \infty)$ be such that $\mathbb{E}(XV_n) \leq c_n$ for all $X \in \mathcal{X}$, and then take $a_0, a_1, \dots > 0$ such that $\sum_{n=0}^\infty a_n < \infty$ and $\sum_{n=0}^\infty a_n c_n < \infty$. Now let $H = \sum_{n=0}^\infty a_n V_n$. This sum converges uniformly; so H is bounded. We have that $H > 0$ a.s., since $\mathbb{P}[V_n = 0] \leq 2^{-n}$. For each $X \in \mathcal{X}$, $\mathbb{E}(XH) = \sum_{n=0}^\infty a_n \mathbb{E}(XV_n)$. So $\sup_{X \in \mathcal{X}} \mathbb{E}(XH) \leq \sum_{n=0}^\infty a_n c_n < \infty$. Therefore define $\tilde{\mathbb{P}}$ by $d\tilde{\mathbb{P}} = cH d\mathbb{P}$, where $c = \mathbb{E}(H)^{-1}$, to give the result. $\square$

(9.15)**Proposition.** Suppose that S is an integrator. Then there exists a probability $\tilde{\mathbb{P}}$ equivalent to $\mathbb{P}$, with bounded density with respect to $\mathbb{P}$, and such that S is an $L^1(\tilde{\mathbb{P}})$ -process with $\widetilde{\text{Var}}_b(S) < \infty$ for all $b \in \mathbb{R}^+$. So S is a local $\tilde{\mathbb{P}}$ -quasimartingale,

and hence also a local $L^1(\tilde{\mathbb{P}})$ -integrator. ($\widetilde{\text{Var}}_b$ denotes Var_b evaluated with respect to $\tilde{\mathbb{P}}$.)

Proof: Recall that $S_t^* = \sup_{0 \leq u \leq t} |S_u|$. Take $a_1, a_2, \dots > 0$ such that $\rho_0(a_n S_n^*) < 2^{-n}$. Then since $\sum_{n=1}^{\infty} \rho_0(a_n S_n^*) < \infty$, it follows that $\sum_{n=1}^{\infty} a_n S_n^* < \infty$ a.s. Let V denote this sum. By replacing $\mathbb{P}$ by the probability with density $c/(1+V)$ with respect to $\mathbb{P}$ ($c \in (0, \infty)$ chosen appropriately), we may assume that $V \in L^1(\mathbb{P})$. But $S_n^* \leq \frac{1}{a_n} V$. So $S_n^* \in L^1(\mathbb{P})$. Therefore $S_t \in L^1(\mathbb{P})$ for all $t \in \mathbb{R}^+$; i.e. S is an $L^1(\mathbb{P})$ -process.

Now since S is an integrator, there exist $\delta_1 \geq \delta_2 \geq \dots > 0$ such that

$$\Phi_S^0(\delta_n \mathbf{1}_{(n-1, n] \times \Omega}) < 2^{-n} \quad \text{for } n = 1, 2, \dots$$

Define a previsible process $W: \mathbb{R}^{++} \times \Omega \rightarrow \mathbb{R}^+$ by

$$W = \sum_{n=1}^{\infty} \delta_n \mathbf{1}_{(n-1, n] \times \Omega}.$$

I claim that W is Φ_S^0 -finite. For suppose that $\varepsilon > 0$. Then take $n \in \mathbb{N}_{>0}$ such that $2^{-n} < \frac{\varepsilon}{2}$, and then take $\lambda \in (0, 1]$ such that $\Phi_S^0(\lambda \delta_0 \mathbf{1}_{(0, n] \times \Omega}) < \frac{\varepsilon}{2}$. Then

$$\begin{aligned} \Phi_S^0(\lambda W) &\leq \Phi_S^0(\lambda \delta_0 \mathbf{1}_{(0, n] \times \Omega}) + \sum_{k=n+1}^{\infty} \Phi_S^0(\delta_k \mathbf{1}_{(k-1, k] \times \Omega}) \\ &< \frac{\varepsilon}{2} + \sum_{k=n+1}^{\infty} 2^{-k} < \varepsilon, \end{aligned}$$

as required.

Now let $\mathcal{H} = \{H \in \mathcal{E} \mid |H| \leq W\}$, and $\mathcal{X} = \{\int H dS \mid H \in \mathcal{H}\}$. Then $\mathcal{X}$ satisfies the hypothesis of Lemma(9.14) ($\mathcal{X}$ is bounded in $L^0(\mathbb{P})$, since $\sup_{X \in \mathcal{X}} \rho_0(\lambda X) \leq \Phi_S^0(\lambda W) \rightarrow 0$ as $\lambda \downarrow 0$). Thus by that lemma there exists a probability $\tilde{\mathbb{P}}$ equivalent

to $\mathbb{P}$, with bounded density with respect to $\mathbb{P}$, and such that $\sup_{X \in \mathcal{X}} \tilde{\mathbb{E}}(X) < \infty$. Let c denote this supremum.

Now suppose that $b \in \mathbb{R}^+$, and take $n \in \mathbb{N}_{>0}$ such that $n \geq b$. Suppose that $0 = t_0 < \dots < t_r = b$, and for $k = 0, \dots, r-1$, C_k is a simple $\mathcal{F}_{t_k}$ -measurable random variable such that $|C_k| \leq 1$. Define $H \in \mathcal{E}$ by

$$H_t(\omega) = \delta_n \sum_{k=0}^{r-1} C_k(\omega) \mathbf{1}_{(t_k, t_{k+1}]}(t).$$

Then $H \in \mathcal{H}$, and so $\int H dS \in \mathcal{X}$. Therefore $\mathbb{E}(\int H dS) \leq c$. I.e.

$$\tilde{\mathbb{E}} \left[\sum_{k=0}^{r-1} C_k(S_{t_{k+1}} - S_{t_k}) \right] \leq c\delta_n^{-1}.$$

Now $\tilde{\mathbb{E}}[C_k(S_{t_{k+1}} - S_{t_k})] = \tilde{\mathbb{E}}[C_k \tilde{\mathbb{E}}(S_{t_{k+1}} - S_{t_k} | \mathcal{F}_{t_k})]$. So taking C_k to be 1 on the set where the conditional expectation is positive and -1 elsewhere gives that

$$\tilde{\mathbb{E}} \left[\sum_{k=0}^{r-1} |\tilde{\mathbb{E}}(S_{t_{k+1}} - S_{t_k} | \mathcal{F}_{t_k})| \right] \leq c\delta_n^{-1}.$$

Now add $\tilde{\mathbb{E}}(|S_b|)$, and take the supremum over all partitions $0 = t_0 < \dots < t_r = b$, to give that $\widetilde{\text{Var}}_b(S) \leq c\delta_n^{-1} + \tilde{\mathbb{E}}(|S_b|) < \infty$ as required. $\square$

(9.16) Theorem (Bichteler, Dellacherie, Mokobodzki and Meyer).

A càdlàg process is an integrator if and only if it is a semimartingale.

Proof: We have already proved (Theorem(9.2)) that all semimartingales are integrators. So suppose that S is an integrator. Then by Proposition(9.15), there is a probability $\tilde{\mathbb{P}}$ equivalent to $\mathbb{P}$ such that S is a local $\tilde{\mathbb{P}}$ -quasimartingale. But by Theorem(9.9), we see that S is a $\tilde{\mathbb{P}}$ -semimartingale, and so by Theorem(9.12), S is a $\mathbb{P}$ -semimartingale. $\square$

It might be hoped that the class of local L^1 -integrators is the same as the class of integrators. This is **not** the case. It is for this reason that we have to change the probability in order to prove Theorem(9.16). From Theorem(9.9) and Theorem(9.16), the only difficulty is that a finite-variation process may not be a local integrable-variation process. This is where the conjecture breaks down. Here is a counter example.

(9.17)Example. Take $(\Omega, \mathcal{F}, \mathbb{P})$ to be a complete probability space with a non-negative random variable U on it such that $\mathbb{E}(U) = \infty$. Then define a standard filtration $(\mathcal{F}_t)_t$ on $(\Omega, \mathcal{F}, \mathbb{P})$ as follows. For $t < 1$, let $\mathcal{F}_t$ be the set of null sets and their compliments. For $t \geq 1$, let $\mathcal{F}_t = \mathcal{F}$. Define an increasing process A by $A_t = \mathbf{1}_{[1, \infty)}(t)U$. Now if σ is a stopping time, then since for $t < 1$, $[\sigma \leq t]$ is $\mathcal{F}_t$ -measurable, and hence has probability 0 or 1, it follows that σ is a.s. equal to a constant or $\sigma \geq 1$ a.s. Thus it is clear that it is impossible for there to exist stopping times $(\sigma_n)_n$ such that $\sigma_n \uparrow \infty$, and such that A^{σ_n} is an L^1 -process. Thus A can not be a local L^1 -integrator. But A is an increasing process and so is an integrator by Theorem(5.30). $\square$

A similar construction will give examples of martingales that are not local L^2 -martingales. It is examples like these that cause problems in the Strasbourg approach. The use of L^0 and upper gauges, as we have seen, totally avoids these problems.

REFERENCES

- [Bi1] : Bichteler, K., *Integration Theory. Lect. Notes in Math. 315*, Springer, Berlin, 1975.
- [Bi2] : Bichteler, K., *Stochastic Integration and L^p -theory of Semimartingales, Ann. Prob.*, **9**(1981)49—89.
- [DM] : Dellacherie, C. and Meyer, P. A., *Probabilités et Potentiel*, Chaps. I–IV, Hermann, Paris, 1975; Chaps. V–VIII, Hermann, Paris, 1980; Chaps. IX–XI, Hermann, Paris, 1983; Chaps. XII–XVI, Hermann, Paris, 1987.
- [Ed] : Edwards, D. A., *A Note on Stochastic Integrators, Math. Proc. Camb. Phil. Soc.*, To appear.
- [KW] : Kunita, H. and Watanabe, S., *On Square Integrable Martingales, Nagoya Math. J.*, **30**(1967)209—245.
- [MP] : Métivier, M. and Pellaumail, J., *Stochastic Integration*, Acad. Press, New York, 1979.
- [Me] : Meyer, P. A., *Un cours sur les intégrales stochastique, Séminaire de Probabilités X: Lect. Notes in Math. 511*, Springer, Berlin, 1976.
- [Pe] : Pellaumail, J., *Sur l'intégrale stochastique et la décomposition de Doob-Meyer, Astérisque*, **9**(1973).
- [RW] : Rogers, L. C. G. and Williams, D., *Diffusions, Markov Processes, and Martingales. Volume 2: Itô Calculus*, Wiley, Chichester, 1987.
- [Ro] : Rolewicz, S., *Metric Linear Spaces*, PWN—Polish Scientific Publications, Warsaw, 1972.
- [Ru] : Rudin, W., *Functional Analysis*, McGraw-Hill, New York, 1973.
- [Sa] : Sadosky, C., *Interpolation of Operators and Singular Integrals*, Marcel Dekker inc., New York, 1979.