Stochastic Calculus

Alan Bain

1. Introduction

The following notes aim to provide a very informal introduction to Stochastic Calculus, and especially to the Itô integral and some of its applications. They owe a great deal to Dan Crisan's *Stochastic Calculus and Applications* lectures of 1998; and also much to various books especially those of L. C. G. Rogers and D. Williams, and Dellacherie and Meyer's multi volume series 'Probabilities et Potentiel'. They have also benefited from insights gained by attending lectures given by T. Kurtz.

The present notes grew out of a set of typed notes which I produced when revising for the Cambridge, Part III course; combining the printed notes and my own handwritten notes into a consistent text. I've subsequently expanded them inserting some extra proofs from a great variety of sources. The notes principally concentrate on the parts of the course which I found hard; thus there is often little or no comment on more standard matters; as a secondary goal they aim to present the results in a form which can be readily extended Due to their evolution, they have taken a very informal style; in some ways I hope this may make them easier to read.

The addition of coverage of discontinuous processes was motivated by my interest in the subject, and much insight gained from reading the excellent book of J. Jacod and A. N. Shiryaev.

The goal of the notes in their current form is to present a fairly clear approach to the Itô integral with respect to *continuous* semimartingales but without any attempt at maximal detail. The various alternative approaches to this subject which can be found in books tend to divide into those presenting the integral directed entirely at Brownian Motion, and those who wish to prove results in complete generality for a semimartingale. Here at all points clarity has hopefully been the main goal here, rather than completeness; although secretly the approach aims to be readily extended to the discontinuous theory. I make no apology for proofs which spell out every minute detail, since on a first look at the subject the purpose of some of the steps in a proof often seems elusive. I'd especially like to convince the reader that the Itô integral isn't that much harder in concept than the Lebesgue Integral with which we are all familiar. The motivating principle is to try and explain every detail, no matter how trivial it may seem once the subject has been understood!

Passages enclosed in boxes are intended to be viewed as digressions from the main text; usually describing an alternative approach, or giving an informal description of what is going on – feel free to skip these sections if you find them unhelpful.

In revising these notes I have resisted the temptation to alter the original structure of the development of the Itô integral (although I have corrected unintentional mistakes), since I suspect the more concise proofs which I would favour today would not be helpful on a first approach to the subject.

These notes contain errors with probability one. I always welcome people telling me about the errors because then I can fix them! I can be readily contacted by email as `alanb@chiark.greenend.org.uk`. Also suggestions for improvements or other additions are welcome.

Alan Bain

2. Contents

3. Stochastic Processes

The following notes are a summary of important definitions and results from the theory of stochastic processes, proofs may be found in the usual books for example [Durrett, 1996].

3.1. Probability Space

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. The set of $\mathbb{P}$-null subsets of Ω is defined by

$$\mathcal{N} := \{N \subset \Omega : N \subset A \text{ for } A \in \mathcal{F}, \text{ with } \mathbb{P}(A) = 0\}.$$

The space $(\Omega, \mathcal{F}, \mathbb{P})$ is said to be *complete* if for $A \subset B \subset \Omega$ with $B \in \mathcal{F}$ and $\mathbb{P}(B) = 0$ then this implies that $A \in \mathcal{F}$.

In addition to the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let $(E, \mathcal{E})$ be a measurable space, called the state space, which in many of the cases considered here will be $(\mathbb{R}, \mathcal{B})$, or $(\mathbb{R}^n, \mathcal{B})$. A *random variable* is a $\mathcal{F}/\mathcal{E}$ measurable function $X : \Omega \to E$.

3.2. Stochastic Process

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a measurable state space $(E, \mathcal{E})$, a stochastic process is a family $(X_t)_{t\geq 0}$ such that X_t is an E valued random variable for each time $t \geq 0$. More formally, a map $X : (\mathbb{R}^+ \times \Omega, \mathcal{B}^+ \otimes \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$, where $\mathcal{B}^+$ are the Borel sets of the time space $\mathbb{R}^+$.

Definition 1. **Measurable Process**
The process $(X_t)_{t\geq 0}$ is said to be measurable if the mapping $(\mathbb{R}^+ \times \Omega, \mathcal{B}^+ \otimes \mathcal{F}) \to (\mathbb{R}, \mathcal{B}) :$ $(t, \omega) \mapsto X_t(\omega)$ is measurable on $\mathbb{R} \times \Omega$ with respect to the product σ-field $\mathcal{B}(\mathbb{R}) \otimes \mathcal{F}$.

Associated with a process is a *filtration*, an increasing chain of σ-algebras i.e.

$$\mathcal{F}_s \subset \mathcal{F}_t \text{ if } 0 \leq s \leq t < \infty.$$

Define $\mathcal{F}_\infty$ by

$$\mathcal{F}_\infty = \bigvee_{t\geq 0} \mathcal{F}_t := \sigma\left(\bigcup_{t\geq 0} \mathcal{F}_t\right).$$

If $(X_t)_{t\geq 0}$ is a stochastic process, then the *natural filtration* of $(X_t)_{t\geq 0}$ is given by

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

The process $(X_t)_{t\geq 0}$ is said to be $(\mathcal{F}_t)_{t\geq 0}$ *adapted*, if X_t is $\mathcal{F}_t$ measurable for each $t \geq 0$. The process $(X_t)_{t\geq 0}$ is obviously adapted with respect to the natural filtration.

Definition 2. **Progressively Measurable Process**
A process is progressively measurable if for each t its restriction to the time interval $[0, t]$, is measurable with respect to $\mathcal{B}_{[0,t]} \otimes \mathcal{F}_t$, where $\mathcal{B}_{[0,t]}$ is the Borel σ algebra of subsets of $[0, t]$.

Why on earth is this useful? Consider a non-continuous stochastic process X_t. From the definition of a stochastic process for each t that $X_t \in \mathcal{F}_t$. Now define $Y_t = \sup_{s\in[0,t]} X_s$. Is Y_s a stochastic process? The answer is not necessarily – sigma fields are only guaranteed closed under countable unions, and an event such as

$$\{Y_s > 1\} = \bigcup_{0\leq s\leq s} \{X_s > 1\}$$

is an uncountable union. If X were progressively measurable then this would be sufficient to imply that Y_s is $\mathcal{F}_s$ measurable. If X has suitable continuity properties, we can restrict the unions which cause problems to be over some dense subset (say the rationals) and this solves the problem. Hence the next theorem.

Theorem 3.3.
Every adapted right (or left) continuous, adapted process is progressively measurable.

Proof
We consider the process X restricted to the time interval $[0, s]$. On this interval for each $n \in \mathbb{N}$ we define

$$X_1^n := \sum_{k=0}^{2^n-1} 1_{(ks/2^n,(k+1)s/2^n]}(t)X_{ks/2^n}(\omega),$$

$$X_2^n := 1_{[0,s/2^n)}(t)X_0(\omega) + \sum_{k=1}^{2^n} 1_{[ks/2^n,(k+1)s/2^n)}(t)X_{(k+1)s/2^n}(\omega)$$

Note that X_1^n is a left continuous process, so if X is left continuous, working pointwise (that is, fix ω), the sequence X_1^n converges to X.

But the individual summands in the definition of X_1^n are by the adpatedness of X clearly $\mathcal{B}_{[0,s]} \otimes \mathcal{F}_s$ measurable, hence X_1^n is also. But the convergence implies X is also; hence X is progressively measurable.

Consideration of the sequence X_2^n yields the same result for right continuous, adapted processes. □

The following extra information about filtrations should probably be skipped on a first reading, since they are likely to appear as excess baggage.

Define

$$\forall t \in (0,\infty) \quad \mathcal{F}_{t-} = \bigvee_{0\le s<t} \mathcal{F}_s$$
$$\forall t \in [0,\infty) \quad \mathcal{F}_{t+} = \bigwedge_{t\le s<\infty} \mathcal{F}_s,$$

whence it is clear that for each t, $\mathcal{F}_{t-} \subset \mathcal{F}_t \subset \mathcal{F}_{t+}$.

Definition 3.2.
The family $\{\mathcal{F}_t\}$ is called right continuous if

$$\forall t \in [0,\infty) \quad \mathcal{F}_t = \mathcal{F}_{t+}.$$

Definition 3.3.
A process $(X_t)_{t\ge 0}$ is said to be bounded if there exists a universal constant K such that for all ω and $t \ge 0$, then $|X_t(\omega)| < K$.

Definition 3.4.
Let $X = (X_t)_{t\ge 0}$ be a stochastic process defined on $(\Omega,\mathcal{F},\mathbb{P})$, and let $X' = (X'_t)_{t\ge 0}$ be a stochastic process defined on $(\Omega,\mathcal{F},\mathbb{P})$. Then X and X' have the same finite dimensional distributions if for all n, $0 \le t_1 < t_2 < \cdots < t_n < \infty$, and $A_1, A_2, \ldots, A_n \in \mathcal{E}$,

$$\mathbb{P}(X_{t_1} \in A_1, X_{t_2} \in A_2, \ldots, X_{t_n} \in A_n) = \mathbb{P}'(X'_{t_1} \in A_1, X'_{t_2} \in A_2, \ldots, X'_{t_n} \in A_n).$$

Definition 3.5.
Let X and X' be defined on $(\Omega,\mathcal{F},\mathbb{P})$. Then X and X' are modifications of each other if and only if

$$\mathbb{P}\left(\{\omega\in\Omega : X_t(\omega) = X'_t(\omega)\}\right) = 1 \qquad \forall t \ge 0.$$

Definition 3.6.
Let X and X' be defined on $(\Omega,\mathcal{F},\mathbb{P})$. Then X and X' are indistinguishable if and only if

$$\mathbb{P}\left(\{\omega\in\Omega : X_t(\omega) = X'_t(\omega) \forall t \ge 0\}\right) = 1.$$

There is a chain of implications

$$\text{indistinguishable} \Rightarrow \text{modifications} \Rightarrow \text{same f.d.d.}$$

The following definition provides us with a special name for a process which is indistinguishable from the zero process. It will turn out to be important because many definitions can only be made up to evanescence.

Definition 3.7.
A process X is evanescent if $\mathbb{P}(X_t = 0\ \forall t) = 1$.

4. Martingales

Definition 4.1.
Let $X = \{X_t, \mathcal{F}_t, t \geq 0\}$ be an integrable process then X is a

(i) **Martingale** *if and only if $\mathbb{E}(X_t|\mathcal{F}_s) = X_s$ a.s. for $0 \leq s \leq t < \infty$*
(ii) **Supermartingale** *if and only if $\mathbb{E}(X_t|\mathcal{F}_s) \leq X_s$ a.s. for $0 \leq s \leq t < \infty$*
(iii) **Submartingale** *if and only if $\mathbb{E}(X_t|\mathcal{F}_s) \geq X_s$ a.s. for $0 \leq s \leq t < \infty$*

Theorem (Kolmogorov) 4.2.
Let $X = \{X_t, \mathcal{F}_t, t \geq 0\}$ be an integrable process. Then define $\mathcal{F}_{t+} := \bigwedge_{\epsilon>0} \mathcal{F}_{t+\epsilon}$ and also the partial augmentation of $\mathcal{F}$ by $\tilde{\mathcal{F}}_t = \sigma(\mathcal{F}_{t+}, \mathcal{N})$. Then if $t \mapsto \mathbb{E}(X_t)$ is continuous there exists an $\tilde{\mathcal{F}}_t$ adapted stochastic process $\tilde{X} = \{\tilde{X}_t, \tilde{\mathcal{F}}_t, t \geq 0\}$ with sample paths which are right continuous, with left limits (CADLAG) such that X and $\tilde{X}$ are modifications of each other.

Definition 4.3.
A martingale $X = \{X_t, \mathcal{F}_t, t \geq 0\}$ is said to be an L^2-martingale or a square integrable martingale if $\mathbb{E}(X_t^2) < \infty$ for every $t \geq 0$.

Definition 4.4.
A process $X = \{X_t, \mathcal{F}_t, t \geq 0\}$ is said to be L^p bounded if and only if $\sup_{t \geq 0} \mathbb{E}(|X_t|^p) < \infty$. The space of L^2 bounded martingales is denoted by $\mathcal{M}_2$, and the subspace of continuous L^2 bounded martingales is denoted $\mathcal{M}_2^c$.

Definition 4.5.
A process $X = \{X_t, \mathcal{F}_t, t \geq 0\}$ is said to be uniformly integrable if and only if

$$\sup_{t \geq 0} \mathbb{E}\left(|X_t| 1_{|X_t| \geq N}\right) \to 0 \text{ as } N \to \infty.$$

Orthogonality of Martingale Increments
A frequently used property of a martingale M is the *orthogonality of increments property* which states that for a square integrable martingale M, and $Y \in \mathcal{F}_s$ with $\mathbb{E}(Y^2) < \infty$ then

$$\mathbb{E}\left[Y(M_t - M_s)\right] = 0 \qquad \text{for } t \geq s.$$

Proof
Via Cauchy Schwartz inequality $\mathbb{E}|Y(M_t - M_s)| < \infty$, and so

$$\mathbb{E}(Y(M_t - M_s)) = \mathbb{E}\left(\mathbb{E}(Y(M_t - M_s)|\mathcal{F}_s)\right) = \mathbb{E}\left(Y\mathbb{E}(M_t - M_s|\mathcal{F}_s)\right) = 0.$$

□

A typical example is $Y = M_s$, whence $\mathbb{E}(M_s(M_t - M_s)) = 0$ is obtained. A common application is to the difference of two squares, let $t \geq s$ then

$$\begin{aligned}\mathbb{E}((M_t - M_s)^2|\mathcal{F}_s) =& \mathbb{E}(M_t^2|\mathcal{F}_s) - 2M_s\mathbb{E}(M_t|\mathcal{F}_s) + M_s^2 \\ =& \mathbb{E}(M_t^2 - M_s^2|\mathcal{F}_s) = \mathbb{E}(M_t^2|\mathcal{F}_s) - M_s^2.\end{aligned}$$

4.1. Stopping Times

A random variable $T : \Omega \to [0, \infty)$ is a stopping (optional) time if and only if $\{\omega : T(\omega) \leq t\} \in \mathcal{F}_t$.

The following theorem is included as a demonstration of checking for stopping times, and may be skipped if desired.

Theorem 4.6.
T is a stopping time with respect to $\mathcal{F}_{t+}$ if and only if for all $t \in [0, \infty)$, the event $\{T < t\}$ if $\mathcal{F}_t$ measurable.

Proof
If T is an $\mathcal{F}_{t+}$ stopping time then for all $t \in (0, \infty)$ the event $\{T \leq t\}$ is $\mathcal{F}_{t+}$ measurable. Thus for $1/n < t$ we have

$$\left\{T \leq t - \frac{1}{n}\right\} \in \mathcal{F}_{(t-1/n)^+} \subset \mathcal{F}_t$$

so

$$\{T < t\} = \bigcup_{n=1}^{\infty} \left\{T \leq t - \frac{1}{n}\right\} \in \mathcal{F}_t.$$

To prove the converse, note that if for each $t \in [0, \infty)$ we have that $\{T < t\} \in \mathcal{F}_t$, then for each such t

$$\left\{T < t + \frac{1}{n}\right\} \in \mathcal{F}_{t+1/n},$$

as a consequence of which

$$\{T \leq t\} = \bigcap_{n=1}^{\infty} \left\{T < t + \frac{1}{n}\right\} \in \bigwedge_{n=1}^{\infty} \mathcal{F}_{t+1/n} = \mathcal{F}_{t+}.$$

□

Given a stochastic process $X = (X_t)_{t \geq 0}$, a stopped process X^T may be defined by

$$\begin{aligned} X^T(\omega) :=& X_{T(\omega)\wedge t}(\omega), \\ \mathcal{F}_T :=& \{A \in \mathcal{F} : A \cap \{T \leq t\} \in \mathcal{F}_t\}. \end{aligned}$$

Theorem (Optional Stopping).
Let X be a right continuous integrable, $\mathcal{F}_t$ adapted process. Then the following are equivalent:
(i) X is a martingale.
(ii) X^T is a martingale for all stopping times T.
(iii) $\mathbb{E}(X_T) = \mathbb{E}(X_0)$ for all bounded stopping times T.
(iv) $\mathbb{E}(X_T|\mathcal{F}_S) = X_S$ for all bounded stopping times S and T such that $S \leq T$. If in addition, X is uniformly integrable then (iv) holds for all stopping times (not necessarily bounded).

The condition which is most often forgotten is that in (iii) that the stopping time T be bounded. To see why it is necessary consider B_t a Brownian Motion starting from zero. Let $T = \inf\{t \geq 0 : X_t = 1\}$, clearly a stopping time. Equally B_t is a martingale with respect to the filtration generated by B itself, but it is also clear that $\mathbb{E}B_T = 1 \neq \mathbb{E}B_0 = 0$. Obviously in this case $T < \infty$ is false.

Theorem (Doob's Martingale Inequalities).
Let $M = \{M_t, \mathcal{F}_t, t \geq 0\}$ be a uniformly integrable martingale, and let $M^* := \sup_{t \geq 0} |M_t|$. Then

(i) Maximal Inequality. For $\lambda > 0$,

$$\lambda \mathbb{P}(M^* \geq \lambda) \leq \mathbb{E}\left[|M_\infty| 1_{M^* < \infty}\right].$$

(ii) L^p maximal inequality. For $1 < p < \infty$,

$$\|M^*\|_p \leq \frac{p}{p-1} \|M_\infty\|_p.$$

Note that the norm used in stating the Doob L^p inequality is defined by

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

Theorem (Martingale Convergence).
Let $M = \{M_t, \mathcal{F}_t, t \geq 0\}$ be a martingale.

(i) If M is L^p bounded then $M_\infty(\omega) := \lim_{t \to \infty} M_t(\omega)$ $\mathbb{P}$-a.s.

(ii) If $p = 1$ and M is uniformly integrable then $\lim_{t\to\infty} M_t(\omega) = M_\infty(\omega)$ in L^1. Then for all $A \in L^1(\mathcal{F}_\infty)$, there exists a martingale A_t such that $\lim_{t\to\infty} A_t = A$, and $A_t = \mathbb{E}(A|\mathcal{F}_t)$. Here $\mathcal{F}_\infty := \lim_{t\to\infty} \mathcal{F}_t$.

(iii) If $p > 1$ i.e. M is L^p bounded $\lim_{t\to\infty} M_t = M_\infty$ in L^p.

Definition 4.7.
Let $\mathcal{M}_2$ denote the set of L^2-bounded CADLAG martingales i.e. martingales M such that

$$\sup_{t \geq 0} \mathbb{E} M_t^2 < \infty.$$

Let $\mathcal{M}_2^c$ denote the set of L^2-bounded CADLAG martingales which are continuous. A norm may be defined on the space $\mathcal{M}_2$ by $\|M\|^2 = \|M_\infty\|_2^2 = \mathbb{E}(M_\infty^2)$.

From the conditional Jensen's inequality, since $f(x) = x^2$ is convex,

$$\begin{aligned} \mathbb{E}\left(M_\infty^2 | \mathcal{F}_t\right) &\geq \left(\mathbb{E}(M_\infty | \mathcal{F}_t)\right)^2 \\ \mathbb{E}\left(M_\infty^2 | \mathcal{F}_t\right) &\geq (\mathbb{E}M_t)^2. \end{aligned}$$

Hence taking expectations

$$\mathbb{E}M_t^2 \leq \mathbb{E}M_\infty^2,$$

and since by martingale convergence in L^2, we get $\mathbb{E}(M_t^2) \to \mathbb{E}(M_\infty^2)$, it is clear that

$$\mathbb{E}(M_\infty^2) = \sup_{t \geq 0} \mathbb{E}(M_t^2).$$

Theorem 4.8.
The space $(\mathcal{M}_2, \|\cdot\|)$ (up to equivalence classes defined by modifications) is a Hilbert space, with $\mathcal{M}_2^c$ a closed subspace.

Proof
We prove this by showing a one to one correspondence between $\mathcal{M}_2$ (the space of square integrable martingales) and $L^2(\mathcal{F}_\infty)$. The bijection is obtained via

$$\begin{aligned} f :& \mathcal{M}_2 \to L^2(\mathcal{F}_\infty) \\ f :& (M_t)_{t\geq 0} \mapsto M_\infty \equiv \lim_{t\to\infty} M_t \\ g :& L^2(\mathcal{F}_\infty) \to \mathcal{M}_2 \\ g :& M_\infty \mapsto M_t \equiv \mathbb{E}(M_\infty|\mathcal{F}_t) \end{aligned}$$

Notice that

$$\sup_t \mathbb{E}M_t^2 = \|M_\infty\|_2^2 = \mathbb{E}(M_\infty^2) < \infty,$$

as M_t is a square integrable martingale. As $L^2(\mathcal{F}_\infty)$ is a Hilbert space, $\mathcal{M}_2$ inherits this structure.

To see that $\mathcal{M}_2^c$ is a closed subspace of $\mathcal{M}_2$, consider a Cauchy sequence $\{M^{(n)}\}$ in $\mathcal{M}_2$, equivalently $\{M_\infty^{(n)}\}$ is Cauchy in $L^2(\mathcal{F}_\infty)$. Hence $M_\infty^{(n)}$ converges to a limit, M_∞ say, in $L^2(\mathcal{F}_\infty)$. Let $M_t := \mathbb{E}(M_\infty|\mathcal{F}_t)$, then

$$\sup_{t\geq 0} \left| M_t^{(n)} - M_t \right| \to 0, \text{ in } L^2,$$

that is $M^{(n)} \to M$ uniformly in L^2. Hence there exists a subsequence $n(k)$ such that $M^{n(k)} \to M$ uniformly; as a uniform limit of continuous functions is continuous, $M \in \mathcal{M}_2^c$. Thus $\mathcal{M}_2^c$ is a closed subspace of $\mathcal{M}$.

5. Basics

5.1. Local Martingales

A martingale has already been defined, but a weaker definition will prove useful for stochastic calculus. Note that I'll often drop references to the filtration $\mathcal{F}_t$, but this nevertheless forms an essential part of the (local) martingale.

Just before we dive in and define a Local Martingale, maybe we should pause and consider the reason for considering them. The important property of local martingales will only be seen later in the notes; and as we frequently see in this subject it is one of *stability* that is, they are a class of objects which are closed under an operation, in this case under the stochastic integral – an integral of a previsible process with a local martingale integrator is a local martingale.

Definition 5.1.
$M = \{M_t, \mathcal{F}_t, 0 \le t \le \infty\}$ is a local martingale if and only if there exists a sequence of stopping times T_n tending to infinity such that M^{T_n} are martingales for all n. The space of local martingales is denotes $\mathcal{M}_{loc}$, and the subspace of continuous local martingales is denotes $\mathcal{M}_{loc}^c$.

Recall that a martingale $(X_t)_{t\ge 0}$ is said to be bounded if there exists a universal constant K such that for all ω and $t \ge 0$, then $|X_t(\omega)| < K$.

Theorem 5.2.
Every bounded local martingale is a martingale.

Proof
Let T_n be a sequence of stopping times as in the definition of a local martingale. This sequence tends to infinity, so pointwise $X_t^{T_n}(\omega) \to X_t(\omega)$. Using the conditional form of the dominated convergence theorem (using the constant bound as the dominating function), for $t \ge s \ge 0$

$$\lim_{n\to\infty} \mathbb{E}(X_t^{T_n}|\mathcal{F}_s) = \mathbb{E}(X_t|\mathcal{F}_s).$$

But as X^{T_n} is a (genuine) martingale, $\mathbb{E}(X_t^{T_n}|\mathcal{F}_s) = X_s^{T_n} = X_{T_n\wedge s}$; so

$$\mathbb{E}(X_t|\mathcal{F}_s) = \lim_{n\to\infty} \mathbb{E}(X_t^{T_n}|\mathcal{F}_s) = \lim_{n\to\infty} X_s^{T_n} = X_s.$$

Hence X_t is a genuine martingale. □

Proposition 5.3.
The following are equivalent
(i) $M = \{M_t, \mathcal{F}_t, 0 \le t \le \infty\}$ is a continuous martingale.
(ii) $M = \{M_t, \mathcal{F}_t, 0 \le t \le \infty\}$ is a continuous local martingale and for all $t \ge 0$, the set $\{M_T : T \text{ a stopping time}, \ T \le t\}$ is uniformly integrable.

Proof
(i) $\Rightarrow$ (ii) By optional stopping theorem, if $T \le t$ then $M_T = \mathbb{E}(M_t|\mathcal{F}_T)$ hence the set is uniformly integrable.

(ii) $\Rightarrow$ (i)It is required to prove that $\mathbb{E}(M_0) = \mathbb{E}(M_T)$ for any bounded stopping time T. Then by local martingale property for any n,

$$\mathbb{E}(M_0) = \mathbb{E}(M_{T\wedge T_n}),$$

uniform integrability then implies that

$$\lim_{n\to\infty} \mathbb{E}(M_{T\wedge T_n}) = \mathbb{E}(M_T).$$

5.2. Local Martingales which are not Martingales

There do exist local martingales which are not themselves martingales. The following is an example Let B_t be a d dimensional Brownian Motion starting from x. It can be shown using Itô's formula that a harmonic function of a Brownian motion is a local martingale (this is on the example sheet). From standard PDE theory it is known that for $d \geq 3$, the function

$$f(x) = \frac{1}{|x|^{d-2}}$$

is a harmonic function, hence $X_t = 1/|B_t|^{d-2}$ is a local martingale. Now consider the L^p norm of this local martingale

$$\mathbb{E}_x|X_t|^p = \int \frac{1}{(2\pi t)^{d/2}} \exp\left(-\frac{|y-x|^2}{2t}\right) |y|^{-(d-2)p} \mathrm{d}y.$$

Consider when this integral converges. There are no divergence problems for $|y|$ large, the potential problem lies in the vicinity of the origin. Here the term

$$\frac{1}{(2\pi t)^{d/2}} \exp\left(-\frac{|y-x|^2}{2t}\right)$$

is bounded, so we only need to consider the remainder of the integrand integrated over a ball of unit radius about the origin which is bounded by

$$C\int_{B(0,1)} |y|^{-(d-2)p} \mathrm{d}y,$$

for some constant C, which on tranformation into polar co-ordinates yields a bound of the form

$$C' \int_0^1 r^{-(d-2)p} r^{d-1} \mathrm{d}r,$$

with C' another constant. This is finite if and only if $-(d-2)p + (d-1) > -1$ (standard integrals of the form $1/r^k$). This in turn requires that $p < d/(d-2)$. So clealry $\mathbb{E}_x|X_t|$ will be finite for all $d \geq 3$.

Now although $\mathbb{E}_x|X_t| < \infty$ and X_t is a local martingale, we shall show that it is not a martingale. Note that $(B_t - x)$ has the same distribution as $\sqrt{t}(B_1 - x)$ under $\mathbb{P}_x$ (the probability measure induced by the BM starting from x). So as $t \to \infty$, $|B_t| \to \infty$ in probability and $X_t \to 0$ in probability. As $X_t \geq 0$, we see that $\mathbb{E}_x(X_t) = \mathbb{E}_x|X_t| < \infty$. Now note that for any $R < \infty$, we can construct a bound

$$\mathbb{E}_x X_t \leq \frac{1}{(2\pi t)^{d/2}} \int_{|y| \leq R} |y|^{-(d-2)} \mathrm{d}y + R^{-(d-2)},$$

which converges, and hence

$$\limsup_{t \to \infty} \mathbb{E}_x X_t \leq R^{-(d-2)}.$$

As R was chosen arbitrarily we see that $\mathbb{E}_x X_t \to 0$. But $\mathbb{E}_x X_0 = |x|^{-(d-2)} > 0$, which implies that $\mathbb{E}_x X_t$ is not constant, and hence X_t is not a martingale.

6. Total Variation and the Stieltjes Integral

Let $A : [0, \infty) \to \mathbb{R}$ be a CADLAG (continuous to right, with left limits) process. Let a *partition* $\Pi = \{t_0, t_1, \ldots, t_m\}$ have $0 = t_0 \le t_1 \le \cdots \le t_m = t$; the mesh of the partition is defined by

$$\delta(\Pi) = \max_{1 \le k \le m} |t_k - t_{k-1}|.$$

The variation of A is then defined as the increasing process V given by,

$$V_t := \sup_{\Pi} \left\{ \sum_{k=1}^{n(\Pi)} \left| A_{t_k \wedge t} - A_{t_{k-1} \wedge t} \right| : 0 = t_0 \le t_1 \le \cdots \le t_n = t \right\}.$$

An alternative definition is given by

$$V_t^0 := \lim_{n \to \infty} \sum_{1}^{n(\Pi)} \left| A_{k2^{-n} \wedge t} - A_{(k-1)2^{-n} \wedge t} \right|.$$

These can be shown to be equivalent (for CADLAG processes), since trivially (use the dyadic partition), $V_t^0 \le V_t$. It is also possible to show that $V_t^0 \ge V_t$ for the total variation of a CADLAG process.

Definition 6.1.
A process A is said to have finite variation if the associated variation process V is finite (i.e. if for every t and every ω, $|V_t(\omega)| < \infty$.

6.1. Why we need a Stochastic Integral

Before delving into the depths of the integral it's worth stepping back for a moment to see why the 'ordinary' integral cannot be used on a path at a time basis (i.e. separately for each $\omega \in \Omega$). Suppose we were to do this i.e. set

$$I_t(X) = \int_0^t X_s(\omega) \mathrm{d}M_s(\omega),$$

for $M \in \mathcal{M}_2^c$; but for an interesting martingale (i.e. one which isn't zero a.s.), the total variation is not finite, even on a bounded interval like $[0, T]$. Thus the Lebesgue-Stieltjes integral definition isn't valid in this case. To generalise we shall see that the quadratic variation is actually the 'right' variation to use (higher variations turn out to be zero and lower ones infinite, which is easy to prove by considering the variation expressed as the limit of a sum and factoring it by a maximum multiplies by the quadratic variation, the first term of which tends to zero by continuity). But to start, we shall consider integrating a previsible process H_t with an integrator which is an increasing finite variation process. First we shall prove that a continuous local martingale of finite variation is zero.

Proposition 6.2.
If M is a continuous local martingale of finite variation, starting from zero then M is identically zero.

Proof
Let V be the variation process of M. This V is a continuous, adapted process. Now define a sequence of stopping times S_n as the first time V exceeds n, i.e. $S_n := \inf_t\{t \geq 0 : V_t \geq n\}$. Then the martingale M^{S_n} is of bounded variation. It therefore suffices to prove the result for a bounded, continuous martingale M of bounded variation.

Fix $t \geq 0$ and let $\{0 = t_0, t_1, \ldots, t_N = t\}$ be a partition of $[0, t]$. Then since $M_0 = 0$ it is clear that, $M_t^2 = \sum_{k=1}^N \left(M_{t_k}^2 - M_{t_{k-1}}^2\right)$. Then via orthogonality of martingale increments

$$\begin{aligned}\mathbb{E}(M_t^2) =& \mathbb{E}\left(\sum_{k=1}^N \left(M_{t_k} - M_{t_{k-1}}\right)^2\right)\\ \leq& \mathbb{E}\left(V_t \sup_k \left|M_{t_k} - M_{t_{k-1}}\right|\right)\end{aligned}$$

The integrand is bounded by n^2 (from definition of the stopping time S_n), hence the expectation converges to zero as the modulus of the partition tends to zero by the bounded convergence theorem. Hence $M \equiv 0$. □

6.2. Previsibility

The term previsible has crept into the discussion earlier. Now is the time for a proper definition.

Definition 6.3.
The previsible (or predictable) σ-field $\mathcal{P}$ is the σ-field on $\mathbb{R}^+ \times \Omega$ generated by the processes $(X_t)_{t\geq 0}$, adapted to $\mathcal{F}_t$, with left continuous paths on $(0, \infty)$.

Remark
The same σ-field is generated by left continuous, right limits processes (i.e. càglàd processes) which are adapted to $\mathcal{F}_{t-}$, or indeed continuous processes $(X_t)_{t\geq 0}$ which are adapted to $\mathcal{F}_{t-}$. It is gnerated by sets of the form $A \times (s, t]$ where $A \in \mathcal{F}_s$. It should be noted that càdlàg processes generate the optional σ field which is usually different.

Theorem 6.4.
The previsible σ fieldis also generated by the collection of random sets $A \times \{0\}$ where $A \in \mathcal{F}_0$ and $A \times (s, t]$ where $A \in \mathcal{F}_s$.

Proof
Let the σ field generated by the above collection of sets be denotes $\mathcal{P}'$. We shall show $\mathcal{P} = \mathcal{P}'$. Let X be a left continuous process, define for $n \in N$

$$X^n = X_0 1_0(t) + \sum_k X_{k/2^n} 1_{(k/2^n,(k+1)/2^n]}(t)$$

It is clear that $X_n \in \mathcal{P}'$. As X is left continuous, the above sequence of left-continuous processes converges pointwise to X, so X is $\mathcal{P}'$ measurable, thus $\mathcal{P} \subset \mathcal{P}'$. Conversely consider the indicator function of $A \times (s,t]$ this can be written as $1_{[0,t_A]\setminus[0,s_A]}$, where $s_A(\omega) = s$ for $\omega \in A$ and $+\infty$ otherwise. These indicator functions are adapated and left continuous, hence $\mathcal{P}' \subset \mathcal{P}$. □

Definition 6.5.
A process $(X_t)_{t\geq 0}$ is said to be previsible, if the mapping $(t,\omega) \mapsto X_t(\omega)$ is measurable with respect to the previsible σ-field $\mathcal{P}$.

6.3. Lebesgue-Stieltjes Integral

[In the lecture notes for this course, the Lebesgue-Stieltjes integral is considered first for functions A and H; here I consider processes on a pathwise basis.]

Let A be an increasing cadlag process. This induces a Borel measure $\mathrm{d}A$ on $(0,\infty)$ such that

$$\mathrm{d}A((s,t])(\omega) = A_t(\omega) - A_s(\omega).$$

Let H be a previsible process (as defined above). The Lebesgue-Stieltjes integral of H is defined with respect to an increasing process A by

$$(H \cdot A)_t(\omega) = \int_0^t H_s(\omega)\mathrm{d}A_s(\omega),$$

whenever $H \geq 0$ or $(|H| \cdot A)_t < \infty$.

As a notational aside, we shall write

$$(H \cdot A)_t \equiv \int_0^t H\mathrm{d}X,$$

and later on we shall use

$$\mathrm{d}(H \cdot X) \equiv H\mathrm{d}X.$$

This definition may be extended to integrator of finite variation which are not increasing, by decomposing the process A of finite variation into a difference of two increasing processes, so $A = A^+ - A^-$, where $A^\pm = (V \pm A)/2$ (here V is the total variation process for A). The integral of H with respect to the finite variation process A is then defined by

$$(H \cdot A)_t(\omega) := (H \cdot A^+)_t(\omega) - (H \cdot A^-)_t(\omega),$$

whenever $(|H| \cdot V)_t < \infty$.

There are no really new concepts of the integral in the foregoing; it is basically the Lebesgue-Stieltjes integral eextended from functions $H(t)$ to processes in a pathwise fashion (that's why ω has been included in those definitions as a reminder).

Theorem 6.6.
If X is a non-negative continuous local martingale and $\mathbb{E}(X_0) < \infty$ then X_t is a supermartingale. If additionally X has constant mean, i.e. $\mathbb{E}(X_t) = \mathbb{E}(X_0)$ for all t then X_t is a martingale.

Proof
As X_t is a continuous local martingale there is a sequence of stopping times $T_n \uparrow \infty$ such that X^{T_n} is a genuine martingale. From this martingale property

$$\mathbb{E}(X_t^{T_n}|\mathcal{F}_s) = X_s^{T_n}.$$

As $X_t \geq 0$ we can apply the conditional form of Fatou's lemma, so

$$\mathbb{E}(X_t|\mathcal{F}_s) = \mathbb{E}(\liminf_{n\to\infty} X_t^{T_n}|\mathcal{F}_s) \leq \liminf_{n\to\infty} \mathbb{E}(X_t^{T_n}|\mathcal{F}_s) = \liminf_{n\to\infty} X_s^{T_n} = X_s.$$

Hence $\mathbb{E}(X_t|\mathcal{F}_s) \leq X_s$, so X_t is a supermartingale.

Given the constant mean property $\mathbb{E}(X_t) = \mathbb{E}(X_s)$. Let

$$A_n := \{\omega : X_s - \mathbb{E}(X_t|\mathcal{F}_s) > 1/n\},$$

so

$$A := \bigcup_{n=1}^{\infty} A_n = \{\omega : X_s - \mathbb{E}(X_t|\mathcal{F}_s) > 0\}.$$

Consider $\mathbb{P}(A) = \mathbb{P}(\cup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n)$. Suppose for some n, $\mathbb{P}(A_n) > \epsilon$, then note that

$$\begin{aligned} \omega \in A_n \quad &: \quad X_s - \mathbb{E}(X_t|\mathcal{F}_s) > 1/n \\ \omega \in \Omega/A_n \quad &: \quad X_s - \mathbb{E}(X_t|\mathcal{F}_s) \geq 0 \end{aligned}$$

Hence

$$X_s - \mathbb{E}(X_t|\mathcal{F}_s) \geq \frac{1}{n} 1_{A_n},$$

taking expectations yields

$$\mathbb{E}(X_s) - \mathbb{E}(X_t) > \frac{\epsilon}{n},$$

but by the constant mean property the left hand side is zero; hence a contradiction, thus all the $\mathbb{P}(A_n)$ are zero, so

$$X_s = \mathbb{E}(X_t|\mathcal{F}_s) \quad \text{a.s.}$$

$\square$

7. The Integral

We would like eventually to extend the definition of the integral to integrands which are *previsible* processes and integrators which are *semimartingales* (to be defined later in **these** notes). In fact in these notes we'll only get as far as *continuous semimartingales*; but it is possible to go the whole way and define the integral of a previsible process with respect to a general semimartingale; but some extra problems are thrown up on the way, in particular as regards the construction of the quadratic variation process of a discontinuous process.

Various special classes of process will be needed in the sequel and these are all defined here for convenience. Naturally with terms like 'elementary' and 'simple' occurring many books have different names for the same concepts – so beware!

7.1. Elementary Processes

An elementary process $H_t(\omega)$ is one of the form

$$H_t(\omega) = Z(\omega)1_{(S(\omega),T(\omega)]}(t),$$

where S, T are stopping times, $S \leq T \leq \infty$, and Z is a bounded $\mathcal{F}_S$ measurable random variable.

Such a process is the simplest non-trivial example of a *previsible* process. Let's prove that it is previsible:

H is clearly a left continuous process, so we need only show that it is adapted. It can be considered as the pointwise limit of a sequence of right continuous processes

$$H_n(t) = \lim_{n\to\infty} Z1_{[S_n,T_n)}, \qquad S_n = S + \frac{1}{n}, \qquad T_n = T + \frac{1}{n}.$$

So it is sufficient to show that $Z1_{[U,V)}$ is adapted when U and V are stopping times which satisfy $U \leq V$, and Z is a bounded $\mathcal{F}_U$ measurable function. Let B be a borel set of $\mathbb{R}$, then the event

$$\{Z1_{[U,V)}(t) \in B\} = [\{Z \in B\} \cap \{U \leq t\}] \cap \{V > t\}.$$

By the definition of U as a stopping time and hence the definition of $\mathcal{F}_U$, the event enclosed by square brackets is in $\mathcal{F}_t$, and since V is a stopping time $\{V > t\} = \Omega/\{V \leq t\}$ is also in $\mathcal{F}_t$; hence $Z1_{[U,V)}$ is adapted. $\square$

7.2. Strictly Simple and Simple Processes

A process H is *strictly simple* ($H \in \mathcal{L}^*$) if there exist $0 \leq t_0 \leq \cdots \leq t_n < \infty$ and uniformly bounded $\mathcal{F}_{t_k}$ measurable random variables Z_k such that

$$H = H_0(\omega)1_0(t) \sum_{k=0}^{n-1} Z_k(\omega)1_{(t_k,t_{k+1}](t)}.$$

This can be extended to H is a *simple processes* ($H \in \mathcal{L}$), if there exists a sequence of stopping times $0 \leq T_0 \leq \cdots \leq T_k \to \infty$, and Z_k uniformly bounded $\mathcal{F}_{T_k}$ measurable random variables such that

$$H = H_0(\omega)1_0(t) + \sum_{k=0}^{\infty} Z_k 1_{(T_k, T_{k+1}]}.$$

Similarly a simple process is also a previsible process. The fundamental result will follow from the fact that the σ-algebra generated by the simple processes is exactly the previsible σ-algebra. We shall see the application of this after the next section.

8. The Stochastic Integral

As has been hinted at earlier the stochastic integral must be built up in stages, and to start with we shall consider integrators which are L^2 bounded martingales, and integrands which are simple processes.

8.1. Integral for $H \in \mathcal{L}$ and $M \in \mathcal{M}_2$

For a simple process $H \in \mathcal{L}$, and M an L^2 bounded martingale then the integral may be defined by the 'martingale transform' (c.f. discrete martingale theory)

$$\int_0^t H_s \mathrm{d}M_s = (H \cdot M)_t := \sum_{k=0}^{\infty} Z_k \left(M_{T_{k+1} \wedge t} - M_{T_k \wedge t} \right)$$

Proposition 8.1.
If H is a simple process, M a L_2 bounded martingale, and T a stopping time. Then
(i) $(H \cdot M)^T = (H 1_{(0,T]}) \cdot M = H \cdot (M^T)$.
(ii) $(H \cdot M) \in \mathcal{M}_2$.
(iii) $E[(H \cdot M)^2_\infty] = \sum_{k=0}^{\infty} \mathbb{E}[Z_k^2 (M^2_{T_{k+1}} - M^2_{T_k})] \le \|H\|^2_\infty \mathbb{E}(M^2_\infty)$.
Proof

Part (i)
As $H \in \mathcal{L}$ we can write

$$H = \sum_{k=0}^{\infty} Z_k 1_{(T_k, T_{k+1}]},$$

for T_k stopping times, and Z_k an $\mathcal{F}_{T_k}$ measurable bounded random variable. By our definition for $M \in \mathcal{M}^2$, we have

$$(H \cdot M)_t = \sum_{k=0}^{\infty} Z_k \left(M_{T_{k+1} \wedge t} - M_{T_k \wedge t} \right),$$

and so, for T a general stopping time consider $(H \cdot M)^T_t = (H \cdot M)_{T \wedge t}$ and so

$$(H \cdot M)^T_t = \sum_{k=0}^{\infty} Z_k \left(M_{T_{k+1} \wedge T \wedge t} - M_{T_k \wedge T \wedge t} \right).$$

Similar computations can be performed for $(H \cdot M^T)$, noting that $M^T_t = M_{T \wedge t}$ and for $(H 1_{(0,T]} \cdot M)$ yielding the same result in both cases. Hence

$$(H \cdot M)^T = (H 1_{(0,T]} \cdot M) = (H \cdot M^T).$$

Part (ii)
To prove this result, first we shall establish it for an elementary process $H \in \mathcal{E}$, and then extend to $\mathcal{L}$ by linearity. Suppose

$$H = Z1_{(R,S]},$$

where R and S are stopping times and Z is a bounded $\mathcal{F}_R$ measurable random variable. Let T be an arbitrary stopping time. We shall prove that

$$\mathbb{E}\left((H \cdot M)_T\right) = \mathbb{E}\left((H \cdot M)_0\right),$$

and hence via optional stopping conclude that $(H \cdot M)_t$ is a martingale.

Note that

$$(H \cdot M)_\infty = Z\left(M_S - M_R\right),$$

and hence as M is a martingale, and Z is $\mathcal{F}_R$ measurable we obtain

$$\begin{aligned}\mathbb{E}(H \cdot M)_\infty =& \mathbb{E}\left(\mathbb{E}\left(Z\left(M_S - M_R\right)\right)|\mathcal{F}_R\right) = \mathbb{E}\left(Z\mathbb{E}\left(\left(M_S - M_R\right)|\mathcal{F}_R\right)\right)\\ =& 0.\end{aligned}$$

Via part (i) note that $\mathbb{E}(H \cdot M)_T = \mathbb{E}(H \cdot M^T)$, so

$$\mathbb{E}(H \cdot M)_T = \mathbb{E}(H \cdot M^T)_\infty = 0.$$

Thus $(H \cdot M)_t$ is a martingale by optional stopping theorem. By linearity, this result extends to H a simple process (i.e. $H \in \mathcal{L}$).

Part (iii)
We wish to prove that $(H \cdot M)$ is an L^2 bounded martingale. We again start by considering $H \in \mathcal{E}$, an elementary process, i.e.

$$H = Z1_{(R,S]},$$

where as before R and S are stopping times, and Z is a bounded $\mathcal{F}_R$ measurable random variable.

$$\begin{aligned}\mathbb{E}\left((H \cdot M)_\infty^2\right) =& \mathbb{E}\left(Z^2(M_S - M_R)^2\right),\\ =& \mathbb{E}\left(Z^2\mathbb{E}\left((M_S - M_R)^2|\mathcal{F}_R\right)\right),\end{aligned}$$

where Z^2 is removed from the conditional expectation since it is and $\mathcal{F}_R$ measurable random variable. Using the same argument as used in the orthogonality of martingale increments proof,

$$\mathbb{E}\left((H \cdot M)_\infty^2\right) = \mathbb{E}\left(Z^2\mathbb{E}\left((M_S^2 - M_R^2)|\mathcal{F}_R\right)\right) = \mathbb{E}\left(Z^2\left(M_S^2 - M_R^2\right)\right).$$

As M is an L^2 bounded martingale and Z is a bounded process,

$$\mathbb{E}\left((H \cdot M)_\infty^2\right) \le \sup_{\omega \in \Omega} 2|Z(\omega)|^2\mathbb{E}\left(M_\infty^2\right).$$

so $(H \cdot M)$ is an L^2 bounded martingale; so together with part (ii), $(H \cdot M) \in \mathcal{M}_2$.

To extend this to simple processes is similar, but requires a little care In general the orthogonality of increments arguments extends to the case where only finitely many of the Z_k in the definition of the simple process H are non zero. Let K be the largest k such that $Z_k \not\equiv 0$.

$$\mathbb{E}\left((H \cdot M)^2_\infty\right) = \sum_{k=0}^{K} \mathbb{E}\left(Z_k^2\left(M^2_{T_{k+1}} - M^2_{T_k}\right)\right),$$

which can be bounded as

$$\begin{aligned}\mathbb{E}\left((H \cdot M)^2_\infty\right) \le& \|H_\infty\|^2 \mathbb{E}\left(\sum_{k=0}^{K}\left(M^2_{T_{k+1}} - M^2_{T_k}\right)\right)\\ \le& \|H_\infty\|^2 \mathbb{E}\left(M^2_{T_{K+1}} - M^2_{T_0}\right) \le \|H_\infty\|^2 \mathbb{E} M^2_\infty,\end{aligned}$$

since we require $T_0 = 0$, and $M \in \mathcal{M}_2$, so the final bound is obtained via the L^2 martingale convergence theorem.

Now extend this to the case of an infinite sum; let $n \le m$, we have that

$$(H \cdot M)_{T_m} - (H \cdot M)_{T_n} = (H 1_{(T_n, T_m]} \cdot M),$$

applying the result just proven for finite sums to the right hand side yields

$$\begin{aligned}\left\|(H \cdot M)^{T_m}_\infty - (H \cdot M)^{T_n}_\infty\right\|_2^2 =& \sum_{k=n}^{m-1} \mathbb{E}\left(Z_k^2\left(M^2_{T_{k+1}} - M^2_{T_k}\right)\right)\\ \le& \|H_\infty\|_2^2\, \mathbb{E}\left(M^2_\infty - M^2_{T_n}\right).\end{aligned}$$

But by the L^2 martingale convergence theorem the right hand side of this bound tends to zero as $n \to \infty$; hence $(H \cdot M)^{T_n}$ converges in $\mathcal{M}_2$ and the limit must be the pointwise limit $(H \cdot M)$. Let $n = 0$ and $m \to \infty$ and the result of part (iii) is obtained. □

8.2. Quadratic Variation

We mentioned earlier that the total variation is the variation which is used by the usual Lebesgue-Stieltjes integral, and that this cannot be used for defining a stochastic integral, since any continuous local martingale of finite variation is indistinguishable from zero. We are now going to look at a variation which will prove fundamental for the construction of the integral. All the definitions as given here aren't based on the partition construction. This is because I shall follow Dellacherie and Meyer and show that the other definitions are equivalent by using the stochastic integral.

Theorem 8.2.

The quadratic variation process $\langle M \rangle_t$ of a **continuous** *L^2 integrable martingale M is the unique process A_t starting from zero such that $M_t^2 - A_t$ is a uniformly integrable martingale.*

Proof
For each n define stopping times

$$S_0^n := 0, \quad S_{k+1}^n := \inf\left\{t > S_k^n : \left|M_t - M_{S_k^n}\right| > 2^{-n}\right\} \text{ for } k \geq 0$$

Define

$$T_k^n := S_k^n \wedge t$$

Then

$$\begin{aligned} M_t^2 =& \sum_{k\geq 1}\left(M_{t\wedge S_k^n}^2 - M_{t\wedge S_{k-1}^n}^2\right) \\ =& \sum_{k\geq 1}\left(M_{T_k^n}^2 - M_{T_{k-1}^n}^2\right) \\ =& 2\sum_{k\geq 1} M_{T_{k-1}^n}\left(M_{T_k^n} - M_{T_{k-1}^n}\right) + \sum_{k\geq 1}\left(M_{T_k^n} - M_{T_{k-1}^n}\right)^2 \end{aligned} \tag{$*$}$$

Now define H^n to be the simple process given by

$$H^n := \sum_{k\geq 1} M_{S_{k-1}^n} 1_{(S_{k-1}^n, S_k^n]}.$$

We can then think of the first term in the decomposition $(*)$ as $(H^n \cdot M)$. Now define

$$A_t^n := \sum_{k\geq 1}\left(M_{T_k^n} - M_{T_{k-1}^n}\right)^2,$$

so the expression $(*)$ becomes

$$M_t^2 = 2(H^n \cdot M)_t + A_t^n. \tag{$**$}$$

From the construction of the stopping times S_k^n we have the following properties

$$\begin{aligned} \|H^n - H^{n+1}\|_\infty =& \sup_t |H_t^n - H_t^{n+1}| \leq 2^{-(n+1)} \\ \|H^n - H^{n+m}\|_\infty =& \sup_t |H_t^n - H_t^{n+m}| \leq 2^{-(n+1)} \text{ for all } m \geq 1 \\ \|H^n - M\|_\infty =& \sup_t |H_t^n - M_t| \leq 2^{-n} \end{aligned}$$

Let $J_n(\omega)$ be the set of all stopping times $S_k^n(\omega)$ i.e.

$$J_n(\omega) := \{S_k^n(\omega) : k \geq 0\}.$$

Clearly $J_n(\omega) \subset J_{n+1}(\omega)$. Now for any $m \geq 1$, using Proposition 8.1(iii) the following result holds

$$\begin{aligned} \mathbb{E}\left(\left[(H^n \cdot M) - (H^{n+m} \cdot M)\right]_\infty^2\right) =& \mathbb{E}\left(\left[\left(\left\{H^n - H^{n+m}\right\} \cdot M\right)\right]_\infty^2\right) \\ \leq& \|H^n - H^{n+m}\|_\infty^2 \mathbb{E}(M_\infty^2) \\ \leq& \left(2^{-(n+1)}\right)^2 \mathbb{E}(M_\infty^2). \end{aligned}$$

Thus $(H^n \cdot M)_\infty$ is a Cauchy sequence in the complete Hilbert space $L^2(\mathcal{F}_\infty)$; hence by completeness of the Hilbert Space it converges to a limit in the same space. As $(H^n \cdot M)$ is a continuous martingale for each n, so is the the limit N say. By Doob's L^2 inequality applied to the continuous martingale $(H^n \cdot M) - N$,

$$\mathbb{E}\left(\sup_{t \geq 0} |(H^n \cdot M) - N|^2\right) \leq 4\mathbb{E}\left([(H \cdot M) - N]_\infty^2\right) \to_{n\to\infty} 0.$$

Hence $(H^n \cdot M)$ converges to N uniformly a.s.. From the relation $(**)$ we see that as a consequence of this, the process A^n converges uniformly a.s. to a process A, where

$$M_t^2 = 2N_t + A_t.$$

Now we must check that this limit process A is increasing. Clearly $A^n(S_k^n) \leq A^n(S_{k+1}^n)$, and since $J_n(\omega) \subset J_{n+1}(\omega)$, it is also true that $A(S_k^n) \leq A(S_{k+1}^n)$ for all n and k, and so A is certainly increasing on the closure of $J(\omega) := \cup_n J_n(\omega)$. However if I is an open interval in the complement of J, then no stopping time S_k^n lies in this interval, so M must be constant throughout I, so the same is true for the process A. Hence the process A is continuous, increasing, and null at zero; such that $M_t^2 - A_t = 2N_t$, where N_t is a UI martingale (since it is L^2 bounded). Thus we have established the existence result. It only remains to consider uniqueness.

Uniqueness follows from the result that a continuous local martingale of finite variation is everywhere zero. Suppose the process A in the above definition were not unique. That is suppose that also for some B_t continuous increasing from zero, $M_t^2 - B_t$ is a UI martingale. Then as $M_t^2 - A_t$ is also a UI martingale by subtracting these two equations we get that $A_t - B_t$ is a UI martingale, null at zero. It clearly must have finite variation, and hence be zero. $\square$

The following corollary will be needed to prove the integration by parts formula, and can be skipped on a first reading; however it is clearer to place it here, since this avoids having to redefine the notation.

Corollary 8.3.
Let M be a bounded continuous martingale, starting from zero. Then

$$M_t^2 = 2\int_0^t M_s \mathrm{d}M_s + \langle M \rangle_t.$$

Proof
In the construction of the quadratic variation process the quadratic variation was constructed as the uniform limit in L^2 of processes A_t^n such that

$$A_t^n = M_t^2 - 2(H^n \cdot M)_t,$$

where each H^n was a bounded previsible process, such that

$$\sup_t |H_t^n - M| \leq 2^{-n},$$

and hence $H^n \to M$ in $L^2(M)$, so the martingales $(H^n \cdot M)$ converge to $(M \cdot M)$ uniformly in L^2, hence it follows immediately that

$$M_t^2 = 2\int_0^t M_s \mathrm{d}M_s + \langle \mathrm{M} \rangle_t,$$

$\square$

Theorem 8.4.
The quadratic variation process $\langle M \rangle_t$ of a **continuous** *local martingale M is the unique increasing process A, starting from zero such that $M^2 - A$ is a local martingale.*

Proof
We shall use a localisation technique to extend the definition of quadratic variation from L^2 bounded martingales to general local martingales.

The mysterious seeming technique of localisation isn't really that complex to understand. The idea is that it enables us to extend a definition which applies for 'X widgets' to one valid for 'local X widgets'. It achieves this by using a sequence of stopping times which reduce the 'local X widgets' to 'X widgets' ; the original definition can then be applied to the stopped version of the 'X widget'. We only need to check that we can sew up the pieces without any holes i.e. that our definition is independent of the choice of stopping times!

Let $T_n = \inf\{t : |M_t| > n\}$, define a sequence of stopping times. Now define

$$\langle \mathrm{M} \rangle_t := \langle M^{T_n} \rangle \text{ for } 0 \leq t \leq T_n$$

To check the consistency of this definition note that

$$\langle M^{T_n} \rangle^{T_{n-1}} = \langle M^{T_{n-1}} \rangle$$

and since the sequence of stopping times $T_n \to \infty$, we see that $\langle \mathrm{M} \rangle$ is defined for all t. Uniqueness follows from the result that any finite variation continuous local martingale starting from zero is identically zero. $\square$

The quadratic variation turns out to be the 'right' sort of variation to consider for a martingale; since we have already shown that all but the zero martingale have infinite total variation; and it can be shown that the higher order variations of a martingale are zero a.s.. Note that the definition given is for a **continuous local martingale**; we shall see later how to extend this to a continuous semimartingale.

8.3. Covariation

From the definition of the quadratic variation of a local martingale we can define the covariation of two local martingales N and M which are locally L^2 bounded via the *polarisation identity*

$$\langle M, N \rangle := \frac{\langle M + N \rangle - \langle M - N \rangle}{4}.$$

We need to generalise this slightly, since the above definition required the quadratic variation terms to be finite. We can prove the following theorem in a straightforward manner using the definition of quadratic variation above, and this will motivate the general definition of the *covariation* process.

Theorem 8.5.
For M and N two local martingales which are locally L^2 bounded then there exists a unique finite variation process A starting from zero such that $MN - A$ is a local martingale. This process A is the covariation of M and N.

This theorem is turned round to give the usual definition of the covariation process of two continuous local martingales as:

Definition 8.6.
For two continuous local martingales N and M, there exists a unique finite variation process A, such that $MN - A$ is a local martingale. The covariance process of N and M is defined as this process A.

It can readily be verified that the covariation process can be regarded as a symmetric bilinear form on the space of local martingales, i.e. for L,M and N continuous local martingales

$$\begin{aligned} \langle M+N, L\rangle =& \langle M, L\rangle + \langle N, L\rangle, \\ \langle M, N\rangle =& \langle N, M\rangle, \\ \langle \lambda M, N\rangle =& \lambda \langle M, N\rangle, \ \lambda \in \mathbb{R}. \end{aligned}$$

8.4. Extension of the Integral to $L^2(M)$

We have previously defined the integral for H a simple process (in $\mathcal{L}$), and $M \in \mathcal{M}_2^c$, and we have noted that $(H \cdot M)$ is itself in $\mathcal{M}_2$. Hence

$$\mathbb{E}\left((H \cdot M)_\infty^2\right) = \mathbb{E}\left(Z_{i-1}^2\left(M_{T_i} - M_{T_{i-1}}\right)^2\right)$$

Recall that for $M \in \mathcal{M}_2$, then $M^2 - \langle \mathrm{M}\rangle$ is a uniformly integrable martingale. Hence for S and T stopping times such that $S \leq T$, then

$$\mathbb{E}\left((M_T - M_S)^2 | \mathcal{F}_S\right) = \mathbb{E}(M_T^2 - M_S^2 | \mathcal{F}_S) = \mathbb{E}(\langle \mathrm{M}\rangle_T - \langle \mathrm{M}\rangle_S | \mathcal{F}_S).$$

So summing we obtain

$$\begin{aligned} \mathbb{E}\left((H \cdot M)_\infty^2\right) =& \mathbb{E} \sum Z_{i-1}^2 \left(\langle \mathrm{M}\rangle_{T_i} - \langle \mathrm{M}\rangle_{T_{i-1}}\right), \\ =& \mathbb{E}\left((H^2 \cdot \langle \mathrm{M}\rangle)_\infty\right). \end{aligned}$$

In the light of this, we define a seminorm $\|H\|_M$ via

$$\|H\|_M = \left[\mathbb{E}\left((H^2 \cdot \langle \mathrm{M}\rangle)_\infty\right)\right]^{1/2} = \left[\mathbb{E}\left(\int_0^\infty H_s^2 \mathrm{d}\langle \mathrm{M}\rangle_s\right)\right]^{1/2}.$$

The space $\mathcal{L}^2(M)$ is then defined as the subspace of the previsible processes, where this seminorm is finite, i.e.

$$\mathcal{L}^2(M) := \{\text{previsible processes } H \text{ such that } \|H\|_M < \infty\}.$$

However we would actually like to be able to treat this as a Hilbert space, and there remains a problem, namely that if $X \in \mathcal{L}^2(M)$ and $\|X\|_M = 0$, this doesn't imply that X is the zero process. Thus we follow the usual route of defining an equivalence relation via $X \sim Y$ if and only if $\|X - Y\|_M = 0$. We now define

$$L^2(M) := \{\text{equivalence classes of previsible processes } H \text{ such that } \|H\|_M < \infty\},$$

and this is a Hilbert space with norm $\|\cdot\|_M$ (it can be seen that it is a Hilbert space by considering it as suitable L^2 space).

This establishes an isometry (called the *Itô isometry*) between the spaces $L^2(M) \cap \mathcal{L}$ and $L^2(\mathcal{F}_\infty)$ given by

$$\begin{aligned} I :& L^2(M) \cap \mathcal{L} \to L^2(\mathcal{F}_\infty) \\ I :& H \mapsto (H \cdot M)_\infty \end{aligned}$$

Remember that there is a basic bijection between the space $\mathcal{M}_2$ and the Hilbert Space $L^2(\mathcal{F}_\infty)$ in which each square integrable martingale M is represented by its limiting value M_∞, so the image under the isometry $(H \cdot M)_\infty$ in $L^2(\mathcal{F}_\infty)$ may be thought of a describing an $\mathcal{M}_2$ martingale. Hence this endows $\mathcal{M}_2$ with a Hilbert Space structure, with an inner product given by

$$(M, N) = \mathbb{E}\left(N_\infty M_\infty\right).$$

We shall now use this Itô isometry to extend the definition of the stochastic integral from $\mathcal{L}$ (the class of simple processes) to the whole of $L^2(M)$. Roughly speaking we shall approximate an element of $L^2(M)$ via a sequence of simple processes converging to it; just as in the construction of the Lebesgue Integral. In doing this, we shall use the *Monotone Class Theorem.*

Recall that in the conventional construction of the Lebesgue integration, and proof of the elementary results the following standard machine is repeatedly invoked. To prove a 'linear' result for all $h \in L^1(S, \Sigma, \mu)$, proceed in the following way:

(i) Show the result is true for h an indicator function.
(ii) Show that by linearity the result extends to all positive step functions.
(iii) Use the Monotone convergence theorem to see that if $h_n \uparrow h$, where the h_n are step functions, then the result must also be true for h a non-negative, Σ measurable function.
(iv) Write $h = h^+ - h^-$ where both h^+ and h^- are non-negative functions and use linearity to obtain the result for $h \in L^1$.

The monotone class lemmas is a replacement for this procedure, which hides away all the 'machinery' used in the constructions.

Monotone Class Theorem.
Let $\mathcal{A}$ be π-system generating the σ-algebra $\mathcal{F}$ (i.e. $\sigma(\mathcal{A}) = \mathcal{F}$). If $\mathcal{H}$ is a linear set of bounded functions from Ω to $\mathbb{R}$ satisfying

(i) $1_A \in \mathcal{H}$, for all $A \in \mathcal{A}$,

(ii) $0 \le f_n \uparrow f$, *where* $f_n \in \mathcal{H}$ *and* f *is a bounded function* $f : \Omega \to \mathbb{R}$, *then this implies that* $f \in \mathcal{H}$,

then $\mathcal{H}$ contains every bounded, $\mathcal{F}$-measurable function $f : \Omega \to \mathbb{R}$.

In order to apply this in our case, we need to prove that the σ-algebra of previsible processes is that generated by the simple processes.

The Previsible σ-field and the Simple Processes

It is fairly simple to show that the space of simple processes $\mathcal{L}$ forms a vector space (exercise: check linearity, constant multiples and zero).

Lemma 8.7.

The σ-algebra generated by the simple processes is the previsible σ-algebra i.e. the previsible σ-algebra us the smallest σ-algebra with respect to which every simple process is measurable.

Proof

It suffices to show that every left continuous right limit process, which is bounded and adapted to $\mathcal{F}_t$ is measurable with respect to the σ-algebra generated by the simple processes. Let H_t be a bounded left continuous right limits process, then

$$H = \lim_{k\to\infty} \lim_{n\to\infty} \sum_{i=2}^{nk} H_{(i-1)/n} \left(\frac{i-1}{n}, \frac{i}{n} \right],$$

and if H_t is adapted to $\mathcal{F}_t$ then $H_{(i-1)/n}$ is a bounded element of $\mathcal{F}_{(i-1)/n}$.

We can now apply the Monotone Class Theorem to the vector space $\mathcal{H}$ of processes with a time parameter in $(0, \infty)$, regarded as maps from $(0, \infty) \times \Omega \to \mathbb{R}$. Then if this vector space contains all the simple processes i.e. $\mathcal{L} \subset \mathcal{H}$, then $\mathcal{H}$ contains every bounded previsible process on $(0, \infty)$.

Assembling the Pieces

Since I is an isometry it has a unique extension to the closure of

$$\mathcal{U} = L^2(M) \cap \mathcal{L},$$

in $L^2(M)$. By the application of the monotone class lemma to $\mathcal{H} = \overline{\mathcal{U}}$, and the π-system of simple processes. We see that $\overline{\mathcal{U}}$ must contain every bounded previsible process; hence $\overline{\mathcal{U}} = L^2(M)$. Thus the Itô Isometry extends to a map from $L^2(M)$ to $L^2(\mathcal{F}_\infty)$.

Let us look at this result more informally. For a previsible $H \in L^2(M)$, because of the density of $\mathcal{L}$ in $L^2(M)$, we can find a sequence of simple processes H_n which converges to H, as $n \to \infty$. We then consider $I(H)$ as the limit of the $I(H_n)$. To verify that this limit is unique, suppose that $H'_n \to H$ as $n \to \infty$ also, where $H'_n \in \mathcal{L}$. Note that $H_n - H'_n \in \mathcal{L}$. Also $H_n - H'_n \to 0$ and so $((H_n - H'_n) \cdot M) \to 0$, and hence by the Itô isometry the limits $\lim_{n\to\infty}(H_n \cdot M)$ and $\lim_{n\to\infty}(H'_n \cdot M)$ coincide.

The following result is essential in order to extend the integral to continuous local martingales.

Proposition 8.8.
For $M \in \mathcal{M}_2$, for any $H \in L^2(M)$ and for any stopping time T then

$$(H \cdot M)^T = (H1_{(0,T]} \cdot M) = (H \cdot M^T).$$

Proof
Consider the following linear maps in turn

$$\begin{aligned} f_1 :& L^2(\mathcal{F}_\infty) \to L^2(\mathcal{F}_\infty) \\ f_1 :& Y \mapsto \mathbb{E}(Y|\mathcal{F}_T) \end{aligned}$$

This map is a contraction on $L^2(\mathcal{F}_\infty)$ since by conditional Jensen's inequality

$$\mathbb{E}(Y_\infty|\mathcal{F}_T)^2 \leq \mathbb{E}(Y_\infty^2|\mathcal{F}_T),$$

and taking expectations yields

$$\|\mathbb{E}(Y|\mathcal{F}_T)\|_2^2 = \mathbb{E}\left(\mathbb{E}(Y_\infty|\mathcal{F}_T)^2\right) \leq \mathbb{E}\left(\mathbb{E}(Y_\infty^2|\mathcal{F}_T)\right) = \mathbb{E}(Y_\infty^2) = \|Y\|_2^2.$$

Hence f_1 is a contraction on $L^2(\mathcal{F}_\infty)$. Now

$$\begin{aligned} f_2 :& L^2(M) \to L^2(M) \\ f_2 :& H \mapsto H1_{(0,T]} \end{aligned}$$

Clearly from the definition of $\|\cdot\|_M$, and from the fact that the quadratic variation process is increasing

$$\left\|H1_{(0,T]}\right\|_M = \int_0^\infty H_s^2 1_{(0,T]} \mathrm{d}\langle \mathrm{M}\rangle_s = \int_0^T H_s^2 \mathrm{d}\langle \mathrm{M}\rangle_s \leq \int_0^\infty H_s^2 \mathrm{d}\langle \mathrm{M}\rangle_s = \|H\|_M .$$

Hence f_2 is a contraction on $L^2(M)$. Hence if I denotes the Itô isometry then $f_1 \circ I$ and $I \circ f_2$ are also contractions from $L^2(M)$ to $L^2(\mathcal{F}_\infty)$, (using the fact that I is an isometry between $L^2(M)$ and $L^2(\mathcal{F}_\infty)$).

Now introduce $I^{(T)}$, the stochastic integral map associated with M^T, i.e.

$$I^{(T)}(H) \equiv (H \cdot M^T)_\infty.$$

Note that

$$\|I^{(T)}(H)\|_2 = \|H\|_{M^T} \leq \|H\|_M.$$

We have previously shown that the maps $f_1 \circ I$ and $I \circ f_2$ and $H \mapsto I^{(T)}(H)$ agree on the space of simple processes by direct calculation. We note that $\mathcal{L}$ is dense in $L^2(M)$ (from application of Monotone Class Lemma to the simple processes). Hence from the three bounds above the three maps agree on $L^2(M)$. □

8.5. Localisation

We've already met the idea of localisation in extending the definition of quadratic variation from L^2 bounded continuous martingales to continuous local martingales. In this context a previsible process $\{H_t\}_{t\geq 0}$, is *locally previsible* if there exists a sequence of stopping times $T_n \to \infty$ such that for all n $H1_{(0,T_n]}$ is a previsible process. Fairly obviously every previsible process has this property. However if in addition we want the process H to be *locally bounded* we need the condition that there exists a sequence T_n of stopping times, tending to infinity such that $H1_{(0,T_n]}$ is uniformly bounded for each n.

For the integrator (a martingale of integrable variation say), the localisation is to a local martingale, that is one which has a sequence of stopping times $T_n \to \infty$ such that for all n, X^{T_n} is a genuine martingale.

If we can prove a result like

$$(H \cdot X)^T = (H1_{(0,T]} \cdot X^T)$$

for H and X in their original (i.e. non-localised classes) then it is possible to *extend* the definition of $(H \cdot X)$ to the local classes.

Note firstly that for H and X local, and T_n a reducing sequence[1] of stopping times for both H and X then we see that $(H1_{(0,T]} \cdot X^T)$ is defined in the existing fashion. Also note that if $T = T_{n-1}$ we can check consistency

$$(H1_{(0,T_n]} \cdot X^{T_n})^{T_{n-1}} = (H \cdot X)^{T_{n-1}} = (H1_{(0,T_{n-1}]} \cdot X^{T_{n-1}}).$$

Thus it is consistent to define $(H \cdot X)_t$ on $t \in [0, \infty)$ via

$$(H \cdot X)^{T_n} = (H1_{(0,T_n]} \cdot X^{T_n}), \quad \forall n.$$

We must check that this is well defined, viz if we choose another regularising sequence S_n, we get the same definition of $(H \cdot X)$. To see this note:

$$(H1_{(0,T_n]} \cdot X^{T_n})^{S_n} = (H1_{(0,T_n\wedge S_n]} \cdot X^{T_n\wedge S_n}) = (H1_{(0,S_n]} \cdot X^{S_n})^{T_n},$$

hence the definition of $(H \cdot X)_t$ is the same if constructed from the regularising sequence S_n as if constructed via T_n.

8.6. Some Important Results

We can now extend most of our results to stochastic integrals of a previsible process H with respect to a **continuous** local martingale M. In fact in these notes we will never drop the continuity requirement. It can be done; but it requires considerably more work, especially with regard to the definition of the quadratic variation process.

---

[1] The reducing sequence is the sequence of stopping times tending to infinity which makes the local version of the object into the non-local version. We can find one such sequence, because if say $\{T_n\}$ reduces H and $\{S_n\}$ reduces X then $T_n \wedge S_n$ reduces both H and X.

Theorem 8.9.
Let H be a locally bounded previsible process, and M a continuous local martingale. Let T be an arbitrary stopping time. Then:
(i) $(H \cdot M)^T = (H1_{(0,T]} \cdot M) = (H \cdot M^T)$
(ii) $(H \cdot M)$ *is a continuous local martingale*
(iii) $\langle H \cdot M \rangle = H^2 \cdot \langle M \rangle$
(iv) $H \cdot (K \cdot M) = (HK) \cdot M$

Proof
The proof of parts (i) and (ii) follows from the result used in the localisation that:

$$(H \cdot M)^T = (H1_{(0,T]} \cdot M) = (H \cdot M^T)$$

for H bounded previsible process in $L^2(M)$ and M an L^2 bounded martingale. Using this result it suffices to prove (iii) and (iv) where M, H and K are **uniformly** bounded (via localisation).

Part (iii)

$$\begin{aligned}\mathbb{E}\left[(H \cdot M)_T^2\right] =& \mathbb{E}\left[\big(H1_{(0,T]} \cdot M\big) \cdot M)_\infty^2\right] \\ =& \mathbb{E}\left[\big(H1_{(0,T]} \cdot \langle \mathrm{M} \rangle\big)_\infty^2\right] \\ =& \mathbb{E}\left[\big(H^2 \cdot \langle \mathrm{M} \rangle\big)_T\right]\end{aligned}$$

Hence we see that $(H \cdot M)^2 - (H^2 \cdot \langle \mathrm{M} \rangle)$ is a martingale (via the optional stopping theorem), and so by uniqueness of the quadratic variation process, we have established

$$\langle H \cdot M \rangle = H^2 \cdot \langle \mathrm{M} \rangle.$$

Part (iv)
The truth of this statement is readily established for H and K simple processes (in $\mathcal{L}$). To extend to H and K bounded previsible processes note that

$$\begin{aligned}\mathbb{E}\left[(H \cdot (K \cdot M))_\infty^2\right] =& \mathbb{E}\left[\big(H^2 \cdot \langle K \cdot M \rangle\big)_\infty\right] \\ =& \mathbb{E}\left[\big(H^2 \cdot (K^2 \cdot \langle \mathrm{M} \rangle)\big)_\infty\right] \\ =& \mathbb{E}\left[\big((HK)^2 \cdot \langle \mathrm{M} \rangle\big)_\infty\right] \\ =& \mathbb{E}\left[((HK) \cdot M)_\infty^2\right]\end{aligned}$$

Also note the following bound

$$\mathbb{E}\left[\big((HK)^2 \cdot \langle \mathrm{M} \rangle\big)_\infty\right] \leq \min\left\{\|H\|_\infty^2 \|K\|_M^2, \|H\|_M^2 \|K\|_\infty^2\right\}.$$

$\square$

9. Semimartingales

I mentioned at the start of these notes that the most general form of the stochastic integral would have a previsible process as the integrand and a semimartingale as an integrator. Now it's time to extend the definition of the Itô integral to the case of semimartingale integrators.

Definition 9.1.
A process X is a semimartingale if X is an adapted CADLAG process which has a decomposition

$$X = X_0 + M + A,$$

where M is a local martingale, null at zero and A is a process null at zero, with paths of finite variation.

Note that the decomposition is **not** necessarily unique as there exist martingales which have finite variation. To remove many of these difficulties we shall impose a continuity condition, since under this most of our problems will vanish.

Definition 9.2.
A continuous semimartingale is a process $(X_t)_{t\geq 0}$ which has a Doob-Meyer decomposition

$$X = X_0 + M + A,$$

where X_0 is $\mathcal{F}_0$-measurable, $M_0 = A_0 = 0$, M_t is a continuous local martingale and A_t is a continuous adapted process of finite variation.

Theorem 9.3.
The Doob-Meyer decomposition in the definition of a continuous semimartingale is unique.

Proof
Let another such decomposition be

$$X = X_0 + M' + A',$$

where M' is a continuous local martingale and A a continuous adapted process of finite variation. Then consider the process N, where

$$N = M' - M = A' - A,$$

by the first equality, N is the difference of two continuous local martingales, and hence is itself a continuous local martingale; and by the second inequality it has finite variation. Hence by an earlier proposition (5.2) it must be zero. Hence $M' = M$ and $A' = A$. $\square$

We **define**† the quadratic variation of the continuous semimartingale as that of the continuous local martingale part i.e. for $X = X_0 + M + A$,

$$\langle \mathrm{X} \rangle := \langle \mathrm{M} \rangle.$$

† These definitions can be made to look natural by considering the quadratic variation defined in terms of a sum of squared increments; but following this approach, these are result which are proved later using the Itô integral, since this provided a better approach to the discontinuous theory.

Similarly if $Y + Y_0 + N + B$ is another semimartingale, where B is finite variation and N is a continuous local martingale, we define

$$\langle X, Y\rangle := \langle M, N\rangle.$$

We can extend the definition of the stochastic integral to continuous semimartingale integrators by defining

$$(H \cdot X) := (H \cdot M) + (H \cdot A),$$

where the first integral is a stochastic integral as defined earlier and the second is a Lebesgue-Stieltjes integral (as the integrator is a process of finite variation).

10. Relations to Sums

This section is optional; and is included to bring together the two approaches to the constructions involved in the stochastic integral.

For example the quadratic variation of a process can either be defined in terms of martingale properties, or alternatively in terms of sums of squares of increments.

10.1. The UCP topology

We shall meet the notion of convergence uniformly on compacts in probability when considering stochastic integrals as limits of sums, so it makes sense to review this topology here.

Definition 10.1.

A sequence $\{H_n\}_{n\geq 1}$ converges to a process H uniformly on compacts in probability (abbreviated u.c.p.) if for each $t > 0$,

$$\sup_{0\leq s\leq t} |H_s^n - H_s| \to 0 \text{ in probability.}$$

At first sight this may seem to be quite an esoteric definition; in fact it is a natural extension of convergence in probability to processes. It would also appear to be quite difficult to handle, however Doob's martingale inequalities provide the key to handling it. Let

$$H_t^* = \sup_{0\leq s\leq t} |H_s|,$$

then for Y^n a CADLAG process, Y^n converges to Y u.c.p. iff $(Y^n - Y)^*$ converges to zero in probability for each $t \geq 0$. Thus to prove that a sequence converges u.c.p. it often suffices to apply Doob's inequality to prove that the supremum converges to zero in L^2, whence it must converge to zero in probability, whence u.c.p. convergence follows.

The space of CADLAG processes with u.c.p. topology is in fact metrizable, a compatible metric is given by

$$\mathrm{d}(X, Y) = \sum_{n=1}^{\infty} \frac{1}{2^n} \mathbb{E}\left(\min(1, (X - Y)_n^*)\right),$$

for X and Y CADLAG processes. The metric space can also be shown to be complete. For details see Protter.

Since we have just met a new kind of convergence, it is helpful to recall the other usual types of convergence on a probability space. For convenience here are the usual definitions:

Pointwise
A sequence of random variables X_n converges to X pointwise if for all ω not in some null set,

$$X_n(\omega) \to X(\omega).$$

Probability
A sequence of r.v.s X_n converges to X in probability, if for any $\epsilon > 0$,

$$\mathbb{P}(|X_n - X| > \epsilon) \to 0, \text{ as } n \to \infty.$$

L^p convergence
A sequence of random variables X_n converges to X in L^p, if

$$\mathbb{E}|X_n - X|^p \to 0, \text{ as } n \to \infty.$$

It is trivial to see that pointwise convergence implies convergence in probability. It is also true that L^p convergence implies convergence in probability as the following theorem shows

Theorem 10.2.
If X_n converges to X in L^p for $p > 0$, then X_n converges to X in probability.

Proof
Apply Chebyshev's inequality to $f(x) = x^p$, which yields for any $\epsilon > 0$,

$$\mathbb{P}(|X_n| \geq \epsilon) \leq \epsilon^{-p}\mathbb{E}(|X_n|^p) \to 0, \text{ as } n \to \infty.$$

□

Theorem 10.3.
If $X_n \to X$ in probability, then there exists a subsequence n_k such that $X_{n_k} \to X$ a.s.

Theorem 10.4.
If $X_n \to X$ a.s., then $X_n \to X$ in probability.

10.2. Approximation via Riemann Sums

Following Dellacherie and Meyer we shall establish the equivalence of the two constructions for the quadratic variation by the following theorem which approximates the stochastic integral via Riemann sums.

Theorem 10.2.
Let X be a semimartingale, and H a locally bounded previsible CADLAG process starting from zero. Then

$$\int_0^t H_s \mathrm{d}X_s = \lim_{n\to\infty} \sum_{k=0}^{\infty} H_{t\wedge k2^{-n}} \left(X_{t\wedge(k+1)2^{-n}} - X_{t\wedge k2^{-n}} \right) \text{ u.c.p.}$$

Proof
Let $K_s = H_s 1_{s\le t}$, and define the following sequence of simple process approximations

$$K_s^n := \sum_{k=0}^{\infty} H_{t\wedge k2^{-n}} 1_{(t\wedge k2^{-n}, t\wedge(k+1)2^{-n}]}(s).$$

Clearly this sequence K_s^n converges pointwise to K_s. We can decompose the semimartingale X as $X = X_0 + A_t + M_t$ where A_t is of finite variation and M_t is a continuous local martingale, both starting from zero. The result that

$$\int_0^t K_s^n \mathrm{d}A_s \to \int_0^t K_s \mathrm{d}A_s, \text{ u.c.p.}$$

is standard from the Lebesgue-Stieltjes theory. Let T_k be a reducing sequence for the continuous local martingale M such that M^{T_k} is a **bounded** martingale. Also since K is locally bounded we can find a sequence of stopping times S_k such that K^{S_k} is a bounded previsible process. It therefore suffices to prove for a sequence of stopping times R_k such that $R_k \uparrow \infty$, then

$$(K^n \cdot M)_s^{R_k} \to (K \cdot M)_s^{R_k}, \text{ u.c.p..}$$

By Doob's L^2 inequality, and the Itô isometry we have

$$\begin{aligned} \mathbb{E}\left[((K^n \cdot M) - (K \cdot M))^*\right]^2 \le & 4\mathbb{E}\left[(K^n \cdot M) - (K \cdot M)\right]^2, && \text{Doob } L^2 \\ \le & 4\|K^n - K\|_M^2, && \text{Itô Isometry} \\ \le & 4\int (K_s^n - K_s)^2 \mathrm{d}\langle \mathrm{M}\rangle_s \end{aligned}$$

As $|K^n - K| \to 0$ pointwise, and K is bounded, clearly $|K^n - K|$ is also bounded uniformly in n. Hence by the Dominated Convergence Theorem for the Lebesgue-Stieltjes integral

$$\int (K_s^n - K_s)^2 \mathrm{d}\langle \mathrm{M}\rangle_s \to 0 \text{ a.s..}$$

Hence, we may conclude

$$\mathbb{E}\left(\left[(K^n \cdot M) - (K \cdot M)\right]^*\right)^2 \to 0, \text{ as } n \to \infty.$$

So

$$\left[(K^n \cdot M) - (K \cdot M)\right]^* \to 0 \text{ in } L^2,$$

as $n \to \infty$; but this implies that

$$[(K^n \cdot M) - (K \cdot M)]^* \to 0 \text{ in probability.}$$

Hence

$$[(K^n \cdot M) - (K \cdot M)] \to 0 \text{ u.c.p.}$$

as required, and putting the two parts together yields

$$\int_0^t K_s^n \mathrm{d}X_s \to \int_0^t K_s \mathrm{d}X_s, \text{ u.c.p.}$$

which is the required result. □

This result can now be applied to the construction of the quadratic variation process, as illustrated by the next theorem.

Theorem 10.3.
The quadratic variation process $\langle X \rangle_t$ is equal to the following limit in probability

$$\langle \mathrm{X} \rangle_t = \lim_{n\to\infty} \sum_{k=0}^{\infty} \left(X_{t\wedge(k+1)2^{-n}} - X_{t\wedge k2^{-n}} \right)^2 \text{ in probability.}$$

Proof
In the theorem (7.2) establishing the existence of the quadratic variation process, we noted in (**) that

$$A_t^n = M_t^2 - 2(H^n \cdot M)_t.$$

Now from application of the previous theorem

$$2\int_0^t X_s \mathrm{d}X_s = \lim_{n\to\infty} \sum_{k=0}^{\infty} X_{t\wedge k2^{-n}} \left(X_{t\wedge(k+1)2^{-n}} - X_{t\wedge k2^{-n}} \right).$$

In addition,

$$X_t^2 - X_0^2 = \sum_{k=0}^{\infty} \left(X_{t\wedge(k+1)2^{-n}}^2 - X_{t\wedge k2^{-n}}^2 \right).$$

The difference of these two equations yields

$$A_t = X_0^2 + \lim_{n\to\infty} \sum_{k=0}^{\infty} \left(X_{t\wedge(k+1)2^{-n}} - X_{t\wedge k2^{-n}} \right)^2,$$

where the limit is taken in probability. Hence the process A is increasing and positive on the rational numbers, and hence on the whole of $\mathbb{R}$ by right continuity. □

Remark
The theorem can be strengthened still further by a result of Doléans-Dade to the effect that for X a continuous semimartingale

$$\langle \mathrm{X} \rangle_t = \lim_{n\to\infty} \sum_{k=0}^{\infty} \left(X_{t\wedge(k+1)2^{-n}} - X_{t\wedge k2^{-n}} \right)^2,$$

where the limit is in the strong sense in L^1. This result is harder to prove (essentially the uniform integrability of the sums must be proven) and this is not done here.

11. Itô's Formula

Itô's Formula is the analog of integration by parts in the stochastic calculus. It is also the first place where we see a major difference creep into the theory, and realise that our formalism has found a new subtlety in the subject.

More importantly, it is the fundamental weapon used to evaluate Itô integrals; we shall see some examples of this shortly.

The Itô isometry provides a clean-cut definition of the stochastic integral; however it was originally defined via the following theorem of Kunita and Watanabe.

Theorem (Kunita-Watanabe Identity) 11.1.
Let $M \in \mathcal{M}_2$ and H and K are locally bounded previsible processes. Then $(H \cdot M)_\infty$ is the unique element of $L^2(\mathcal{F}_\infty)$ such that for every $N \in M_2$ we have:

$$\mathbb{E}\left[(H \cdot M)_\infty N_\infty\right] = \mathbb{E}\left[(H \cdot \langle M, N\rangle)_\infty\right]$$

Moreover we have

$$\langle (H \cdot M), N\rangle = H \cdot \langle M, N\rangle.$$

Proof
Consider an elementary function H, so $H = Z1_{(S,T]}$, where Z is an $\mathcal{F}_S$ measurable bounded random variable, and S and T are stopping times such that $S \leq T$. It is clear that

$$\begin{aligned}\mathbb{E}\left[(H \cdot M)_\infty N_\infty\right] =& \mathbb{E}\left[Z\left(M_T - M_S\right) N_\infty\right]\\ =& \mathbb{E}\left[Z(M_T N_T - M_S N_S)\right]\\ =& \mathbb{E}\left[M_\infty (H \cdot N)_\infty\right]\end{aligned}$$

Now by linearity this can be extended to establish the result for all simple functions (in $\mathcal{L}$). We finally extend to general locally bounded previsible H, by considering a sequence (provided it exists) of simple functions H^n such that $H^n \to H$ in $L^2(M)$. Then there exists a subsequence n_k such that H^{n_k} converges to H is $L^2(N)$. Then

$$\begin{aligned}\mathbb{E}\big((H^{n_k} \cdot M)_\infty N_\infty\big) - \mathbb{E}\big((H \cdot M)_\infty N_\infty\big) =& \mathbb{E}\Big(\big((H^{n_k} - H) \cdot M\big) N_\infty\Big)\\ \leq& \sqrt{\mathbb{E}\left(\left[(H^{n_k} - H) \cdot M)\right]^2\right)}\sqrt{\mathbb{E}(N_\infty^2)}\\ \leq& \sqrt{\mathbb{E}\left(\left[(H^{n_k} \cdot M) - (H \cdot M)\right]^2\right)}\sqrt{\mathbb{E}(N_\infty^2)}\end{aligned}$$

By construction $H^{n_k} \to H$ in $L^2(M)$ which means that

$$\|H^{n_k} - H\|_M \to 0, \text{ as } k \to \infty.$$

By the Itô isometry

$$\mathbb{E}\left[\left((H^{n_k} - H) \cdot M\right)^2\right] = \|H^{n_k} - H\|_M^2 \to 0, \text{ as } k \to \infty,$$

that is $(H^{n_k} \cdot M)_\infty \to (H \cdot M)_\infty$ in L^2. Hence as N is an L^2 bounded martingale, the right hand side of the above expression tends to zero as $k \to \infty$. Similarly as $(H^{n_k} \cdot N)_\infty \to (H \cdot N)_\infty$ in L^2, we see also that

$$\mathbb{E}\left((H^{n_k} \cdot N)_\infty M_\infty\right) \to 0, \text{ as } k \to \infty.$$

Hence we can pass to the limit to obtain the result for H.

To prove the second part of the theorem, we shall first show that

$$\langle (H \cdot N), (K \cdot M)\rangle + \langle (K \cdot N), (H \cdot M)\rangle = 2HK\langle M, N\rangle.$$

By polarisation

$$\langle M, N\rangle = \frac{\langle M+N\rangle - \langle M-N\rangle}{4},$$

also

$$HK = \frac{(H+K)^2 - (H-K)^2}{4}.$$

Hence

$$2(HK \cdot \langle M, N\rangle) = \frac{1}{8}\Big(\left[(H+K)^2 - (H-K)^2\right] \cdot \{\langle \text{M+N}\rangle - \langle \text{M-N}\rangle\}\Big).$$

Now we use the result that $\langle (H \cdot M)\rangle = (H^2 \cdot \langle \text{M}\rangle)$ which has been proved previously in theorem (7.9(iii)), to see that

$$\begin{aligned}2(HK \cdot \langle M, N\rangle) =& \frac{1}{8}\Big(\langle (H+K) \cdot (M+N)\rangle - \langle (H+K) \cdot (M-N)\rangle \\ &- \langle (H-K) \cdot (M+N)\rangle + \langle (H-K) \cdot (M-N)\rangle\Big).\end{aligned}$$

Considering the first two terms

$$\begin{aligned}\langle (H+K) \cdot (M+N)\rangle - \langle (H+K) \cdot (M-N)\rangle =& \\ =&\langle (H+K) \cdot M + (H+K) \cdot N\rangle - \langle (H+K) \cdot M - (H+K) \cdot N\rangle \\ =&4\langle (H+K) \cdot M, (H+K) \cdot N\rangle \qquad \text{by polarisation} \\ =&4\left(\langle H \cdot M, H \cdot N\rangle + \langle H \cdot M, K \cdot N\rangle + \langle K \cdot M, H \cdot N\rangle + \langle K \cdot M, K \cdot N\rangle\right).\end{aligned}$$

Similarly for the second two terms

$$\begin{aligned}\langle (H-K) \cdot (M+N)\rangle - \langle (H-K) \cdot (M-N)\rangle =& \\ =&\langle (H-K) \cdot M + (H-K) \cdot N\rangle - \langle (H-K) \cdot M - (H-K) \cdot N\rangle \\ =&4\langle (H-K) \cdot M, (H-K) \cdot N\rangle \text{ by polarisation} \\ =&4\left(\langle H \cdot M, H \cdot N\rangle - \langle H \cdot M, K \cdot N\rangle - \langle K \cdot M, H \cdot N\rangle + \langle K \cdot M, K \cdot N\rangle\right).\end{aligned}$$

Adding these two together yields

$$2(HK \cdot \langle M,N\rangle) = \langle (H\cdot N),(K\cdot M)\rangle + \langle (K\cdot N),(H\cdot M)\rangle$$

Putting $K \equiv 1$ yields

$$2H\cdot\langle M,N\rangle = \langle H\cdot M,N\rangle + \langle M,H\cdot N\rangle.$$

So it suffices to prove that $\langle (H\cdot M),N\rangle = \langle M,(H\cdot N)\rangle$, which is equivalent to showing that

$$(H\cdot M)N - (H\cdot N)M$$

is a local martingale (from the definition of covariation process). By localisation it suffices to consider M and N bounded martingales, whence we must check that for all stopping times T,

$$\mathbb{E}\left((H\cdot M)_T N_T\right) = \mathbb{E}\left((H\cdot N)_T M_T\right),$$

but by the first part of the theorem

$$\mathbb{E}\left((H\cdot M)_\infty N_\infty\right) = \mathbb{E}\left((H\cdot N)_\infty M_\infty\right),$$

which is sufficient to establish the result, since

$$\begin{aligned}(H\cdot M)_T N_T =&(H\cdot M)_\infty^T N_\infty^T\\ (H\cdot N)_T M_T =&(H\cdot N)_\infty^T M_\infty^T\end{aligned}$$

□

Corollary 11.2.
Let N,M be continuous local martingales and H and K locally bounded previsible processes, then

$$\langle (H\cdot N),(K\cdot M)\rangle = (HK\cdot\langle N,M\rangle).$$

Proof
Note that the covariation is symmetric, hence

$$\begin{aligned}\langle (H\cdot N),(K\cdot M)\rangle =&(H\cdot\langle X,(K\cdot M)\rangle)\\ =&(H\cdot\langle (K\cdot M),X\rangle)\\ =&(HK\cdot\langle M,N\rangle).\end{aligned}$$

□

We can prove a stochastic calculus analogue of the usual integration by parts formula. However note that there is an extra term on the right hand side, the *covariation* of the processes X and Y. This is the first major difference we have seen between the Stochastic Integral and the usual Lebesgue Integral.

Before we can prove the general theorem, we need a lemma.

Lemma (Parts for Finite Variation Process and a Martingale) 11.3.
Let M be a bounded continuous martingale starting from zero, and V a bounded variation process starting from zero. Then

$$M_tV_t = \int_0^t M_s \mathrm{d}V_s + \int_0^t V_s \mathrm{d}M_s.$$

Proof
For n fixed, we can write

$$\begin{aligned} M_tV_t &= \sum_{k\geq 1} M_{k2^{-n}\wedge t}\left(V_{k2^{-n}\wedge t} - V_{(k-1)2^{-n}\wedge t}\right) + \sum_{k\geq 1} V_{(k-1)2^{-n}\wedge t}\left(M_{k2^{-n}\wedge t} - M_{(k-1)2^{-n}\wedge t}\right) \\ &= \sum_{k\geq 1} M_{k2^{-n}\wedge t}\left(V_{k2^{-n}\wedge t} - V_{(k-1)2^{-n}\wedge t}\right) + \int_0^t H_s^n \mathrm{d}M_s, \end{aligned}$$

where H^n is the previsible simple process

$$H_s^n = \sum_{k\geq 1} V_{k2^{-n}\wedge t} 1_{((k-1)2^{-n}\wedge t, k2^{-n}\wedge t]}(s).$$

These H^n are bounded and converge to V by the continuity of V, so as $n \to \infty$ the second term tends to

$$\int_0^t V_s \mathrm{d}M_s,$$

and by the Dominated Convergence Theorem for Lebesgue-Stieltjes integrals, the second term converges to

$$\int_0^t M_s \mathrm{d}V_s,$$

as $n \to \infty$. □

Theorem (Integration by Parts) 11.4.
For X and Y continuous semimartingales, then the following holds

$$X_tY_t - X_0Y_0 = \int_0^t X_s \mathrm{d}Y_s + \int_0^t Y_s \mathrm{d}X_s + \langle X, Y\rangle_t.$$

Proof
It is trivial to see that it suffices to prove the result for processes starting from zero. Hence let $X_t = M_t + A_t$ and $Y_t = N_t + B_t$ in Doob-Meyer decomposition, so N_t and M_t are continuous local martingales and A_t and B_t are finite variation processes, all starting from zero. By localisation we can consider the local martingales M and N to be bounded martingales and the FV processes A and B to have bounded variation. Hence by the usual (finite variation) theory

$$A_tB_t = \int_0^t A_s \mathrm{d}B_s + \int_0^t B_s \mathrm{d}A_s.$$

It only remains to prove for bounded martingales N and M starting from zero that

$$M_tN_t = \int_0^t M_s \mathrm{d}N_s + \int_0^t N_s \mathrm{d}M_s + \langle M, N\rangle_t.$$

This follows by application of polarisation to corollary (7.3) to the quadratic variation existence theorem. Hence

$$\begin{aligned}(M_t + A_t)(N_t + B_t) =& M_tN_t + M_tB_t + N_tA_t + A_tB_t \\ =& \int_0^t M_s \mathrm{d}N_s + \int_0^t N_s \mathrm{d}M_s + \langle M, N\rangle_t \\ &+ \int_0^t M_s \mathrm{d}B_s + \int_0^t B_s \mathrm{d}M_s + \int_0^t N_s \mathrm{d}A_s + \int_0^t A_s \mathrm{d}N_s \\ &+ \int_0^t A_s \mathrm{d}B_s + \int_0^t B_s \mathrm{d}A_s \\ =& \int_0^t (M_s + A_s)\mathrm{d}(N_s + B_s) + \int_0^t (N_s + B_s)\mathrm{d}(M_s + A_s) + \langle M, N\rangle.\end{aligned}$$

□

Reflect for a moment that this theorem is implying another useful closure property of continuous semimartingales. It implies that the product of two continuous semimartingales X_tY_t is a continuous semimartingale, since it can be written as a stochastic integrals with respect to continuous semimartingales and so it itself a continuous semimartingale.

Theorem (Itô's Formula) 11.5.
Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a twice continuously differentiable function, and also let $X = (X^1, X^2, \ldots, X^n)$ be a continuous semimartingale in $\mathbb{R}^n$. Then

$$f(X_t) - f(X_0) = \sum_{i=1}^n \int_0^t \frac{\partial f}{\partial x^i}(X_s)\mathrm{d}X_s^i + \frac{1}{2}\sum_{i,j=1}^n \int_0^t \frac{\partial f}{\partial x^i \partial x^j}(X_s)\mathrm{d}\langle X^i, X^j\rangle_s.$$

Proof

To prove Itô's formula; first consider the $n = 1$ case to simplify the notation. Then let $\mathcal{A}$ be the collection of C^2 (twice differentiable) functions $f : \mathbb{R} \to \mathbb{R}$ for which it holds. Clearly $\mathcal{A}$ is a vector space; in fact we shall show that it is also an algebra. To do this we must check that if f and g are in $\mathcal{A}$, then their product fg is also in $\mathcal{A}$. Let $F_t = f(X_t)$ and $G_t = g(X_t)$ be the associated semimartingales. From the integration by parts formula

$$F_tG_t - F_0G_0 = \int_0^t F_s \mathrm{d}G_s + \int_0^t G_s \mathrm{d}F_s + \langle F_s, G_s\rangle.$$

However since by assumption f and g are in $\mathcal{A}$, Itô's formula may be applied to them individually, so

$$\int_0^t F_s \mathrm{d}G_s = \int_0^t f(X_s)\frac{\partial f}{\partial x}(X_s)\mathrm{d}X_s.$$

Also by the Kunita-Watanabe formula extended to continuous local martingales we have

$$\langle F, G\rangle_t = \int_0^t f'(X_s)g'(X_s)\mathrm{d}\langle X, X\rangle_s.$$

Thus from the integration by parts,

$$\begin{aligned} F_tG_t - F_0G_0 &= \int_0^t F_s\mathrm{d}G_s + \int_0^t G_s\mathrm{d}F_s + \int_0^t f'(X_s)g'(X_s)\mathrm{d}\langle X, X\rangle_s, \\ &= \int_0^t \left(F_sg'(X_s) + f'(X_s)G_s\right)\mathrm{d}X_s \\ &+ \quad \frac{1}{2}\int_0^t \left(F_sg''(X_s) + 2f'g'(X_s) + f''(X_s)G_s\right)\mathrm{d}\langle \mathrm{M}\rangle_s. \end{aligned}$$

So this is just what Itô's formula states for fg and so Itô's formula also applies to fg; hence $fg \in \mathcal{A}$.

Since trivially $f(x) = x$ is in $\mathcal{A}$, then as $\mathcal{A}$ is an algebra, and a vector space this implies that $\mathcal{A}$ contains all polynomials. So to complete the proof, we must approximate f by polynomials (which we can do by standard functional analysis), and check that in the limit we obtain Itô's formula.

Introduce a sequence $U_n := \inf\{t : |X_t| + \langle \mathrm{X}\rangle_t > n\}$. Hence $\{U_n\}$ is a sequence of stopping times tending to infinity. Now we shall prove Itô's formula for twice continuously differentiable f restricted to the interval $[0, U_n]$, so we can consider X as a bounded martingale. Consider a polynomial sequence f_k approximating f, in the sense that for $r = 0, 1, 2$, $f_k^{(r)} \to f^{(r)}$ uniformly on a compact interval. We have proved that Itô's formula holds for all polynomial, so it holds for f_k and hence

$$f_k(X_{t\wedge U_n}) - f_k(X_0) = \int_0^{t\wedge U_n} f'(X_s)\mathrm{d}X_s + \frac{1}{2}\int_0^{t\wedge U_n} f_k''(X_s)\mathrm{d}\langle \mathrm{X}\rangle_s.$$

Let the continuous semimartingale X have Doob-Meyer decomposition

$$X_t = X_0 + M_t + A_t,$$

where M is a continuous local martingale and A is a finite variation process. We can rewrite the above as

$$f_k(X_{t\wedge U_n}) - f_k(X_0) = \int_0^{t\wedge U_n} f'(X_s)\mathrm{d}M_s + \int_0^{t\wedge U_n} f'(X_s)\mathrm{d}A_s + \frac{1}{2}\int_0^{t\wedge U_n} f_k''(X_s)\mathrm{d}\langle \mathrm{M}\rangle_s.$$

since $\langle \mathrm{X}\rangle = \langle \mathrm{M}\rangle$. On $(0, U_n]$ the process $|X|$ is uniformly bounded by n, so for $r = 0, 1, 2$ from the convergence (which is uniform on the compact interval $[0, U_n]$) we obtain

$$\sup_{|x|\le n} \left|f_k^{(r)} - f^{(r)}\right| \to 0 \text{ as } k \to \infty$$

And from the Itô isometry we get the required convergence.

$\square$

11.1. Applications of Itô's Formula

Let B_t be a standard Brownian motion; the aim of this example is to establish that:

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

This example gives a nice simple demonstration that all our hard work has achieved something. The result isn't the same as that which would be given by the 'logical' extension of the usual integration rules.

To prove this we apply Itô's formula to the function $f(x) = x^2$. We obtain

$$f(B_t) - f(B_0) = \int_0^t \frac{\partial f}{\partial x}(B_s)\mathrm{d}B_s + \frac{1}{2}\int_0^t \frac{\partial^2 f}{\partial x^2}(B_s)\mathrm{d}\langle B, B\rangle_s,$$

noting that $B_0 = 0$ for a standard Brownian Motion we see that

$$B_t^2 = 2\int_0^t B_s \mathrm{d}B_s + \frac{1}{2}2\mathrm{d}s,$$

whence we derive that

$$\int_0^t B_s \mathrm{d}B_s = \frac{B_t^2}{2} - \frac{t}{2}.$$

For those who have read the foregoing material carefully, there are grounds to complain that there are simpler ways to establish this result, notably by consideration of the definition of the quadratic variation process. However the point of this example was to show how Itô's formula can help in the actual evaluation of stochastic integrals; not to establish a totally new result.

11.2. Exponential Martingales

Exponential martingales play an important part in the theory. Suppose X is a continuous semimartingale starting from zero. Define:

$$Z_t = \exp\left(X_t - \frac{1}{2}\langle \mathrm{X}\rangle_t\right).$$

This Z_t is called the **exponential semimartingale** associated with X_t, and it is the solution of the stochastic differential equation

$$\mathrm{d}Z_t = Z_t \mathrm{d}X_t,$$

that is

$$Z_t = 1 + \int_0^t Z_s \mathrm{d}X_s,$$

so clearly if X is a continuous local martingale, i.e. $X \in \mathcal{M}_{loc}^c$ then this implies, by the stability property of stochastic integration, that $Z \in \mathcal{M}_{loc}^c$ also.

Proof
For existence, apply Itô's formula to $f(x) = \exp(x)$ to obtain

$$\mathrm{d}\left(\exp(Y_t)\right) = \exp(Y_t)\mathrm{d}Y_t + \frac{1}{2}\exp(Y_t)\mathrm{d}\langle Y, Y\rangle_t.$$

Hence

$$\begin{aligned}
\mathrm{d}\left(\exp(X_t - \frac{1}{2}\langle \mathrm{X}\rangle_t)\right) =& \exp(X_t - \frac{1}{2}\langle \mathrm{X}\rangle_t)\mathrm{d}\left(X_t - \frac{1}{2}\langle \mathrm{X}\rangle_t\right) \\
&+ \frac{1}{2}\exp\left(X_t - \frac{1}{2}\langle \mathrm{X}\rangle_t\right)\mathrm{d}\left\langle X_t - \frac{1}{2}\langle \mathrm{X}\rangle_t, X_t - \frac{1}{2}\langle \mathrm{X}\rangle_t\right\rangle \\
=& \exp(X_t - \frac{1}{2}\langle \mathrm{X}\rangle_t)\mathrm{d}X_t - \frac{1}{2}\exp\left(X_t - \frac{1}{2}\langle \mathrm{X}\rangle_t\right)\mathrm{d}\langle \mathrm{X}\rangle_t \\
&+ \frac{1}{2}\exp\left(X_t - \frac{1}{2}\langle \mathrm{X}\rangle_t\right)\mathrm{d}\langle \mathrm{X}\rangle_t \\
=& Z_t \mathrm{d}X_t
\end{aligned}$$

Hence Z_t certainly solves the equation. Now to check uniqueness, define

$$Y_t = \exp\left(-X_t + \frac{1}{2}\langle \mathrm{X}\rangle_t\right),$$

we wish to show that for every solution of the Stochastic Differential Equation $Z_t Y_t$ is a constant. By a similar application of Itô's formula

$$\mathrm{d}Y_t = -Y_t \mathrm{d}X_t + Y_t \mathrm{d}\langle \mathrm{X}\rangle_t,$$

whence by integration by parts (alternatively consider Itô applied to $f(x, y) = xy$),

$$\begin{aligned}
\mathrm{d}(Z_t Y_t) =& Z_t \mathrm{d}Y_t + Y_t \mathrm{d}Z_t + \langle Z, Y\rangle_t, \\
=& Z_t(-Y_t \mathrm{d}X_t + Y_t \mathrm{d}\langle \mathrm{X}\rangle_t) + Y_t Z_t \mathrm{d}X_t + (-Y_t Z_t)\mathrm{d}\langle \mathrm{X}\rangle_t, \\
=& 0.
\end{aligned}$$

So $Z_t Y_t$ is a constant, hence the unique solution of the stochastic differential equation $\mathrm{d}Z_t = Z_t \mathrm{d}X_t$, with $Z_0 = 1$, is

$$Z_t = \exp\left(X_t - \frac{1}{2}\langle \mathrm{X}\rangle_t\right).$$

□

Example
Let $X_t = \lambda B_t$, for an arbitrary scalar λ. Clearly X_t is a continuous local martingale, so the associated exponential martingale is

$$M_t = \exp\left(\lambda B_t - \frac{1}{2}\lambda^2 t\right).$$

It is clear that the exponential semimartingale of a real valued martingale must be non-negative, and thus by application of Fatou's lemma we can show that it is a supermartingale, thus $\mathbb{E}(M_t) \leq 1$ for all t.

Theorem 11.6.
Let M be a non-negative local martingale, such that $\mathbb{E}M_0 < \infty$ then M is a supermartingale.

Proof
Let T_n be a reducing sequence for $M_n - M_0$, then for $t > s \geq 0$,

$$\begin{aligned}\mathbb{E}(M_{t\wedge T_n}|\mathcal{F}_s) &= \mathbb{E}(M_0|\mathcal{F}_s) + \mathbb{E}(M_{t\wedge T_n} - M_0|\mathcal{F}_s)\\ &= M_0 + M_{s\wedge T_n} - M_0 = M_{s\wedge T_n}.\end{aligned}$$

Now by application of the conditional form of Fatou's lemma

$$\mathbb{E}(M_t|F_s) = \mathbb{E}(\liminf_{n\to\infty} M_{t\wedge T_n}|\mathcal{F}_s) \leq \liminf_{n\to\infty} \mathbb{E}(M_{t\wedge T_n}|\mathcal{F}_s) = M_s.$$

Thus M is a supermartingale as required. □

Checking for a Martingale
The general exponential semimartingale is useful, but in many applications, not the least of which will be Girsanov's formula an actual Martingale will be needed. How do we go about checking if a local martingale is a martingale anyway? It will turn out that there are various methods, some of which crop up in the section on filtration. First I shall present a simple example and then prove a more general theorem. A common error is to think that it is sufficient to show that a local martingale is locally bounded in L^2 to show that it is a martingale – this is not sufficient as should be made clear by this example!

The exponential of λB_t
We continue our example from above. Let $M_t = \exp(\lambda B_t - 1/2\lambda^2 t)$ be the exponential semimartingale associated with a standard Brownian Motion B_t, starting from zero. By the previous argument we know that

$$M_t = 1 + \int_0^t \lambda M_s \mathrm{d}B_s,$$

hence M is a local martingale. Fix T a constant time, which is of course a stopping time, then B^T is an L^2 bounded martingale ($\mathbb{E}(B_t^T)^2 = t \wedge T \leq T$). We then show that M^T is in $L^2(B^T)$ as follows

$$\begin{aligned}\|M^T\|_B &= \mathbb{E}\left(\int_0^T M_s^2 \mathrm{d}\langle B\rangle_s\right), \\ &= \mathbb{E}\left(\int_0^T M_s^2 \mathrm{d}s\right), \\ &\leq \mathbb{E}\left(\int_0^T \exp(2\lambda B_s)\mathrm{d}s\right), \\ &= \int_0^T \mathbb{E}\left(\exp(2\lambda B_s)\right)\mathrm{d}s, \\ &= \int_0^T \exp(2\lambda^2 s^2)\mathrm{d}s < \infty.\end{aligned}$$

In the final equality we use that fact that B_s is distributed as $N(0,s)$, and we use the characteristic function of the normal distribution. Thus by the integration isometry theorem we have that $(M^T \cdot B)_t$ is an L^2 bounded martingale. Thus for every such T, Z^T is an L^2 bounded martingale, which implies that M is a martingale.

The Exponential Martingale Inequality

We have seen a specific proof that a certain (important) exponential martingale is a true martingale, we now show a more general argument.

Theorem 11.7.

Let M be a continuous local martingale, starting from zero. Suppose for each t, there exists a constant K_t such that $\langle M\rangle_t < \infty$ a.s., then for every t, and every $y > 0$,

$$\mathbb{P}\left[\sup_{0\leq s\leq t} M_s > y\right] \leq \exp(-y^2/2K_t).$$

Furthermore, the associated exponential semimartingale $Z_t = \exp(\theta M_t - 1/2\theta^2\langle M\rangle_t)$ is a true martingale.

Proof

We have already noted that the exponential semimartingale Z_t is a supermartingale, so $\mathbb{E}(Z_t) \leq 1$ for all $t \geq 0$, and hence for $\theta > 0$ and $y > 0$,

$$\begin{aligned}\mathbb{P}\left[\sup_{s\leq t} M_s > y\right] &\leq \mathbb{P}\left[\sup_{s\leq t} Z_s > \exp(\theta y - 1/2\theta^2 K_t)\right], \\ &\leq \exp(-\theta y + 1/2\theta^2 K_t).\end{aligned}$$

Optimizing over θ now gives the desired result. For the second part, we establish the following bound

$$\mathbb{E}\left(\sup_{0\le s\le t} Z_s\right) \le \mathbb{E}\left(\exp\left[\sup_{0\le s\le t} Z_s\right]\right),$$

$$\le \int_0^\infty \mathbb{P}\left[\sup 0 \le s \le tZ_s \ge \log\lambda\right] \mathrm{d}\lambda, \qquad (*)$$

$$\le 1 + \int_1^\infty \exp\left(-(\log\lambda)^2/2K_t\right) \mathrm{d}\lambda < \infty.$$

We have previously noted that Z is a local martingale; let T_n be a reducing sequence for Z, hence Z^{T_n} is a martingale, hence

$$\mathbb{E}\left[Z_{t\wedge T_n}|\mathcal{F}_s\right] = Z_{s\wedge T_n}. \qquad (**)$$

We note that Z_t is dominated by $\exp(\theta \sup_{0\le s\le t} Z_s)$, and thus by our bound we can apply the dominated convergence theorem to $(**)$ as $n\to\infty$ to establish that Z is a true martingale. □

Corollary 11.8.
For all $\epsilon, \delta > 0$,

$$\mathbb{P}\left[\sup_{t\ge 0} M_t \ge \epsilon \ \&\ \langle M\rangle_\infty \le \delta\right] \le \exp(-\epsilon^2/2\delta).$$

Proof
Set $T = \inf\{t\ge 0 : M_t \ge \epsilon\}$, the conditions of the previous theorem now apply to M^T, with $K_t = \epsilon$. □

From this corollary, it is clear that if H is any bounded previsible process, then

$$\exp\left(\int_0^t H_s \mathrm{d}B_s - \frac{1}{2}\int_0^t |H_s|^2 \mathrm{d}s\right)$$

is a true martingale, since this is the exponential semimartingale associated with the process $\int H\mathrm{d}B$.

Corollary 11.9.
If the bounds K_t on $\langle M\rangle$ are uniform, that is if $K_t \le C$ for all t, then the exponential martingale is Uniformly Integrable. We shall use the useful result

$$\int_0^\infty \mathbb{P}(X\ge \log\lambda)\mathrm{d}\lambda = \int_0^\infty \mathbb{E}\left(1_{e^X\ge\lambda}\right)\mathrm{d}\lambda = \mathbb{E}\int_0^\infty 1_{e^X\ge\lambda}\mathrm{d}\lambda = \mathbb{E}(e^X).$$

Proof
Note that the bound $(*)$ extends to a uniform bound

$$\mathbb{E}\left(\sup_{t\ge 0} Z_t\right) \le 1 + \int_1^\infty \exp\left(-(\log\lambda)^2/2C\right)\mathrm{d}\lambda < \infty.$$

Hence Z is bounded in L^∞ and thus a uniformly integrable martingale. □

12. Lévy Characterisation of Brownian Motion

A very useful result can be proved using the Itô calculus about the characterisation of Brownian Motion.

Theorem 12.1.
Let $\{B^i\}_{t\geq 0}$ be continuous local martingales starting from zero for $i = 1, \ldots, n$. Then $B_t = (B_t^1, \ldots, B_t^n)$ is a Brownian motion with respect to $(\Omega, \mathcal{F}, \mathbb{P})$ adapted to the filtration $\mathcal{F}_t$, if and only iff

$$\langle B^i, B^j \rangle_t = \delta_{ij} t \quad \forall i, j \in \{1, \ldots, n\}.$$

Proof
In these circumstances it follows that the statement B_t is a Brownian Motion is by definition equivalent to stating that $B_t - B_s$ is independent of $\mathcal{F}_s$ and is distributed normally with mean zero and covariance matrix $(t-s)I$.

Clearly if B_t is a Brownian motion then the covariation result follows trivially from the definitions. Now to establish the converse, we assume $\langle B^i, B^j \rangle_t = \delta_{ij} t$ for $i, j \in \{1, \ldots, n\}$, and shall prove B_t is a Brownian Motion.

Observe that for fixed $\theta \in \mathbb{R}^n$ we can define M_t^θ by

$$M_t^\theta := f(B_t, t) = \exp\left(i(\theta, x) + \frac{1}{2}|\theta|^2 t\right).$$

By application of Itô's formula to f we obtain (in differential form using the Einstein summation convention)

$$\begin{aligned}
\mathrm{d}\left(f(B_t, t)\right) =& \frac{\partial f}{\partial x^j}(B_t, t)\mathrm{d}B_t^j + \frac{\partial f}{\partial t}(B_t, t)\mathrm{d}t + \frac{1}{2}\frac{\partial^2 f}{\partial x^j \partial x^k}(B_t, t)\mathrm{d}\langle B^j, B^k \rangle_t, \\
=& i\theta_j f(B_t, t)\mathrm{d}B_t^j + \frac{1}{2}|\theta|^2 f(B_t, t)\mathrm{d}t - \frac{1}{2}\theta_j\theta_k\delta_{jk} f(B_t, t)\mathrm{d}t \\
=& i\theta_j f(B_t, t)\mathrm{d}B_t^j.
\end{aligned}$$

Hence

$$M_t^\theta = 1 + \int_0^t \mathrm{d}(f(B_t, t)),$$

and is a sum of stochastic integrals with respect to continuous local martingales and is hence itself a continuous local martingale. But note that for each t,

$$|M_t^\theta| = \left(e^{\frac{1}{2}|\theta|^2 t}\right) < \infty$$

Hence for any fixed time t_0, $(M^{t_0})_t$ satisfies

$$|(M^{t_0})_t| \leq |(M^{t_0})_\infty| < \infty,$$

and so is a bounded local martingale; hence $(M^{t_0})_t)$ is a martingale. Hence M^{t_0} is a genuine martingale. Thus for $0 \leq s < t$ we have

$$\mathbb{E}\left(\exp\left(i(\theta, B_t - B_s)\right) | \mathcal{F}_s\right) = \exp\left(-\frac{1}{2}(t-s)\,|\theta|^2\right) \quad \text{a.s.}$$

However this is just the characteristic function of a normal random variable following $N(0, t-s)$; so by the Lévy character theorem $B_t - B_s$ is a $N(0, t-s)$ random variable. $\square$

13. Time Change of Brownian Motion

This result is one of frequent application, essentially it tells us that any continuous local martingale starting from zero, can be written as a time change of Brownian motion. So modulo a time change a Brownian motion is the most general kind of continuous local martingale.

Proposition 13.1.
Let M be a continuous local martingale starting from zero, such that $\langle M\rangle_t \to \infty$ as $t \to \infty$. Then define

$$\tau_s := \inf\{t > 0 : \langle M\rangle_t > s\}.$$

Then define

$$\tilde{A}_s := M_{\tau_s}.$$

(i) This τ_s is an $\mathcal{F}$ stopping time.
(ii) $\langle M\rangle_{\tau_s} = s$.
(iii) The local martingale M can be written as a time change of Brownian Motion as $M_t = B_{\langle M\rangle_t}$. Moreover the process $\tilde{A}_s$ is an $\tilde{\mathcal{F}}_s$ adapted Brownian Motion, where $\tilde{\mathcal{F}}_s$ is the time-changed σ algebra i.e. $\tilde{\mathcal{F}}_s = \mathcal{F}_{\tau_s}$.

Proof
We may assume that the map $t \mapsto \langle M\rangle_t$ is strictly increasing. Note that the map $s \mapsto \tau_s$ is the inverse to $t \mapsto \langle M\rangle_t$. Hence the results (i),(ii) and (iii).

Define

$$T_n := \inf\{t : |M|_t > n\},$$
$$[U_n := \langle M\rangle_{T_n}.$$

Note that from these definitions

$$\begin{aligned}\tau_{t\wedge U_n} &= \inf\{s > 0 : \langle M\rangle_s > t \wedge U_n\}\\ &= \inf\{s > 0 : \langle M\rangle_s > t \wedge \langle M\rangle_{T_n}\}\\ &= T_n \wedge \tau_t\end{aligned}$$

So

$$\tilde{A}_s^{U_n} = \tilde{A}_{s\wedge U_n} = M_{\tau_t}^{T_n}.$$

Now note that U_n is an $\tilde{\mathcal{F}}_t$ stopping time, since consider

$$\Lambda \equiv \{U_n \le t\} \equiv \{\langle M\rangle_{T_n} \le t\} \equiv \{T_n \le \tau_t\},$$

the latter event is clearly $\mathcal{F}_{\tau_t}$ measurable i.e. it is $\tilde{F}_t$ measurable, so U_n is a $\tilde{\mathcal{F}}_t$ stopping time. We may now apply the optional stopping theorem to the UI martingale M^{T_n}, which yields

$$\begin{aligned}\mathbb{E}\left(\tilde{A}_t^{U_n}|\mathcal{F}_s\right) &= \mathbb{E}\left(\tilde{A}_{t\wedge U_n}|\tilde{\mathcal{F}}_s\right) = \mathbb{E}\left(M_{\tau_t}^{T_n}|\tilde{\mathcal{F}}_s\right)\\ &= \mathbb{E}\left(M_{\tau_t}^{T_n}|\mathcal{F}_{\tau_s}\right) = M_{\tau_s}^{T_n} = \tilde{A}_s^{U_n}.\end{aligned}$$

So $\tilde{A}_t$ is a $\tilde{\mathcal{F}}_t$ local martingale. But we also know that $(M^2 - \langle \mathrm{M} \rangle)^{T_n}$ is a UI martingale, since M^{T_n} is a UI martingale. By the optional stopping theorem, for $0 < r < s$ we have

$$\begin{aligned}
\mathbb{E}\left(\tilde{A}^2_{s\wedge U_n} - (s \wedge U_n)|\tilde{\mathcal{F}}_r\right) = \mathbb{E}\left(\left(\left(M^{T_n}_{\tau_s}\right)^2 - \langle \mathrm{M} \rangle_{\tau_s \wedge T_n}\right)|\mathcal{F}_{\tau_r}\right) \\
=&\mathbb{E}\left(\left(M^2_{\tau_s} - \langle \mathrm{M} \rangle_{\tau_s}\right)^{T_n}|\mathcal{F}_{\tau_r}\right) = \left(M^2_{\tau_r} - \langle \mathrm{M} \rangle_{\tau_r}\right)^{T_n} \\
=&\tilde{A}^2_{r\wedge U_n} - (r \wedge U_n).
\end{aligned}$$

Hence $\tilde{A}^2 - t$ is a $\tilde{\mathcal{F}}_t$ local martingale. Before we can apply Lévy's characterisation theorem we must check that $\tilde{A}$ is continuous; that is we must check that for almost every ω that M is constant on each interval of constancy of $\langle \mathrm{M} \rangle$. By localisation it suffices to consider M a square integrable martingale, now let q be a positive rational, and define

$$S_q := \inf\{t > q : \langle \mathrm{M} \rangle_t > \langle \mathrm{M} \rangle_q\},$$

then it is enough to show that M is constant on $[q, S_q)$. But $M^2 - \langle \mathrm{M} \rangle$ is a martingale, hence

$$\begin{aligned}
\mathbb{E}\left[\left(M^2_{S_q} - \langle \mathrm{M} \rangle_{S_q}\right)^2|\mathcal{F}_q\right] =&M^2_q - \langle \mathrm{M} \rangle_q \\
=&M^2_q - \langle \mathrm{M} \rangle_{S_q}, \text{ as } \langle \mathrm{M} \rangle_q = \langle \mathrm{M} \rangle_{S_q}.
\end{aligned}$$

Hence

$$\mathbb{E}\left[\left(M_{S_q} - M_q\right)^2|\mathcal{F}_q\right] = 0,$$

which establishes that $\tilde{A}$ is continuous.

Thus $\tilde{A}$ is a continuous $\tilde{\mathcal{F}}_t$ adapted martingale with $\langle \tilde{A}_s \rangle_s = s$ and so by the Lévy characterisation theorem $\tilde{A}_s$ is a Brownian Motion. □

13.1. Gaussian Martingales

The time change of Brownian Motion can be used to prove the following useful theorem.

Theorem 13.2.

If M is a continuous local martingale starting from zero, and $\langle M \rangle_t$ is deterministic, that is if we can find a deterministic function f taking values in the non-negative real numbers such that $\langle \mathrm{M} \rangle_t = f(t)$ a.e., then M is a Gaussian Martingale (i.e. M_t has a Gaussian distribution for almost all t).

Proof

Note that by the time change of Brownian Motion theorem, we can write M_t as a time change of Brownian Motion through

$$M_t = B_{\langle \mathrm{M} \rangle_t},$$

where B is a standard Brownian Motion. By hypothesis $\langle \mathrm{M} \rangle_t = f(t)$, a deterministic function for almost all t, hence for almost all t,

$$M_t = B_{f(t)},$$

but the right hand side is a Gaussian random variable following $N(0, f(t))$. Hence M is a Gaussian Martingale, and at time t it has distribution given by $N(0, \langle \mathrm{M} \rangle_t)$. $\square$

As a corollary consider the stochastic integral of a purely deterministic function with respect to a Brownian motion.

Corollary 13.3.
Let $g(t)$ be a deterministic function of t, then M defined by

$$M_t := \int_0^t f(s) \mathrm{d}B_s,$$

satisfies

$$M_t \sim N\left(0, \int_0^t |f(s)|^2 \mathrm{d}s\right).$$

Proof
From the definition of M via a stochastic integral with respect to a continuous martingale, it is clear that M is a continuous local martingale, and by the Kunita-Watanabe result, the quadratic variation of M is given by

$$\langle \mathrm{M} \rangle_t = \int_0^t |f(s)| \mathrm{d}s,$$

hence the result follows. $\square$

This result can also be established directly in a fashion which is very similar to the proof of the Lévy characterisation theorem. Consider Z defined via

$$Z_t = \exp\left(i\theta M_t + \frac{1}{2}\theta^2 \langle \mathrm{M} \rangle_t\right),$$

as in the Lévy characterisation proof, we see that this is a continuous local martingale, and by boundedness furthermore is a martingale, and hence

$$\mathbb{E}(Z_0) = \mathbb{E}(Z_t),$$

whence

$$\mathbb{E}\left(\exp(i\theta M_t)\right) = \mathbb{E}\left(\exp\left(-\frac{1}{2}\theta^2 \int_0^t f(s)^2 \mathrm{d}s\right)\right)$$

which is the characteristic function of the appropriate normal distribution.

14. Girsanov's Theorem

Girsanov's theorem is an element of stochastic calculus which does not have an analogue in standard calculus.

14.1. Change of measure

When we wish to compare two measures $\mathbb{P}$ and $\mathbb{Q}$, we don't want either of them simply to throw information away; since when they are positive they can be related by the *Radon-Nikodym* derivative; this motivates the following definition of *equivalence* of two measures.

Definition 14.1.
Two measures $\mathbb{P}$ and $\mathbb{Q}$ are said to be equivalent if they operate on the same sample space, and if A is any event in the sample space then

$$\mathbb{P}(A) > 0 \Leftrightarrow \mathbb{Q}(A) > 0.$$

In other words $\mathbb{P}$ is absolutely continuous with respect to $\mathbb{Q}$ and $\mathbb{Q}$ is absolutely continuous with respect to $\mathbb{P}$.

Theorem 14.2.
If $\mathbb{P}$ and $\mathbb{Q}$ are equivalent measures, and X_t is an $\mathcal{F}_t$-adapted process then the following results hold

$$\mathbb{E}_{\mathbb{Q}}(X_t) = \mathbb{E}_{\mathbb{P}}\left(\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} X_t\right),$$

$$\mathbb{E}_{\mathbb{Q}}(X_t|\mathcal{F}_s) = L_s^{-1}\mathbb{E}_{\mathbb{P}}\left(L_t X_t|\mathcal{F}_s\right),$$

where

$$L_s = \mathbb{E}_{\mathbb{P}}\left(\left.\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\right|\mathcal{F}_s\right).$$

Here L_t is the Radon-Nikodym derivative of $\mathbb{Q}$ with respect to $\mathbb{P}$. The first result basically shows that this is a martingale, and the second is a continuous time version of Bayes theorem.

Proof
The first part is basically the statement that the Radon-Nikodym derivative is a martingale. This follows because the measures $\mathbb{P}$ and $\mathbb{Q}$ are equivalent, but this will not be proved in detail here. Let Y be an $\mathcal{F}_t$ measurable random variable, such that $\mathbb{E}_{\mathbb{Q}}(|Y|) < \infty$. We shall prove that

$$\mathbb{E}_Q(Y|\mathcal{F}_s) = \frac{1}{L_s}\mathbb{E}_{\mathbb{P}}\left[Y L_t|\mathcal{F}_s\right] \text{ a.s. } (\mathbb{P} \text{ and } \mathbb{Q}).$$

Then for any $A \in \mathcal{F}_s$, using the definition of conditional expectation we have that

$$\begin{aligned}\mathbb{E}_{\mathbb{Q}}\left(1_A \frac{1}{L_s}\mathbb{E}_{\mathbb{P}}\left[Y L_t|\mathcal{F}_s\right]\right) =& \mathbb{E}_{\mathbb{P}}\left(1_A \mathbb{E}_{\mathbb{P}}\left[Y L_t|\mathcal{F}_s\right]\right)\\ =& \mathbb{E}_{\mathbb{P}}\left[1_A Y L_t\right] = \mathbb{E}_Q\left[1_A Y\right].\end{aligned}$$

Substituting $Y = X_t$ gives the desired result. □

Theorem (Girsanov).
Let M be a continuous local martingale, and let Z be the associated exponential martingale

$$Z_t = \exp\left(M_t - \frac{1}{2}\langle M\rangle_t\right).$$

If Z is uniformly integrable, then a new measure $\mathbb{Q}$, equivalent to $\mathbb{P}$ may be defined by

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = Z_\infty.$$

Then if X is a continuous $\mathbb{P}$ local martingale, $X - \langle X, M\rangle$ is a $\mathbb{Q}$ local martingale.

Proof
Since Z_∞ exists a.s. it defines a uniformly integrable martingale (the exponential martingale), a version of which is given by $Z_t = \mathbb{E}(Z_\infty|\mathcal{F}_t)$. Hence $\mathbb{Q}$ constructed thus is a probability measure which is equivalent to $\mathbb{P}$. Now consider X, a $\mathbb{P}$ local martingale. Define a sequence of stopping times which tend to infinity via

$$T_n := \inf\{t \geq 0 : |X_t| \geq n, \text{ or } |\langle X, M\rangle_t| \geq n\}.$$

Now consider the process Y defined via

$$Y := X^{T_n} - \langle X^{T_n}, M\rangle.$$

By Itô's formula for $0 \leq t \leq T_n$, remembering that $\mathrm{d}Z_t = Z_t\mathrm{d}M_t$ as Z is the exponential martingale associated with M,

$$\begin{aligned}
\mathrm{d}(Z_tY_t) =& Z_t\mathrm{d}Y_t + Y_t\mathrm{d}Z_t + \mathrm{d}\langle Z, Y\rangle_t\\
=& Z_t(\mathrm{d}X_t - \mathrm{d}\langle X, M\rangle_t) + Y_tZ_t\mathrm{d}M_t + \mathrm{d}\langle Z, Y\rangle_t\\
=& Z_t(\mathrm{d}X_t - \mathrm{d}\langle X, M\rangle_t) + (X_t - \langle X, M\rangle_t)Z_t\mathrm{d}M_t + Z_t\mathrm{d}\langle X, M\rangle_t\\
=& (X_t - \langle X, M\rangle_t)Z_t\mathrm{d}M_t + Z_t\mathrm{d}X_t
\end{aligned}$$

Where the result $\mathrm{d}\langle Z, Y\rangle_t = Z_t\mathrm{d}\langle X, M\rangle_t$ follows from the Kunita-Watanabe theorem. Hence ZY is a $\mathbb{P}$-local martingale. But since Z is uniformly integrable, and Y is bounded (by construction of the stopping time T_n), hence ZY is a genuine $\mathbb{P}$-martingale. Hence for $s < t$ and $A \in \mathcal{F}_s$, we have

$$\mathbb{E}_\mathbb{Q}\left[(Y_t - Y_s)1_A\right] = \mathbb{E}\left[Z_\infty(Y_t - Y_s)1_A\right] = \mathbb{E}\left[(Z_tY_t - Z_sY_s)1_A\right] = 0,$$

hence Y is a $\mathbb{Q}$ martingale. Thus $X - \langle X, M\rangle$ is a $\mathbb{Q}$ *local* martingale, since T_n is a reducing sequence such that $(X - \langle X, M\rangle)^{T_n}$ is a $\mathbb{Q}$-martingale, and $T_n \uparrow \infty$ as $n \to \infty$. □

Corollary 14.3.
Let W_t be a $\mathbb{P}$ Brownian motion, then $\tilde{W}_t := W_t - \langle W, M\rangle_t$ is a $\mathbb{Q}$ Brownian motion.

Proof
Use Lévy's characterisation of Brownian motion to see that since $\tilde{W}_t$ is continuous and $\langle \tilde{W}, \tilde{W}\rangle_t = \langle W, W\rangle_t = t$, since W_t is a $\mathbb{P}$ Brownian motion, then $\tilde{W}$ is a $\mathbb{Q}$ Brownian motion. □

15. Brownian Martingale Representation Theorem

The following theorem has many applications, for example in the rigorous study of mathematical finance, even though the result is purely an existence theorem. The Malliavin calculus offers methods by which the process $\mathbf{H}$ in the following theorem can be stated explicitly, but these methods are beyond the scope of these notes!

Theorem 15.1.
Let $\mathbf{B}_t$ be a Brownian Motion on $\mathbb{R}^n$ and $\mathcal{G}_t$ is the usual augmentation of the filtration generated by $\mathbf{B}_t$. If Y is L^2 integrable and is measurable with respect to $\mathcal{G}_\infty$ then there exists a previsible $\mathcal{G}_t$ measurable process $\mathbf{H}_s$ uniquely defined up to evanescence such that

$$\mathbb{E}(Y|\mathcal{G}_t) = \mathbb{E}(Y) + \int_0^t \mathbf{H}_s \cdot \mathrm{d}\mathbf{B}_s \tag{1}$$

The proof of this result can seem hard if you are not familiar with functional analysis style arguments. The outline of the proof is to describe all Ys which are $\mathcal{G}_T$ measurable which cannot be represented in the form (1) as belonging to the orthogonal complement of a space. Then we show that for Z in this orthogonal complement that $\mathbb{E}(ZX) = 0$ for all X in a large space of $\mathcal{G}_T$ measurable functions. Finally we show that this space is sufficiently big that actually we have proved this for all $\mathcal{G}_T$ measurable functions, which includes Z so $\mathbb{E}(Z^2) = 0$ and hence $Z = 0$ a.s. and we are done!

Proof
Without loss of generality prove the result in the case $\mathbb{E}Y = 0$ where Y is L^2 integrable and measurable with respect to $\mathcal{G}_T$ for some constant $T > 0$.

Define the space

$$L^2_T(B) = \left\{ \mathbf{H} : \mathbf{H} \text{ is } \mathcal{G}_t \text{ previsible and } \mathbb{E}\left(\int_0^T \|\mathbf{H}_s\|^2 \mathrm{d}s \right) < \infty \right\}$$

Consider the stochastic integral map

$$I : L^2_T(\mathbf{B}) \to L^2(\mathcal{G}_T)$$

defined by

$$I(\mathbf{H}) = \int_0^T \mathbf{H}_s \cdot \mathrm{d}\mathbf{B}_s.$$

As a consequence of the Itô isometry theorem, this map is an isometry. Hence the image V under I of the Hilbert space $L^2_T(\mathbf{B})$ is complete and hence a closed subspace of $L^2_0(\mathcal{G}_T) = \{H \in L^2(\mathcal{G}_T) : \mathbb{E}H = 0\}$. The theorem will be proved if we can establish that the image is the whole space.

We follow the usual approach in such proofs; consider the orthogonal complement of V in $L^2_0(\mathcal{G}_T)$ and we aim to show that every element of this orthogonal complement is zero. Suppose that Z is in the orthogonal complement of $L^2_0(\mathcal{G}_T)$, thus

$$\mathbb{E}(ZX) = 0 \text{ for all } X \in L^2_0(\mathcal{G}_T) \tag{2}$$

We can define $Z_t = \mathbb{E}(Z|\mathcal{G}_t)$ which is an L^2 bounded martingale. We know that the sigma field $\mathcal{G}_0$ is trivial by the 0-1 law therefore

$$Z_0 = \mathbb{E}(Z|\mathcal{G}_0) = \mathbb{E}Z = 0.$$

Let $\mathbf{H} \in L^2(\mathbf{B})$ let $N_T = I(\mathbf{H})$; we may define $N_t = \mathbb{E}(N_T|\mathcal{G}_t)$ for $0 \leq t \leq T$. Clearly $N_T \in V$ as it is the image under I of some $\mathbf{H}$.

Let S be a stopping time such that $S \leq T$ then

$$N_S = \mathbb{E}(N_T|\mathcal{G}_S) = \mathbb{E}\left(\int_0^S \mathbf{H}_s \cdot \mathrm{d}\mathbf{B}_s + \int_S^T \mathbf{H}_s \cdot \mathrm{d}\mathbf{B}_s \middle| \mathcal{G}_S \right) = I(\mathbf{H}1_{(0,S]}),$$

so consequently $N_S \in V$. The orthogonality relation (2) then implies that $\mathbb{E}(ZN_S) = 0$. Thus using the martingale property of Z,

$$\mathbb{E}(ZN_S) = \mathbb{E}(\mathbb{E}(ZN_S|\mathcal{G}_S)) = \mathbb{E}(N_S\mathbb{E}(Z|\mathcal{G}_S)) = \mathbb{E}(Z_S N_S) = 0$$

Since Z_T and N_T are square integrable, it follows that $Z_t N_t$ is a uniformly integrable martingale.

Since the stochastic exponential of a process may be written as

$$M_t = \mathcal{E}(i\boldsymbol{\theta} \cdot \mathbf{B}_t) = \exp\left(i\boldsymbol{\theta} \cdot \mathbf{B}_t + \frac{1}{2}|\boldsymbol{\theta}|^2 t \right) = \int_0^t iM_t\boldsymbol{\theta} \cdot \mathrm{d}\mathbf{B}_t,$$

such a process can be taken as $\mathbf{H} = i\boldsymbol{\theta} M_t$ in the definition of N_T and by the foregoing argument we see that $Z_t M_t$ is a martingale. Thus

$$Z_s M_s = \mathbb{E}(Z_t M_t|\mathcal{G}_s) = \mathbb{E}\left(Z_t \exp\left(i\boldsymbol{\theta} \cdot \mathbf{B}_t + \frac{1}{2}|\boldsymbol{\theta}|^2 t \right) \middle| \mathcal{G}_s \right)$$

Thus

$$Z_s \exp\left(-\frac{1}{2}|\boldsymbol{\theta}|^2 (t-s) \right) = \mathbb{E}\left(Z_t \exp\left(i\boldsymbol{\theta} \cdot (\mathbf{B}_t - \mathbf{B}_s) \right) \middle| \mathcal{G}_s \right).$$

Consider a partition $0 < t_1 < t_2 < \cdots < t_m \leq T$, and by repeating the above argument, conditioning on each $\mathcal{G}_{t_j}$ we establish that

$$\mathbb{E}\left(Z_T \exp\left(i \sum_j \boldsymbol{\theta}_j \cdot (\mathbf{B}_{t_j} - \mathbf{B}_{t_{j-1}}) \right) \right) = \mathbb{E}\left(Z_0 \exp\left(-\frac{1}{2} \sum (t_j - t_{j-1})|\boldsymbol{\theta}_j|^2 \right) \right) = 0, \tag{3}$$

where the last equality follows since $Z_0 = 0$.

This is true for any choices of $\boldsymbol{\theta}_j \in \mathbb{R}^n$ for $j = 1, \ldots n$. The complex valued functions defined on $(\mathbb{R}^n)^m$ by

$$P^{(r)}(\mathbf{x}_1, \ldots, \mathbf{x}_m) = \sum_{k=1}^{K^{(n)}} c_k^{(r)} \exp\left(i \sum_{j=1}^m \mathbf{a}_{j,k}^{(r)} \cdot \mathbf{x}_j \right)$$

clearly separate points (i.e. for given distinct points we can choose coefficients such that the functions have distinct values at these points), form a linear space and are closed under complex conjugation. Therefore by the Stone-Weierstass theorem (see [Bollobas, 1990]), their uniform closure is the space of complex valued functions (recall that the complex variable form of this theorem only requires local compactness of the domain).

Therefore we can approximate any continuous bounded complex valued function $f : (\mathbb{R}^n)^m \to \mathbb{C}$ by a sequence of such Ps. But we have already shown in (3) that

$$\mathbb{E}\left(Z_T P^{(r)}(\mathbf{B}_{t_1}, \dots, \mathbf{B}_{t_n})\right) = 0$$

Hence by uniform approximation we can extend this to any f continuous, bounded

$$\mathbb{E}\left(Z_T f(\mathbf{B}_{t_1}, \dots, \mathbf{B}_{t_n})\right) = 0.$$

Now we use the monotone class framework; consider the class $\mathcal{H}$ such that for $H \in \mathcal{H}$,

$$\mathbb{E}(Z_T H) = 0$$

This $\{calH\}$ is a vector space, and contains the constant one since $\mathbb{E}(Z) = 0$. The foregoing argument shows that it contains all H measurable with respect to the sigma field $\sigma(\mathbf{B}_{t_1}, \dots, \mathbf{B}_{t_n})$ with $0 < t_1 < t_2 < \cdots < t_n \leq T$. Thus the monotone class theorem implies that it contains all functions which are measurable with respect to $\mathcal{G}_T$.

The function $Z_T \in \mathcal{G}_T$, and we have shown $\mathbb{E}(Z_T X) = 0$ for $X \in \mathcal{G}_T$. Thus we can take $X = Z_T$ whence $\mathbb{E}(Z_T^2) = 0$ which implies that $Z_T = 0$ a.s.. This establishes the desired result. □

The reader should examine the latter part of the proof carefully; it is in fact related to the proof that the set

$$\left\{\exp\left(i\int_0^t \boldsymbol{\theta}_s \cdot \mathrm{d}\mathbf{B}_s\right) : \boldsymbol{\theta} \in L^\infty([0,t], \mathbb{R}^m)\right\}$$

is total in L^1. A set S is said to be total if $\mathbb{E}(af) = 0$ for all $a \in S$ implies $a = 0$ a.s.. The full proof of this result will reappear in a more abstract form in the stochastic filtering section of these notes.

16. Stochastic Differential Equations

Stochastic differential equations arise naturally in various engineering problems, where the effects of random 'noise' perturbations to a system are being considered. For example in the problem of tracking a satelite, we know that it's motion will obey Newton's law to a very high degree of accuracy, so in theory we can integrate the trajectories from the initial point. However in practice there are other random effects which perturb the motion.

The variety of SDE to be considered here describes a *diffusion* process and has the form

$$\mathrm{d}X_t = b(t, X_t) + \sigma(t, X_t)\mathrm{d}B_t, \tag{$*$}$$

where $b_i(x,t)$, and $\sigma_{ij}(t,x)$ for $1 \le i \le d$ and $1 \le j \le r$ are Borel measurable functions.

In practice such SDEs generally occur written in the Statonowich form, but as we have seen the Itô form has numerous calculational advantages (especially the fact that local martinagles are a closed class under the Itô integral), so it is conventional to transform the SDE to the Itô form before proceeding.

Strong Solutions

A strong solution of the SDE ($*$) on the given probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with initial condition ζ is a process $(X_t)_{t\ge 0}$ which has continuous sample paths such that

(i) X_t is adapted to the augmented filtration generated by the Brownian motion B and initial condition ζ, which is denoted $\mathcal{F}_t$.

(ii) $\mathbb{P}(X_0 = \zeta) = 1$

(iii) For every $0 \le t < \infty$ and for each $1 \le i \le d$ and $1 \le j \le r$, then the folllowing holds almost surely

$$\int_0^t \left(|b_i(s, X_s)| + \sigma_{ij}^2(s, X_s)\right) \mathrm{d}s < \infty,$$

(iv) Almost surely the following holds

$$X_t = X_0 + \int_0^t b(s, X_s)\mathrm{d}s + \int_0^t \sigma(s, X_s)\mathrm{d}B_s.$$

Lipshitz Conditions

Let $\|\cdot\|$ denote the usual Euclidean norm on $\mathbb{R}^d$. Recall that a function f is said to be Lipshitz if there exists a constant K such that

$$\|f(\mathbf{x}) - f(\mathbf{y})\| \le K\|\mathbf{x} - \mathbf{y}\|,$$

we shall generalise the norm concept to a $(d \times r)$ matrix σ by defining

$$\|\sigma\|^2 = \sum_{i=1}^d \sum_{j=1}^r \sigma_{ij}^2.$$

The concept of Liphshitz continuity can be extended to that of local Lipshitz continuity, by requiring that for each n there exists K_n, such that for all x and y such that $\|x\| \le n$ and $\|y\| \le n$ then

$$\|f(\mathbf{x}) - f(\mathbf{y})\| \le K_n\|\mathbf{x} - \mathbf{y}\|.$$

Strong Uniqueness of Solutions

Theorem (Uniqueness) 16.1.
Suppose that $b(t, x)$ and $\sigma(t, x)$ are locally Lipshitz continuous in the spatial variable (x). That is for every $n \geq 1$ there exists a constant $K_n > 0$ such that for every $t \geq 0$, $\|x\| \leq n$ and $\|y\| \leq n$ the following holds

$$\|b(t,x) - b(t,y)\| + \|\sigma(t,x) - \sigma(t,y)\| \leq K_n \|x - y\|.$$

Then strong uniqueness holds for the pair (b, σ), which is to say that if X and $\tilde{X}$ are two strong solutions of $(*)$ relative to B with initial condition ζ then X and $\tilde{X}$ are indistinguishable, that is

$$\mathbb{P}\left[X_t = \tilde{X}_t \forall t : 0 \leq t < \infty\right] = 1.$$

The proof of this result is importanty inasmuch as it illustrates the first example of a technique of bounding which recurs again and again throughout the theory of stochastic differential equations. Therefore I make no apology for spelling the proof out in excessive detail, as it is most important to understand exactly where each step comes from!

Proof
Suppose that X and $\tilde{X}$ are strong solutions of $(*)$, relative to the same brownian motion B and initial condition ζ on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Define a sequence of stopping times

$$\tau_n = \inf\{t \geq 0 : \|X_t\| \geq n\}, \text{ and } \tilde{\tau}_n = \inf\{t \geq 0 : \|Y_t\| \geq n\}.$$

Now set $S_n = \min(\tau_n, \tilde{\tau}_n)$. Clearly S_n is also a stopping time, and $S_n \to \infty$ a.s. $(\mathbb{P})$ as $n \to \infty$. These stopping times are only needed because b and σ are being assumed merely to be locally Lipshitz. If they are assumed to be Lipshitz, as will be needed in the existence part of the proof, then this complexity may be ignored.

Hence

$$\begin{aligned} X_{t\wedge S_n} - \tilde{X}_{t\wedge S_n} = &\int_0^{t\wedge S_n} \left(b(u, X_u) - b(u, \tilde{X}_u)\right) \mathrm{d}u \\ &+ \int_0^{t\wedge S_n} \left(\sigma(u, X_u) - \sigma(u, \tilde{X}_u)\right) \mathrm{d}W_u. \end{aligned}$$

Now we consider evaluating $\mathbb{E}\|X_{t\wedge S_n} - \tilde{X}_{t\wedge S_n}\|^2$, the first stage follows using the identity $(a+b)^2 \leq 4(a^2 + b^2)$,

$$\begin{aligned} \mathbb{E}\|X_{t\wedge S_n} - \tilde{X}_{t\wedge S_n}\|^2 \leq & 4\mathbb{E}\left[\int_0^{t\wedge S_n} \left(b(u, X_u) - b(u, \tilde{X}_u)\right) \mathrm{d}u\right]^2 \\ &+ 4\mathbb{E}\left[\int_0^{t\wedge S_n} \left(\sigma(u, X_u) - \sigma(u, \tilde{X}_u)\right) \mathrm{d}W_u\right]^2 \end{aligned}$$

Considering the second term, we use the *Itô isometry* which we remember states that $\|(H\cdot M)\|_2 = \|H\|_M$, so

$$\mathbb{E}\left[\int_0^{t\wedge S_n}\left(\sigma(u,X_u)-\sigma(u,\tilde{X}_u)\right)\mathrm{d}W_u\right]^2 = \mathbb{E}\left[\int_0^{t\wedge S_n}|\sigma(u,X_u)-\sigma(u,\tilde{X}_u)|^2\mathrm{d}u\right]$$

The classical Hölder inequality (in the form of the Cauchy Schwartz inequality) for Lebsgue integrals which states that for $p,q\in(1,\infty)$, with $p^{-1}+q^{-1}=1$ the following inequality is satisfied.

$$\int|f(x)g(x)|\mathrm{d}\mu(x)\le\left(\int|f(x)|^p\mathrm{d}\mu(x)\right)^{1/p}\left(\int|g(x)|^q\mathrm{d}\mu(x)\right)^{1/q}$$

This result may be applied to the other term, taking $p=q=2$ which yields

$$\begin{aligned}\mathbb{E}\left[\int_0^{t\wedge S_n}\left(b(u,X_u)-b(u,\tilde{X}_u)\right)\mathrm{d}u\right]^2 &\le \mathbb{E}\left[\int_0^{t\wedge S_n}\left|b(u,X_u)-b(u,\tilde{X}_u)\right|\mathrm{d}u\right]^2\\ &\le \mathbb{E}\left[\int_0^{t\wedge S_n}1\mathrm{d}s\int_0^{t\wedge S_n}\left(b(u,X_u)-b(u,\tilde{X}_u)\right)^2\mathrm{d}s\right]\\ &\le \mathbb{E}\left[t\int_0^{t\wedge S_n}\left(b(u,X_u)-b(u,\tilde{X}_u)\right)^2\right]\end{aligned}$$

Thus combining these two useful inequalities and using the nth local Lipshitz relations we have that

$$\begin{aligned}\mathbb{E}\|X_{t\wedge S_n}-\tilde{X}_{t\wedge S_n}\|^2 \le& 4t\mathbb{E}\left[\int_0^{t\wedge S_n}\left(b(u,X_u)-b(u,\tilde{X}_u)\right)^2\right]\\ &+4\mathbb{E}\left[\int_0^{t\wedge S_n}|\sigma(u,X_u)-\sigma(u,\tilde{X}_u)|^2\mathrm{d}u\right]\\ &\le 4(T+1)K_n^2\mathbb{E}\int_0^t\left(X_{u\wedge S_n}-\tilde{X}_{u\wedge S_n}\right)^2\mathrm{d}u\end{aligned}$$

Now by Gronwall's lemma, which in this case has a zero term outside of the integral, we see that $\mathbb{E}\|X_{t\wedge S_n}-\tilde{X}_{t\wedge S_n}\|^2=0$, and hence that $\mathbb{P}(X_{t\wedge S_n}=\tilde{X}_{t\wedge S_n})=1$ for all $t<\infty$. That is these two processes are modifications, and thus indistinguishable. Letting $n\to\infty$ we see that the same is true for $\{X_t\}_{t\ge0}$ and $\{\tilde{X}_t\}_{t\ge0}$. $\square$

Now we impose Lipshitz conditions on the functions b and σ to produce an existence result. The following form omits some measure theoretic details which are very important; for a clear treatment see Chung & Williams chapter 10.

Theorem (Existence) 16.2.
If the coefficients b and σ satisfy the global Lipshitz conditions that for all u, t

$$b(u,x) - b(u,y) \leq K|x-y|, \quad |\sigma(t,x) - \sigma(t,y)| \leq K|x-y|,$$

and additionally the bounded growth condition

$$|b(t,x)|^2 + |\sigma(t,x)|^2 \leq K^2(1+|x|^2)$$

Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let ξ be a random valued vector, independent of the Brownian Motion B_t, with finite second moment. Let $\mathcal{F}_t$ be the augmented filtration associated with the Brownian Motion B_t and ξ. Then there exists a continuous, adapted process X which is a strong solution of the SDE with initial condition ξ. Additionally this process is square integrable: for each $T > 0$ there exists $C(K,T)$ such that

$$\mathbb{E}|X_t|^2 \leq C\left(1 + \mathbb{E}|\xi|^2\right) e^{Ct},$$

for $0 \leq t \leq T$.

Proof

This proof proceeds by Picard iteration through a map F, analogously to the deterministic case to prove the existence of solutions to first order ordinary differential equations. This is a departure from the more conventional proof of this result. Let F be a map from the space $\mathcal{C}_T$ of continuous adapted processes X from $\Omega \times [0,T]$ to $\mathbb{R}$, such that $\mathbb{E}\left[\left(\sup_{t\leq T} X_t\right)^2\right] < \infty$. Define $X_t^{(k)}$ recursively, with $X_t^{(0)} = \xi$, and

$$X_t^{(k+1)} = F(X^k)_t = \xi + \int_0^t b(s, X_s^{(k)})\mathrm{d}s + \int_0^t \sigma(s, X_s^{(k)})\mathrm{d}B_s$$

[Note: we have left out checking that the image of X under F is indeed adapted!] Now note that using $(a+b)^2 \leq 2a^2 + ab^2$, we have using the same bounds as in the uniqueness result that

$$\begin{aligned}
\mathbb{E}\left[\left(\sup_{0\leq t\leq T} F(X)_t - F(Y)_t\right)^2\right] &\leq 2\mathbb{E}\left(\sup_{t\leq T}\left|\int_0^T \left(\sigma(X_s) - \sigma(Y_s)\right)\mathrm{d}B_s\right|^2\right) \\
&\quad + 2\mathbb{E}\left(\sup_{t\leq T}\left|\int_0^T \left(b(X_s) - b(Y_s)\right)\mathrm{d}s\right|^2\right) \\
&\leq 2K^2(4+T)\int_0^T \mathbb{E}\left[\left(\sup_{t\leq T}|X_t - Y_t|^2\right)^2\right]\mathrm{d}t.
\end{aligned}$$

By induction we see that for each T we can prove

$$\mathbb{E}\left[\left(\sup_{t\leq T} F^n(X) - F^n(Y)\right)^2\right] \leq \frac{C^n T^n}{n!}\mathbb{E}\left[\left(\sup_{t\leq T} X_t - Y_t\right)^2\right]$$

So by taking n sufficiently large we have that F^n is a contraction mapping and so by the contraction mapping theorem, F^n mapping $\mathcal{C}_T$ to itself has a fixed point, which must be unique, call it $X^{(T)}$. Clearly from the uniqueness part $X_t^{(T)} = X_t^{(T')}$ for $t \leq T \wedge T'$ a.s., and so we may consistently define $X \in \mathcal{C}$ by

$$X_t = X_t^{(N)} \text{ for } t \leq N, \; N \in \mathbb{N},$$

which solves the SDE, and has already been shown to be unique. $\square$

17. Relations to Second Order PDEs

The aim of this section is to show a rather surprising connection between stochastic differential equations and the solution of second order partial differential equations. Surprising though the results may seem they often provide a viable route to calculating the solutions of explicit PDEs (an example of this is solving the Black-Scholes Equation in Option Pricing, which is much easier to solve via stochastic methods, than as a second order PDE). At first this may well seem to be surprising since one problem is entirely deterministic and the other in inherently stochastic!

17.1. Infinitesimal Generator

Consider the following d-dimensional SDE,

$$\begin{aligned} \mathrm{d}\mathbf{X}_t =& \mathbf{b}(\mathbf{X}_t)\mathrm{d}t + \sigma(\mathbf{X}_t)\mathrm{d}\mathbf{B}_t, \\ \mathbf{X}_0 =& x_0 \end{aligned}$$

where σ is a $d \times d$ matrix with elements $\sigma = \{\sigma_{ij}\}$. This SDE has infinitesimal generator A, where

$$A = \sum_{j=1}^d b^k(X_t)\frac{\partial}{\partial x^j} + \frac{1}{2}\sum_{i=1}^d\sum_{j=1}^d\sum_{k=1}^d \sigma_{ik}(\mathbf{X}_t)\sigma_{kj}(\mathbf{X}_t)\frac{\partial^2}{\partial x^i \partial x^j}.$$

It is conventional to set

$$a_{ij} = \sum_{k=1}^d \sigma_{ik}\sigma_{kj},$$

whence A takes the simpler form

$$A = \sum_{j=1}^d b^j(X_t)\frac{\partial}{\partial x^j} + \frac{1}{2}\sum_{i=1}^d\sum_{j=1}^d a_{ij}(\mathbf{X}_t)\frac{\partial^2}{\partial x^i \partial x^j}.$$

Why is the definition useful? Consider application of Itô's formula to $f(\mathbf{X}_t)$, which yields

$$f(\mathbf{X}_t) - f(\mathbf{X}_0) = \int_0^t \sum_{j=1}^d \frac{\partial f}{\partial x^j}(\mathbf{X}_s)\mathrm{d}\mathbf{X}_s + \frac{1}{2}\int_0^t \sum_{i=1}^d\sum_{j=1}^d \frac{\partial^2 f}{\partial x^i \partial x^j}(\mathbf{X}_s)\mathrm{d}\langle X^i, X^j\rangle_s.$$

Substituting for $\mathrm{d}\mathbf{X}_t$ from the SDE we obtain,

$$\begin{aligned} f(\mathbf{X}_t) - f(\mathbf{X}_0) =& \int_0^t \left(\sum_{j=1}^d b^j(\mathbf{X}_s)\frac{\partial f}{\partial x^j}(\mathbf{X}_s) + \frac{1}{2}\sum_{i=1}^d\sum_{j=1}^d\sum_{k=1}^d \sigma_{ik}\sigma_{kj}\frac{\partial^2 f}{\partial x^i \partial x^j}(\mathbf{X}_s)\right)\mathrm{d}t \\ &+ \int_0^t \sum_{j=1}^d \sigma_{ij}(\mathbf{X}_s)\frac{\partial f}{\partial x^j}(\mathbf{X}_s)\mathrm{d}\mathbf{B}_s \\ =& \int_0^t Af(\mathbf{X}_s)\mathrm{d}s + \int_0^t \sum_{j=1}^d \sigma_{ij}(\mathbf{X}_s)\frac{\partial f}{\partial x^j}(\mathbf{X}_s)\mathrm{d}\mathbf{B}_s \end{aligned}$$

Definition 17.1.
We say that X_t satisfies the martingale problem for A, if X_t is $\mathcal{F}_t$ adapted and

$$M_t = f(\mathbf{X}_t) - f(\mathbf{X}_0) - \int_0^t Af(X_s)\mathrm{d}s,$$

is a martingale for each $f \in C_c^2(\mathbb{R}^d)$.

It is simple to verify from the foregoing that any solution of the associated SDE solves the martingale problem for A. This can be generalised if we consider test functions $\phi \in C^2(\mathbb{R}^+ \times \mathbb{R}^d, \mathbb{R})$, and define

$$M_t^\phi := \phi(t, \mathbf{X}_t) - \phi(0, \mathbf{X}_0) - \int_0^t \left(\frac{\partial}{\partial s} + A\right) \phi(s, \mathbf{X}_s)\mathrm{d}s.$$

then M_t^ϕ is a local martingale, for $\mathbf{X}_t$ a solution of the SDE associated with the infinitesimal generator A. The proof follows by an application of Itô's formula to M_t^ϕ, similar to that of the above discussion.

17.2. The Dirichlet Problem

Let Ω be a subspace of $\mathbb{R}^d$ with a smooth boundary $\partial\Omega$. The *Dirichlet Problem* for f is defined as the solution of the system

$$\begin{aligned} &Au + \phi = 0 \text{ on } \Omega, \\ &u = f \text{ on } \partial\Omega. \end{aligned}$$

Where A is a second order partial differential operator of the form

$$A = \sum_{j=1}^d b^j(\mathbf{X}_t)\frac{\partial}{\partial x^j} + \frac{1}{2}\sum_{i=1}^d \sum_{j=1}^d a_{ij}(\mathbf{X}_t)\frac{\partial^2}{\partial x^i \partial x^j},$$

which is associated as before to an SDE. This SDE will play an important role in what is to follow.

A simple example of a Dirichlet Problem is the solution of the Laplace equation in the disc, with Dirichlet boundary conditions on the boundary, i.e.

$$\begin{aligned} &\nabla^2 u = 0 \text{ on } D, \\ &u = f \text{ on } \partial D. \end{aligned}$$

Theorem 17.2.
For each $f \in C_b^2(\partial\Omega)$ there exists a unique $u \in C_b^2(\overline{\Omega})$ solving the Dirichlet problem for f. Moreover there exists a continuous function $m : \overline{\Omega} \to (0, \infty)$ such that for all $f \in C_b^2(\partial\Omega)$ this solution is given by

$$u(\mathbf{x}) = \int_{\partial\Omega} m(\mathbf{x}, \mathbf{y}) f(\mathbf{y}) \sigma(\mathrm{d}\mathbf{y}).$$

Now remember the SDE which is associated with the infinitesimal generator A:

$$\begin{aligned} \mathrm{d}\mathbf{X}_t =& \mathbf{b}(\mathbf{X}_t)\mathrm{d}t + \sigma(\mathbf{X}_t)\mathrm{d}\mathbf{B}_t, \\ \mathbf{X}_0 =& x_0 \end{aligned}$$

Often in what follows we shall want to consider the conditional expectation and probability measures, conditional on $x_0 = x$, these will be denoted $\mathbb{E}_x$ and $\mathbb{P}_x$ respectively.

Theorem (Dirichlet Solution).
Define a stopping time via

$$T := \inf\{t \geq 0 : X_t \notin \Omega\}.$$

Then $u(\mathbf{x})$ given by

$$u(\mathbf{x}) := \mathbb{E}_x \left[\int_0^T \phi(\mathbf{X}_s)\mathrm{d}s + f(\mathbf{X}_T) \right],$$

solves the Dirichlet problem for f.

Proof
Define

$$M_t := u(\mathbf{X}_{T\wedge t}) + \int_0^{t\wedge T} \phi(\mathbf{X}_s)\mathrm{d}s.$$

We shall now show that this M_t is a martingale. For $t \geq T$, it is clear that $\mathrm{d}M_t = 0$. For $t < T$ by Itô's formula

$$\mathrm{d}M_t = \mathrm{d}u(\mathbf{X}_t) + \phi(\mathbf{X}_t)\mathrm{d}t.$$

Also, by Itô's formula,

$$\begin{aligned} \mathrm{d}u(\mathbf{X}_t) =& \sum_{j=1}^d \frac{\partial u}{\partial x^i}(\mathbf{X}_t)\mathrm{d}X_t^i + \frac{1}{2}\sum_{i=1}^d\sum_{j=1}^d \frac{\partial^2 u}{\partial x^i \partial x^j}(\mathbf{X}_t)\mathrm{d}\langle X^i, X^j\rangle_t \\ =& \sum_{j=1}^d \frac{\partial u}{\partial x^j}(\mathbf{X}_t)\left[\mathbf{b}(\mathbf{X}_t)\mathrm{d}t + \sigma(\mathbf{X}_t)\mathrm{d}\mathbf{B}_t\right] + \frac{1}{2}\sum_{i=1}^d\sum_{j=1}^d\sum_{k=1}^d \sigma_{ik}\sigma_{kj}\frac{\partial^d u}{\partial x^i \partial x^j}(\mathbf{X}_t)\mathrm{d}t \\ =& Au(\mathbf{X}_t)\mathrm{d}t + \sum_{j=1}^d \sigma(\mathbf{X}_t)\frac{\partial u}{\partial x^j}(\mathbf{X}_t)\mathrm{d}\mathbf{B}_t. \end{aligned}$$

Putting these two applications of Itô's formula together yields

$$\mathrm{d}M_t = \left(Au(\mathbf{X}_t) + \phi(\mathbf{X}_t)\right)\mathrm{d}t + \sum_{j=1}^d \sigma(\mathbf{X}_t)\frac{\partial u}{\partial x^j}(\mathbf{X}_t)\mathrm{d}\mathbf{B}_t.$$

but since u solves the Dirichlet problem, then

$$(Au + \phi)(\mathbf{X}_t) = 0,$$

hence

$$\begin{aligned}
\mathrm{d}M_t &= (Au(\mathbf{X}_t) + \phi(\mathbf{X}_t))\,\mathrm{d}t + \sum_{j=1}^{d} \sigma(\mathbf{X}_t)\frac{\partial u}{\partial x^j}(\mathbf{X}_t)\mathrm{d}\mathbf{B}_t^j, \\
&= \sum_{j=1}^{d} \sigma(\mathbf{X}_t)\frac{\partial u}{\partial x^j}(\mathbf{X}_t)\mathrm{d}\mathbf{B}_t^j.
\end{aligned}$$

from which we conclude by the stability property of the stochastic integral that M_t is a local martingale. However M_t is uniformly bounded on $[0, t]$, and hence M_t is a martingale.

In particular, let $\phi(x) \equiv 1$, and $f \equiv 0$, by the optional stopping theorem, since $T \wedge t$ is a bounded stoppping time, this gives

$$u(\mathbf{x}) = \mathbb{E}_x(M_0) = \mathbb{E}_x(M_{T\wedge t}) = \mathbb{E}_x\left[u(\mathbf{X}_{T\wedge t}) + (T \wedge t)\right].$$

Letting $t \to \infty$, we have via monotone convergence that $\mathbb{E}_x(T) < \infty$, since we know that the solutions u is bounded from the PDE solution existance theorem; hence $T < \infty$ a.s.. We cannot simply apply the optional stopping theorem directly, since T is not necessarily a bounded stopping time. But for arbitrary ϕ and f, we have that

$$|M_t| \le \|u\|_\infty + T\|\phi\|_\infty = \sup_{\mathbf{x}\in\overline{\Omega}} |u(\mathbf{x})| + T\sup_{\mathbf{x}\in\overline{\Omega}} |\phi(\mathbf{x})|,$$

whence as $\mathbb{E}_x(T) < \infty$, the martingale M is uniformly integrable, and by the martingale convergence theorem has a limit M_∞. This limiting random variable is given by

$$M_\infty = f(\mathbf{X}_T) + \int_0^T \phi(\mathbf{X}_s)\mathrm{d}s.$$

Hence from the identity $\mathbb{E}_x M_0 = \mathbb{E}_x M_\infty$ we have that,

$$u(\mathbf{x}) = \mathbb{E}_x(M_0) = \mathbb{E}_x(M_\infty) = \mathbb{E}_x\left[f(\mathbf{X}_T) + \int_0^T \phi(\mathbf{X}_s)\mathrm{d}s\right].$$

$\square$

17.3. The Cauchy Problem

The *Cauchy Problem* for f, a C_b^2 function, is the solution of the system

$$\begin{aligned}
&\frac{\partial u}{\partial t} = Au \text{ on } \Omega \\
&u(0,\mathbf{x}) = f(\mathbf{x}) \text{ on } \mathbf{x} \in \Omega \\
&u(t,\mathbf{x}) = f(\mathbf{x}) \ \forall t \geq 0, \text{ on } \mathbf{x} \in \partial\Omega
\end{aligned}$$

A typical problem of this sort is to solve the heat equation,

$$\frac{\partial u}{\partial t} = \frac{1}{2}\nabla^2 u,$$

where the function u represents the temperature in a region Ω, and the boundary condition is to specify the temperature field over the region at time zero, i.e. a condition of the form

$$u(0,\mathbf{x}) = f(\mathbf{x}) \text{ for } x \in \Omega,$$

In addition the boundary has its temperature fixed at zero,

$$u(0,\mathbf{x}) = 0 \text{ for } x \in \partial\Omega.$$

If Ω is just the real line, then the solution has the beautifully simple form

$$u(t,x) = \mathbb{E}_x\left(f(B_t)\right),$$

where B_t is a standard Brownian Motion.

Theorem (Cauchy Existence) 17.3.
For each $f \in C_b^2(\mathbb{R}^d)$ there exists a unique u in $C_b^{1,2}(\mathbb{R}\times\mathbb{R}^d)$ such that u solves the Cauchy Problem for f. Moreover there exists a continuous function (the heat kernel)

$$p : (0,\infty)\times\mathbb{R}^d\times\mathbb{R}^d \to (0,\infty),$$

such that for all $f \in C_b^2(\mathbb{R}^d)$, the solution to the Cauchy Problem for f is given by

$$u(t,\mathbf{x}) = \int_{\mathbb{R}^d} p(t,\mathbf{x},\mathbf{y}) f(\mathbf{y}) \mathrm{d}\mathbf{y}.$$

Theorem 17.4.
Let $u \in C_b^{1,2}(\mathbb{R}\times\mathbb{R}^d)$ be the solution of the Cauchy Problem for f. Then define

$$T := \inf\{t \geq 0 : \mathbf{X}_t \notin \Omega\},$$

a stopping time. Then

$$u(t,\mathbf{x}) = \mathbb{E}_x\left[f(\mathbf{X}_{T\wedge t})\right]$$

Proof
Fix $s \in (0,\infty)$ and consider the time reversed process

$$M_t := u((s-t)\wedge T, \mathbf{X}_{t\wedge T}).$$

There are three cases now to consider; for $0 \leq T \leq t \leq s$, $M_t = u((s-t)\wedge T, \mathbf{X}_T)$, where $X_T \in \partial\Omega$, so from the boundary condition, $M_t = f(\mathbf{X}_T)$, and hence it is clear that $\mathrm{d}M_t = 0$. For $0 \leq s \leq T \leq t$ and for $0 \leq t \leq s \leq T$, the argument is similar; in the latter case by Itô's formula we obtain

$$\begin{aligned}
\mathrm{d}M_t =& -\frac{\partial u}{\partial t}(s-t,\mathbf{X}_t)\mathrm{d}t + \sum_{j=1}^{d} \frac{\partial u}{\partial x^j}(s-t,\mathbf{X}_t)\mathrm{d}X_t^j \\
&+ \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} \frac{\partial^2 u}{\partial x^i \partial x^j}(s-t,\mathbf{X}_t)\mathrm{d}\langle X^i, X^j\rangle_t, \\
=& \left(-\frac{\partial u}{\partial t} + Au\right)(s-t,\mathbf{X}_t)\mathrm{d}t + \sum_{j=1}^{d}\frac{\partial u}{\partial x^j}(s-t,\mathbf{X}_t)\sum_{k=1}^{d}\sigma_{jk}(\mathbf{X}_t)\mathrm{d}\mathbf{B}_t^k.
\end{aligned}$$

We obtain a similar result in the $0 \leq t \leq T \leq s$, case but with s replaced by T. Thus for u solving the Cauchy Problem for f, we have that

$$\left(-\frac{\partial u}{\partial t} + Au\right) = 0,$$

we see that M_t is a local martingale. Boundedness implies that M_t is a martingale, and hence by optional stopping

$$u(s,\mathbf{x}) = \mathbb{E}_x(M_0) = \mathbb{E}_x(M_s) = \mathbb{E}_x(f(\mathbf{X}_{s\wedge T})),$$

□

17.4. Feynman-Kač Representation

Feynman observed the following representation for the representation of the solution of a PDE via the expectation of a suitable function of a Brownian Motion 'intuitively' and the theory was later made rigorous by Kač.

In what context was Feynman interested in this problem? Consider the Schrödinger Equation,

$$-\frac{\hbar^2}{2m}\nabla^2\Phi(\mathbf{x},t) + V(\mathbf{x})\Phi(\mathbf{x},t) = i\hbar\frac{\partial}{\partial t}\Phi(\mathbf{x},t),$$

which is a second order PDE. Feynman introduced the concept of a path-integral to express solutions to such an equation. In a manner which is analogous to the 'Hamiltonian' principle in classical mechanics, there is an action integral which is minimised over all 'permissible paths' that the system can take.

We have already considered solving the Cauchy problem

$$\begin{aligned}
&\frac{\partial u}{\partial t} = Au \text{ on } \Omega \\
&u(0, \mathbf{x}) = f(\mathbf{x}) \text{ on } \mathbf{x} \in \Omega \\
&u(t, \mathbf{x}) = f(\mathbf{x}) \ \forall t \geq 0, \text{ on } \mathbf{x} \in \partial\Omega
\end{aligned}$$

where A is the generator of an SDE and hence of the form

$$A = \sum_{j=1}^{d} b^j(\mathbf{X}_t)\frac{\partial}{\partial x^j} + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} a_{ij}(\mathbf{X}_t)\frac{\partial^2}{\partial x^i \partial x^j}.$$

Now consider the more general form of the same Cauchy problem where we consider a Cauchy Problem with generator $\mathcal{L}$ of the form:

$$\mathcal{L} \equiv A + v = \sum_{j=1}^{d} b^j(\mathbf{X}_t)\frac{\partial}{\partial x^j} + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} a_{ij}(\mathbf{X}_t)\frac{\partial^2}{\partial x^i \partial x^j} + v(\mathbf{X}_t).$$

For example let $\mathbf{X}_t = \mathbf{B}_t$ corresponding to

$$A = \frac{1}{2}\nabla^2, \qquad \mathcal{L} = \frac{1}{2}\nabla^2 + v(\mathbf{X}_t),$$

so in this example we are solving the problem

$$\begin{aligned}
&\frac{\partial u}{\partial t} = \frac{1}{2}\nabla^2 u(t, \mathbf{X}_t) + v(\mathbf{X}_t)u(\mathbf{X}_t) \text{ on } \mathbb{R}^d. \\
&u(0, \mathbf{x}) = f(\mathbf{x}) \text{ on } \partial\mathbb{R}^d.
\end{aligned}$$

The Feynman-Kač Representation Theorem expresses the solution of a general second order PDE in terms of an expectation of a function of a Brownian Motion. To simplify the statement of the result, we shall work on $\Omega = \mathbb{R}^d$, since this removes the problem of considering the Brownian Motion hitting the boundary.

Theorem (Feynman-Kač Representation).
Let $u \in C_b^{1,2}(\mathbb{R} \times \mathbb{R}^d)$ be a solution of the Cauchy Problem with a generator of the above form with initial condition f, and let $\mathbf{X}_t$ be a solution of the SDE with infinitessimal generator A in $\mathbb{R}^d$ starting at $\mathbf{x}$. Then

$$u(t, \mathbf{x}) = \mathbb{E}_x\left[f(\mathbf{X}_t)\exp\left(\int_0^t v(\mathbf{X}_s)\mathrm{d}s\right)\right].$$

Proof
Note that the SDE for $\mathbf{X}_t$ has the form

$$\mathrm{d}\mathbf{X}_t = a(\mathbf{X}_t)\mathrm{d}\mathbf{B}_t + \mathbf{b}(\mathbf{X}_t)\mathrm{d}t.$$

Fix $s \in (0,\infty)$ and apply Itô's formula to

$$M_t = u(s-t, \mathbf{X}_t) \exp\left(\int_0^t v(\mathbf{X}_r) \mathrm{d}r\right).$$

For notational convenience, let

$$E_t = \exp\left(\int_0^t v(\mathbf{X}_r) \mathrm{d}r\right),$$

whence by Itô's formula $\mathrm{d}E_t = E_t v(\mathbf{X}_t)\mathrm{d}t$. For $0 \le t \le s$, we have

$$\begin{aligned}
\mathrm{d}M_t =& \sum_{j=1}^d \frac{\partial u}{\partial x^j}(s-t, \mathbf{X}_t) E_t \mathrm{d}B_t^j + \left(-\frac{\partial u}{\partial t} + Au + vu\right)(s-t, \mathbf{X}_t) E_t \mathrm{d}t \\
=& \sum_{j=1}^d \frac{\partial u}{\partial x^j}(s-t, \mathbf{X}_t) E_t \mathrm{d}B_t^j.
\end{aligned}$$

Hence M_t is a local martingale; since it it bounded, M_t is a martingale and hence by optional stopping

$$u(s, \mathbf{x}) = \mathbb{E}_x(M_0) = \mathbb{E}_x(M_s) = \mathbb{E}_x\left(f(\mathbf{X}_s)E_s\right).$$

□

18. Stochastic Filtering

Engineers have for many years studied system problems arising in the design of control systems to be placed between user inputs to control system parameters in order to achieve a desired response. For example the connection between the flight controls on a modern aircraft such as the Harrier Jump Jet and the aircraft control surfaces go via a complex control system rather than simple wire cables as found in early aircraft.

Even a basic control system such as a central heating thermostat is based upon an element of feedback – observing the state of the system (room temperature) and adjusting the system inputs (turning the boiler off or on) to keep the system state close to the desired value (the temperature set on the dial).

In the example of a high performance aircraft, sudden movement of an aircraft control surface communicated directly to the control surfaces might cause uncontrolled growing oscillations – which was not the response required and the control system would have to modify the control surfaces to avoid this happening.

In the real world the behaviour of these systems is affected by extra random unknown factors which can not be known; for example wind effects on the airframe, engine vibration. Also all measurements have inherent random errors which cannot be eliminated completely. Attempts to solve control problems in this context require that we have some idea how to use the observations made to give a best possible estimate of the system state at a given time.

In this basic introduction to stochastic non-linear filtering we shall consider a system described by a stochastic process called the signal process whose behaviour is governed by a dynamical system containing a Brownian motion noise input. This is observed, together with unavoidable observation noise. The goal of the stochastic filtering problem is to determine the law of the signal given the observations. This must be formalised in the language of stochastic calculus. The notation $\|\cdot\|$ will be used in this section to denote the Euclidean norm in d-dimensional space.

An important modern example of stochastic filtering is the use of GPS measurements to correct an inertial navigation system (INS). An INS system has transducers which measure acceleration. Starting from a known point in space the acceleration can be integrated twice to compute a position. But since there are error in acceleration measurements this position will gradually diverge from the true position. This is a phenomenon familiar to anyone who has navigated with a map by dead-reckoning (From the station I walk NE for a mile, E for 2 miles – where am I?).

The simple navigator checks his position as frequently as possible by other means (e.g. observing the bearing to several landmarks, taking a star-sight or by looking at a handheld GPS receiver).

In a similar fashion we might try and correct our INS position estimates by using periodic GPS observations; these GPS position measurements can not be made continuously and are subject to random errors; additionally the observer continues to move while taking the readings. What is the optimal way to use these GPS measurements to correct our INS estimate? The answer lies in the solution of a stochastic filtering problem. The *signal process* is the true position and the *observation process* is the GPS position process.

18.1. Signal Process

Let $\{\mathbf{X}_t, \mathcal{F}_t, t \geq 0\}$ be the *signal process*. It is defined to be the solution of the stochastic differential equation

$$\mathrm{d}\mathbf{X}_t = \mathbf{f}(t, \mathbf{X}_t)\mathrm{d}t + \sigma(t, \mathbf{X}_t)\mathrm{d}\mathbf{V}_t, \tag{4}$$

where V is a d dimensional standard Brownian motion. The coefficients satisfy the conditions

$$\begin{aligned} &\|\mathbf{f}(t,\mathbf{x}) - \mathbf{f}(t,\mathbf{y})\| + \|\sigma(t,\mathbf{x}) - \sigma(t,\mathbf{y})\| \leq k\|\mathbf{x} - \mathbf{y}\| \\ &\|\mathbf{f}(t,\mathbf{x})\|^2 + \|\sigma(t,\mathbf{x})\|^2 \leq k^2(1 + \|\mathbf{x}\|^2), \end{aligned}$$

which ensure that the equation for X has a unique solution.

18.2. Observation Process

The observation process satisfies the stochastic differential equation

$$\mathrm{d}\mathbf{Y}_t = \mathbf{h}(t, \mathbf{X}_t)\mathrm{d}t + \mathrm{d}\mathbf{W}_t, \quad Y_0 = 0, \tag{5}$$

where $\mathbf{W}_t$ is an m dimensional standard Brownian motion independent of $\mathbf{V}_t$. The function h taking values in $\mathbb{R}^m$ satisfies a linear growth condition

$$\|\mathbf{h}(t,\mathbf{x})\|^2 \leq k(1 + \|\mathbf{x}\|^2).$$

A consequence of this condition is that for any $T > 0$,

$$\mathbb{E}\left(\|\mathbf{X}_t\|^2\right) \leq C\left(1 + \mathbb{E}\left[\|\mathbf{X}_0\|^2\right]\right)e^{ct}, \tag{6}$$

for $t \in [0, T]$ with suitable constants C and c which may be functions of T. As a result of this bound

$$\mathbb{E}\left[\int_0^T \|\mathbf{h}(s, \mathbf{X}_s)\|^2 \,\mathrm{d}s\right] < \infty.$$

Given this process a sequence of observation σ-algebras may be defined

$$\mathcal{Y}_t^o := \sigma\left(\mathbf{Y}_s : 0 \leq s \leq t\right).$$

These must be augmented in the usual fashion to give

$$\mathcal{Y}_t = \mathcal{Y}_t^o \cup \mathcal{N},$$

where $\mathcal{N}$ is the set of $\mathbb{P}$ null subsets of Ω, and

$$\mathcal{Y} := \bigvee_{t \geq 0} \mathcal{Y}_t.$$

18.3. The Filtering Problem

The above has set the scene, we have a real physical system whose state at time t is represented by the vector $\mathbf{X}_t$. The system state is governed by an SDE with a noise term representing random perturbations to the system. This is observed to give the observation process $\mathbf{Y}_t$ which includes new independent noise (represented by $\mathbf{W}_t$).

The filtering problem is to find the conditional law of the signal process given the observations to date, i.e. to find

$$\pi_t(\phi) := \mathbb{E}\left(\phi(\mathbf{X}_t)|\mathcal{Y}_t\right) \quad \forall t \in [0,T]. \tag{7}$$

The following proofs are in many cases made complex by the fact that we do not assume that the functions f, h and σ are bounded.

18.4. Change of Measure

To solve the filtering problem Girsanov's theorem will be used to make a change of measure to a new measure under which the observation process Y is a Brownian motion. Let

$$Z_t := \exp\left(-\int_0^t \mathbf{h}^T(s,\mathbf{X}_s)\mathrm{d}\mathbf{W}_s - \frac{1}{2}\int_0^t \|h(s,\mathbf{X}_s)\|^2\,\mathrm{d}s\right), \tag{8}$$

be a real valued process.

Proposition 18.1.
The process Z_t is a martingale with respect to the filtration $\{\mathcal{F}_t, t \geq 0\}$.

Proof
The process Z_t satisfies

$$Z_t = 1 - \int_0^t Z_s\mathbf{h}^T(s,\mathbf{X}_s)\mathrm{d}\mathbf{W}_s,$$

that is Z_t solves the SDE

$$\begin{aligned}\mathrm{d}Z_t &= -Z_t\mathbf{h}^T(t,\mathbf{X}_t)\mathrm{d}W_t,\\ Z_0 &= 1.\end{aligned}$$

Given this expression, Z_t is a positive continuous local martingale and hence is a supermartingale. To prove that it is a true martingale we must prove in addition that it has constant mean.

Let T_n be a reducing sequence for the local martingale Z_t, i.e. an increasing sequence of stopping times tending to infinity as $n \to \infty$ such that for each n, Z^{T_n} is a genuine martingale. By Fatou's lemma, and the local martingale property

$$\mathbb{E}Z_t = \mathbb{E}\lim_{n\to\infty} Z_t^{T_n} \leq \liminf \mathbb{E}Z_t^{T_n} = \mathbb{E}Z_0^{T_n} = 1,$$

so

$$\mathbb{E}Z_t \leq 1 \quad \forall t. \tag{9}$$

This will be used as an upper bound in an application of the dominated convergence theorem. By application of Itô's formula to the function f defined by

$$f(x) = \frac{x}{1+\epsilon x},$$

we obtain that

$$f(Z_t) = f(Z_0) + \int_0^t f'(Z_s)\mathrm{d}Z_s + \frac{1}{2}\int_0^t f''(Z_s)\mathrm{d}\langle Z_s, Z_s\rangle,$$

Hence

$$\frac{Z_t}{1+\epsilon Z_t} = \frac{1}{1+\epsilon} - \int_0^t \frac{Z_s\mathbf{h}^T(s,\mathbf{X}_s)}{(1+\epsilon Z_s)^2}\mathrm{d}\mathbf{W}_s - \int_0^t \frac{\epsilon Z_s^2\|\mathbf{h}(s,\mathbf{X}_s)\|^2}{(1+\epsilon Z_s)^3}\mathrm{d}s. \tag{10}$$

Consider the term

$$\int_0^t \frac{Z_s\mathbf{h}^T(s,\mathbf{X}_s)}{(1+\epsilon Z_s)^2}\mathrm{d}\mathbf{W}_s,$$

clearly this a local martingale, since it is a stochastic integral. The next step in the proof is explained in detail as it is one which occurs frequently. We wish to show that the above stochastic integral is in fact a genuine martingale. From the earlier theory (for L^2 integrable martingales) it suffices to show that integrand is in $L^2(W)$. We therefore compute

$$\left\|\frac{Z_s\mathbf{h}^T(s,\mathbf{X}_s)}{(1+\epsilon Z_s)^2}\right\|_W = \mathbb{E}\left[\int_0^t \left\|\frac{Z_s\mathbf{h}^T(s,\mathbf{X}_s)}{(1+\epsilon Z_s)^2}\right\|^2 \mathrm{d}s\right],$$

with a view to showing that it is finite, whence the integrand must be in $L^2(W)$ and hence the integral a genuine martingale. We note that since $Z_s \geq 0$, that

$$\frac{1}{(1+\epsilon Z_s)} \leq 1, \text{ and } \frac{Z_s^2}{(1+\epsilon Z_s)^2} \leq \frac{1}{\epsilon^2},$$

so,

$$\begin{aligned}\left\|\frac{Z_s\mathbf{h}^T(s,\mathbf{X}_s)}{(1+\epsilon Z_s)^2}\right\|^2 &= \frac{Z_s^2}{(1+\epsilon Z_s)^4}\left\|\mathbf{h}(s,\mathbf{X}_s)\right\|^2 \\ &= \left[\frac{Z_s^2}{(1+\epsilon Z_s)^2}\right] \times \left[\frac{1}{(1+\epsilon Z_s)^2}\right] \times \left\|\mathbf{h}(s,\mathbf{X}_s)\right\|^2 \\ &\leq \frac{1}{\epsilon^2}\left\|\mathbf{h}(s,\mathbf{X}_s)\right\|^2.\end{aligned}$$

Using this inequality we obtain

$$\mathbb{E}\left[\int_0^t \left\|\frac{Z_s\mathbf{h}^T(s,\mathbf{X}_s)}{(1+\epsilon Z_s)^2}\right\|^2 \mathrm{d}s\right] \leq \frac{1}{\epsilon^2}\mathbb{E}\left[\int_0^T \left\|\mathbf{h}(s,\mathbf{X}_s)\right\|^2 \mathrm{d}s\right],$$

however as a consequence of our linear growth condition on h, the last expectation is finite. Therefore

$$\int_0^t \frac{Z_s\mathbf{h}^T(s,\mathbf{X}_s)}{(1+\epsilon Z_s)^2}\mathrm{d}\mathbf{W}_s$$

is a genuine martingale, and hence taking the expectation of (10) we obtain

$$\mathbb{E}\left[\frac{Z_t}{1+\epsilon Z_t}\right] = \frac{1}{1+\epsilon} - \mathbb{E}\left[\int_0^t \frac{\epsilon Z_s^2\|\mathbf{h}(s,X_s)\|^2}{(1+\epsilon Z_s)^3}\mathrm{d}s\right].$$

Consider the integrand on the right hand side; clearly it tends to zero as $\epsilon \to 0$. But also

$$\begin{aligned}\frac{\epsilon Z_s^2 \|\mathbf{h}(s, X_s)\|^2}{(1+\epsilon Z_s)^3} =& \frac{\epsilon Z_s}{(1+\epsilon Z_s)^3} \cdot Z_s \cdot \|\mathbf{h}(s, \mathbf{X}_s)\|^2 \\ \leq& \frac{1+\epsilon Z_s}{(1+\epsilon Z_s)^3} \cdot Z_s \cdot \|\mathbf{h}(s, \mathbf{X}_s)\|^2 \\ \leq& k Z_s \left(1 + \|\mathbf{X}_s\|^2\right),\end{aligned}$$

where we have used the fact that $\|\mathbf{h}(s, X_s)\|^2 \leq k(1 + \|\mathbf{X}_s\|^2)$. So from the fact that $\mathbb{E} Z_s \leq 1$ and using the next lemma we shall see that $\mathbb{E}\left(Z_s \|\mathbf{X}_s\|^2\right) \leq C$. Hence we conclude that $k Z_s(1 + \|\mathbf{X}_s\|^2)$ is an integrable dominating function for the integrand on interest. Hence by the Dominated Convergence Theorem, as $\epsilon \to 0$,

$$\mathbb{E}\left[\int_0^t \frac{\epsilon Z_s^2 \|\mathbf{h}(s, \mathbf{X}_s)\|^2}{(1+\epsilon Z_s)^3} \mathrm{d}s\right] \to 0.$$

In addition, since $\mathbb{E}(Z_s) \leq 1$, the bounded convergence theorem yields

$$\mathbb{E}\left(\frac{Z_s}{1+\epsilon Z_s}\right) \to \mathbb{E}(Z_s), \text{ as } \epsilon \to 0.$$

Hence we conclude that

$$\mathbb{E}(Z_t) = 1 \quad \forall t,$$

and so Z_t is a genuine martingale. □

Lemma 18.2.

$$\mathbb{E}\left[\int_0^t Z_s \|\mathbf{X}_s\|^2 \mathrm{d}s\right] < C(T) \quad \forall t \in [0, T].$$

Proof

To establish this result Itô's formula is used to derive two important results

$$\begin{aligned}\mathrm{d}(\|\mathbf{X}_t\|^2) =& 2\mathbf{X}_t^T(\mathbf{f}\mathrm{d}t + \sigma \mathrm{d}\mathbf{W}_s) + \mathrm{tr}(\sigma^T \sigma)\mathrm{d}t \\ \mathrm{d}(Z_t \|\mathbf{X}_t\|^2) =& -Z_t \|\mathbf{X}_t\|^2 \mathbf{h}^T \mathrm{d}\mathbf{V}_s + Z_t \left(2\mathbf{X}_t^T(\mathbf{f}\mathrm{d}t + \sigma \mathrm{d}\mathbf{W}_s) + \mathrm{tr}(\sigma\sigma^T)\right)\mathrm{d}t\end{aligned}$$

By Itô's formula,

$$\mathrm{d}\left(\frac{Z_t \|\mathbf{X}_t\|^2}{1+\epsilon Z_t \|\mathbf{X}_t\|^2}\right) = \frac{1}{(1+\epsilon Z_t \|\mathbf{X}_t\|^2)^2} \mathrm{d}\left(Z_t \|\mathbf{X}_t\|^2\right) + \frac{\epsilon}{(1+\epsilon Z_t \|\mathbf{X}_t\|^2)^3} \mathrm{d}\langle Z_t \|\mathbf{X}_t\|^2 \rangle.$$

Substituting for $d(Z_t \|\mathbf{X}_t\|^2)$ into this expression yields

$$\begin{aligned}\mathrm{d}\left(\frac{Z_t \|\mathbf{X}_t\|^2}{1+\epsilon Z_t \|\mathbf{X}_t\|^2}\right) =& \frac{1}{\left(1+\epsilon Z_t \|\mathbf{X}_t\|^2\right)^2} \left[-Z_t \|\mathbf{X}_t\|^2 \mathbf{h}^T \mathrm{d}\mathbf{W}_t + Z_t 2\mathbf{X}_t^T \sigma \mathrm{d}\mathbf{V}_s\right] \\ &+ \left[\frac{Z_t\left(2\mathbf{X}_t^T \mathbf{f} + \mathrm{tr}(\sigma\sigma^T)\right)}{\left(1+\epsilon Z_t \|\mathbf{X}_t\|^2\right)^2} - \frac{\epsilon\left(Z_t^2 \|\mathbf{X}_t\|^4 \mathbf{h}^T \mathbf{h} + 4Z_t^2 2\mathbf{X}_t^T \sigma\sigma^T \mathbf{X}_t\right)}{\left(1+\epsilon Z_t \|\mathbf{X}_t\|^2\right)^3}\right] \mathrm{d}t.\end{aligned}$$

After integrating this expression from 0 to t, the terms which are stochastic integrals are local martingales. In fact they can be shown to be genuine martingales. For example consider the term

$$\int_0^t \frac{Z_s 2\mathbf{X}_s^T \sigma}{(1+\epsilon Z_s \|\mathbf{X}_s\|^2)^2}\,\mathrm{d}\mathbf{V}_s,$$

to show that this is a martingale we must therefore establish that

$$\mathbb{E}\left[\int_0^t \left[\frac{Z_s 2\mathbf{X}_s^T \sigma}{(1+\epsilon Z_s\|\mathbf{X}_s\|^2)^2}\right]^2 \mathrm{d}s\right] = 4\mathbb{E}\left[\int_0^t \frac{Z_s^2 \mathbf{X}_s^T \sigma\sigma^T \mathbf{X}_s}{(1+\epsilon Z_s\|\mathbf{X}_s\|^2)^4}\mathrm{d}s\right] < \infty.$$

In order to establish this inequality consider the term $\mathbf{X}_t^T\sigma\sigma^T\mathbf{X}_t$ which is a sum over terms of the form

$$|X_t^i \sigma_{ij}(t,\mathbf{X}_t)\sigma_{kj}(t,\mathbf{X}_t)X_t^k| \le \|\mathbf{X}_t\|^2\|\sigma\|^2,$$

of which there are d^3 terms (in $\mathbb{R}^d$). But by the linear increase condition on σ, we have for some constant κ

$$\|\sigma\|^2 \le \kappa(1+\|\mathbf{X}\|^2),$$

and hence

$$|X_t^i \sigma_{ij}(t,\mathbf{X}_t)\sigma_{kj}(t,\mathbf{X}_t)X_t^k| \le \kappa\|\mathbf{X}_t\|^2\left(1+\|\mathbf{X}_t\|^2\right),$$

so the integral may be bounded by

$$\begin{aligned}\int_0^t \frac{Z_s^2\mathbf{X}_s^T\sigma\sigma^T\mathbf{X}_s}{(1+\epsilon Z_s\|\mathbf{X}_s\|^2)^4}\mathrm{d}s &\le \kappa d^3\int_0^t \frac{Z_s^2\|\mathbf{X}_s\|^2\left(1+\|\mathbf{X}_s\|^2\right)}{(1+\epsilon Z_s\|\mathbf{X}_s\|^2)^4}\mathrm{d}s\\ &= \kappa d^3\left[\int_0^t \frac{Z_s^2\|\mathbf{X}_s\|^2}{(1+\epsilon Z_s\|\mathbf{X}_s\|^2)^4}\mathrm{d}s + \int_0^t \frac{Z_s^2\|\mathbf{X}_s\|^4}{(1+\epsilon Z_s\|\mathbf{X}_s\|^2)^4}\right]\mathrm{d}s\end{aligned}$$

Considering each term separately, the first satisfies

$$\begin{aligned}\int_0^t \frac{Z_s^2\|\mathbf{X}_s\|^2}{(1+\epsilon Z_s\|\mathbf{X}_s\|^2)^4}\mathrm{d}s &\le \int_0^t Z_s \times \frac{Z_s\|\mathbf{X}_s\|^2}{(1+\epsilon Z_s\|\mathbf{X}_t\|^2)} \times \frac{1}{(1+\epsilon Z_s\|\mathbf{X}_s\|^2)^3}\,\mathrm{d}s\\ &\le \int_0^t \frac{Z_s}{\epsilon}\mathrm{d}s \le \frac{1}{\epsilon}\int_0^t Z_s.\end{aligned}$$

The last integral has a bounded expectation since $\mathbb{E}(Z_s) \le 1$. Similarly for the second term,

$$\int_0^t \frac{Z_s^2\|\mathbf{X}_s\|^4}{(1+\epsilon Z_s\|\mathbf{X}_s\|^2)^4}\mathrm{d}s \le \int_0^t \frac{Z_s^2\|\mathbf{X}_s\|^4}{(1+\epsilon Z_s\|\mathbf{X}_s\|^2)^2} \times \frac{1}{(1+\epsilon Z_s\|\mathbf{X}_s\|^2)^2}\mathrm{d}s \le \frac{1}{\epsilon^2}t < \infty.$$

A similar argument holds for the other stochastic integrals, so they must both be genuine martingales, and hence if we take the expectations of the integrals they are zero. Integrating the whole expression from 0 to t, then taking the expectation and finally differentiation with respect to t yields

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\left(\frac{Z_t\|\mathbf{X}_t\|^2}{1+\epsilon Z_t\|\mathbf{X}_t\|^2}\right) &\leq \mathbb{E}\left(\frac{Z_t(2\mathbf{X}_t^T\mathbf{f}+\mathrm{tr}(\sigma\sigma^T))}{(1+\epsilon Z_t\|\mathbf{X}_t\|^2)^2}\right)\\ &\leq k\left(1+\mathbb{E}\left[\frac{Z_t\|\mathbf{X}_t\|^2}{1+\epsilon Z_t\|\mathbf{X}_t\|^2}\right]\right).\end{aligned}$$

An application of Gronwall's inequality (see the next section for a proof of this useful result) establishes,

$$\mathbb{E}\left(\frac{Z_t\|\mathbf{X}_t\|^2}{1+\epsilon Z_t\|\mathbf{X}_t\|^2}\right) \leq C(T); \quad \forall t \in [0,T].$$

Applying Fatou's lemma the desired result is obtained. □

Given that Z_t has been shown to be a martingale a new probability measure $\tilde{\mathbb{P}}$ may be defined by the Radon-Nikodym derivative:

$$\left.\frac{\mathrm{d}\tilde{\mathbb{P}}}{\mathrm{d}\mathbb{P}}\right|_{\mathcal{F}_T^o} = Z_T.$$

This change of measure allows us to establish one of the most important results in stochastic filtering theory.

Proposition 18.3.
Under the probability measure $\tilde{\mathbb{P}}$, the observations process Y_t is a Brownian motion, which is independent of X_t.

Proof
Define a process M_t taking values in $\mathbb{R}$ by

$$M_t = -\int_0^t \mathbf{h}^T(s, X_s)\,\mathrm{d}\mathbf{W}_s, \tag{11}$$

is a martingale. Since $\mathbf{W}_t$ is a $\mathbb{P}$ Brownian Motion, which is a martingale under $\mathbb{P}$, it follows that M_t is a $\mathbb{P}$ local martingale. The previous lemma then establishes that M_t is in fact a $\mathbb{P}$ martingale. Thus we may apply Girsanov's theorem, to construct a new measure where

$$\mathbf{Y}_t = \left.\frac{\mathrm{d}\tilde{\mathbb{P}}}{\mathrm{d}\mathbb{P}}\right|_t = \mathcal{E}(M_t) = \exp\left(-\int_0^t h^T(s, X_s)\,\mathrm{d}\mathbf{W}_s - \frac{1}{2}\int_0^t \|\mathbf{h}(s, X_s)\|^2\,\mathrm{d}s\right) = Z_t.$$

As a corollary of Girsanov's theorem $W_t - \langle W, M\rangle_t$ is a continuous $\tilde{\mathbb{P}}$ local martingale. Lévy's characterisation theorem then implies that under $\tilde{\mathbb{P}}$ the process

$$\mathbf{W}_t - \langle \mathbf{W}, M\rangle_t = \mathbf{W}_t + \int_0^t \mathbf{h}(s, \mathbf{X}_s)\,\mathrm{d}s,$$

is an m dimensional Brownian Motion. But this process is just the observation process $\mathbf{Y}_t$ by definition. Thus we have proven that under $\tilde{\mathbb{P}}$, the observation process $\mathbf{Y}_t$ is a Brownian Motion. To establish the second part of the proposition we must prove the independence of $\mathbf{Y}$ and $\mathbf{X}$. The law of the pair $(\mathbf{X},\mathbf{Y})$ on $[0,T]$ is absolutely continuous with respect to that of the pair $(\mathbf{X},\mathbf{V})$ on $[0,T]$, since the latter is equal to the former plus a drift term.

Their Radon-Nikodym derivative (which exists since they are absolutely continuous with respect to each other) is Ψ_T, which means that for f any bounded measurable function

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}\left[f(\mathbf{X}_t,\mathbf{Y}_t)\Psi_\infty\right] &= \mathbb{E}_{\mathbb{P}}\left[\mathbb{E}_{\mathbb{P}}\left[f(\mathbf{X}_t,\mathbf{Y}_t)\Psi_t|\mathcal{F}_t\right]\right]\\ &= \mathbb{E}_{\mathbb{P}}\left[f(\mathbf{X}_t,\mathbf{Y}_t)\mathbb{E}_{\mathbb{P}}\left[\Psi_\infty|\mathcal{F}_t\right]\right]\\ &= \mathbb{E}_{\mathbb{P}}\left[f(\mathbf{X}_t,\mathbf{Y}_t)\Psi_t\right] = \mathbb{E}_{\mathbb{P}}\left[f(\mathbf{X}_t,\mathbf{V}_t)\right].\end{aligned}$$

Hence in terms of the probability measure $\tilde{\mathbb{P}}$

$$\mathbb{E}_{\tilde{\mathbb{P}}}\left[f(\mathbf{X}_t,\mathbf{Y}_t)\right] = \mathbb{E}_{\mathbb{P}}\left[f(\mathbf{X}_t,\mathbf{Y}_t)Z_\infty\right] = \mathbb{E}_{\mathbb{P}}\left[f(\mathbf{X}_t,\mathbf{V}_t)\right],$$

and hence under $\tilde{\mathbb{P}}$, the variables $\mathbf{X}$ and $\mathbf{Y}$ are independent. □

Proposition 18.4.
Let U be an integrable $\mathcal{F}_t$ measurable random variable. Then

$$\mathbb{E}_{\tilde{\mathbb{P}}}\left[U|\mathcal{Y}_t\right] = \mathbb{E}_{\tilde{\mathbb{P}}}\left[U|\mathcal{Y}\right].$$

Proof
Define

$$\mathcal{Y}'_t := \sigma\left(\mathbf{Y}_{t+u} - \mathbf{Y}_t : u \geq 0\right),$$

then $\mathcal{Y} = \mathcal{Y}_t \cup \mathcal{Y}'_t$. Under the measure $\tilde{\mathbb{P}}$, the σ-algebra $\mathcal{Y}'_t$ is independent of $\mathcal{F}_t$, since Y is an $\mathcal{F}_t$ adapted Brownian Motion which implies that it satisfies the independent increment property. Hence

$$\mathbb{E}_{\tilde{\mathbb{P}}}\left[U|\mathcal{Y}_t\right] = \mathbb{E}_{\tilde{\mathbb{P}}}\left[U|\mathcal{Y}_t \cup \mathcal{Y}'_t\right] = \mathbb{E}_{\tilde{\mathbb{P}}}\left[U|\mathcal{Y}\right].$$

□

18.5. The Unnormalised Conditional Distribution

To simplify the notation in what follows define $\tilde{Z}_t = 1/Z_t$. Then under the measure $\tilde{\mathbb{P}}$ we have

$$\mathrm{d}\tilde{Z}_t = \tilde{Z}_t\mathbf{h}^T(t,\mathbf{X}_t)\mathrm{d}\mathbf{Y}_t, \qquad (12)$$

so $\tilde{Z}_t$ is a $\tilde{\mathbb{P}}$ local martingale. Hence

$$\tilde{Z}_t = \exp\left(\int_0^t \mathbf{h}^T(s,X_s)\mathrm{d}\mathbf{Y}_t - \frac{1}{2}\int_0^t \|\mathbf{h}(s,X_s)\|^2\mathrm{d}s\right).$$

Also for any stopping time T, $\mathbb{E}_{\tilde{\mathbb{P}}}\left(\tilde{Z}_T\right) = \mathbb{E}_{\mathbb{P}}\left(Z_T\tilde{Z}_T\right) = 1$, the constant mean condition implies that $\tilde{Z}_T$ is a UI $\tilde{\mathbb{P}}$ martingale and thus by the martingale convergence theorem $Z_t \to Z_\infty$ in L^1. So we may write

$$\left.\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\tilde{\mathbb{P}}}\right|_{\mathcal{F}_\infty} = \tilde{Z}_\infty,$$

whence

$$\left.\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\tilde{\mathbb{P}}}\right|_{\mathcal{F}_t} = \tilde{Z}_t, \text{ for all } t \geq 0.$$

Now **define**, for every bounded measurable function ϕ, the *unnormalised conditional distribution* of $\mathbf{X}$ via

$$\begin{aligned}\rho_t(\phi) :=& \mathbb{E}_{\tilde{\mathbb{P}}}\left[\phi(\mathbf{X}_t)\tilde{Z}_t|\mathcal{Y}_t\right]\\ =& \mathbb{E}_{\tilde{\mathbb{P}}}\left[\phi(\mathbf{X}_t)\tilde{Z}_t|\mathcal{Y}\right],\end{aligned} \tag{13}$$

using the result of proposition 18.4. □

The following result within the context of stochastic filtering is sometimes called the Kallianpur-Striebel formula.

Proposition 18.5.
For every bounded measurable function ϕ, we have

$$\pi_t(\phi) := \mathbb{E}_{\mathbb{P}}\left[\phi(\mathbf{X}_t)|\mathcal{Y}_t\right] = \frac{\rho_t(\phi)}{\rho_t(1)}, \ \mathbb{P} \text{ and } \tilde{\mathbb{P}} \text{ a.s..}$$

Proof
Let b be a bounded $\mathcal{Y}_t$ measurable function. From the definition of π_t viz

$$\pi_t(\phi) := \mathbb{E}_{\mathbb{P}}\left[\phi(\mathbf{X}_t)|\mathcal{Y}_t\right],$$

we deduce from the definition of conditional expectation that

$$\mathbb{E}_{\mathbb{P}}\left(\pi_t(\phi)b\right) = \mathbb{E}_{\mathbb{P}}\left(\phi(\mathbf{X}_t)b\right),$$

hence by the tower property of conditional expectation

$$\mathbb{E}_{\tilde{\mathbb{P}}}\left(\tilde{Z}_t\pi_t(\phi)b\right) = \mathbb{E}_{\tilde{\mathbb{P}}}\left(\tilde{Z}_t\phi(\mathbf{X}_t)b\right).$$

Whence

$$\mathbb{E}_{\tilde{\mathbb{P}}}\left(\pi_t(\phi)\mathbb{E}_{\tilde{\mathbb{P}}}\left[\tilde{Z}_t|\mathcal{Y}_t\right]b\right) = \mathbb{E}_{\tilde{\mathbb{P}}}\left(\mathbb{E}_{\tilde{\mathbb{P}}}\left[\phi(\mathbf{X}_t)\tilde{Z}_t|\mathcal{Y}_t\right]b\right).$$

which is equivalent to

$$\mathbb{E}_{\tilde{\mathbb{P}}}\left(\pi_t(\phi)\rho_t(1)b\right) = \mathbb{E}_{\tilde{\mathbb{P}}}\left(\rho_t(\phi)b\right).$$

This implies that

$$\mathbb{E}_{\tilde{\mathbb{P}}}\left(\left.\pi_t(\phi)\rho_t(1)\right|\mathcal{Y}_t\right) = \mathbb{E}_{\tilde{\mathbb{P}}}\left(\left.\rho_t(\phi)\right|\mathcal{Y}_t\right)$$

As the random variables $\pi_t(\phi)$, $\rho_t(\phi)$ and $\rho_t(1)$ from their definitions are all $\mathcal{Y}_t$ measurable this implies that

$$\pi_t(\phi)\rho_t(1) = \rho_t(\phi)$$

Since $\tilde{Z}_t > 0$ a.s. which implies that $\rho_t(1) > 0$ a.s. this is sufficient to prove the required result. □

Remark
The above proof is simply a specific proof of the abstract form of Bayes' theorem, the general form of which is a standard result, see e.g. A.0.3 in [Musiela and Rutkowski, 2005].

18.6. The Zakai Equation

We wish to derive a stochastic differential equation which is satisfied by the unnormalised density ρ_t.

Definition 18.6.
A set S is said to be total in $L^1(\Omega, \mathcal{F}_t, \mathbb{P})$, if for every $a \in L^1(\Omega, \mathcal{F}_t, \mathbb{P})$

$$\mathbb{E}\left(a\epsilon_t\right) = 0 \quad \forall \epsilon \in S \Rightarrow a = 0.$$

Lemma 18.7.
Let

$$S_t := \left\{ \epsilon_t = \exp\left(i \int_0^t \mathbf{r}_s^T \mathrm{d}\mathbf{Y}_s + \frac{1}{2} \int_0^t \|r_s\|^2 \mathrm{d}s \right) \quad : \quad \mathbf{r}_s \in L^\infty\left([0,t], \mathbb{R}^m\right) \right\},$$

then the set S_t is a total set in $L^1(\Omega, \mathcal{Y}_t, \tilde{\mathbb{P}})$. The following proof of this standard result is taken from [Bensoussan, 1982].

Proof
It suffices to show that if $a \in L^1(\Omega, \mathcal{Y}_t, \tilde{\mathbb{P}})$ and $\tilde{\mathbb{E}}(a\epsilon_t) = 0$ for all $\epsilon_t \in S_t$ that $a = 0$. Take $t_1, t_2, \ldots, t_p \in (0,t)$ such that $t_1 < t_2 < \cdots < t_p$, then given $\mathbf{k}_1, \ldots, \mathbf{k}_p$

$$\sum_{j=1}^p \mathbf{k}_j^T \mathbf{Y}_{t_j} = \sum_{j=1}^p \mu_i^T (\mathbf{Y}_{t_j} - \mathbf{Y}_{t_{j-1}}) = \int_0^t \mathbf{b}_s^T \mathrm{d}\mathbf{Y}_s$$

where

$$\mu_p = \mathbf{k}_p, \quad \mu_{p-1} = \mathbf{k}_p + \mathbf{k}_{p-1}, \ldots, \quad \mu_1 = \mathbf{k}_p + \cdots \mathbf{k}_1.$$

and

$$\mathbf{b}_s = \begin{cases} \mu_j & \text{for } t \in (t_{j-1}, t_j) \\ 0 & \text{for } t \in (t_p, T) \end{cases}$$

Hence

$$\tilde{E}\left(a \exp\left(i \int_0^t \mathbf{b}_s \, \mathrm{d}\mathbf{Y}_s \right)\right) = \tilde{\mathbb{E}}\left(a \exp\left(i \sum_{j=1}^p \mathbf{k}_j^T \mathbf{Y}_{t_j} \right)\right) = 0$$

and the same holds for linear combinations by the linearity of $\tilde{\mathbb{E}}$ viz

$$\tilde{\mathbb{E}}\left(a \sum_{k=1}^K c_k \exp\left(i \sum_{j=1}^p \mathbf{k}_{j,k}^T \mathbf{Y}_{t_j} \right)\right) = 0$$

where $c_1, \ldots, c_K$ are complex coefficients. If $F(\mathbf{x}_1, \ldots, \mathbf{x}_p)$ is a continuous bounded complex valued function defined on $(\mathbb{R}^m)^p$ then since the set

$$\left\{ f(\mathbf{x}_1, \ldots, \mathbf{x}_p) = \sum_{k=1}^{K} c_k \exp\left(i \sum_{j=1}^{p} \mathbf{k}_{j,k}^T \mathbf{x}_j \right) \ : \ c_k \in \mathbb{C}\ \forall k,\ \mathbf{k}_{j,k} \in \mathbb{C}\ \forall j,k \right\}$$

is closed under complex conjugation, and separates points in $(\mathbb{R}^m)^p$, the Stone-Weierstrass approximation theorem for complex valued functions implies that there exists a uniformly bounded sequence of functions of the form

$$P^{(n)}(\mathbf{x}_1, \ldots, \mathbf{x}_p) = \sum_{k=1}^{K^{(n)}} c_k^{(n)} \exp\left(i \sum_{j=1}^{p} \left(\mathbf{k}_{j,k}^{(n)}\right)^T \mathbf{x}_j \right)$$

such that

$$\lim_{n\to\infty} P^{(n)}(\mathbf{x}_1, \ldots, \mathbf{x}_p) = F(\mathbf{x}_1, \ldots, \mathbf{x}_p)$$

hence

$$\tilde{\mathbb{E}}\left(aF(\mathbf{Y}_{t_1}, \ldots, \mathbf{Y}_{t_p}) \right) = 0$$

for every continuous bounded F and by the usual approximation argument this extends to every F bounded Borel measurable with respect to $\sigma(\mathbf{Y}_{t_1}, \ldots, \mathbf{Y}_{t_p})$. As $0 < t_1 < \ldots < t_p < t$ were arbitrary it follows that $\tilde{\mathbb{E}}(af) = 0$ for every f bounded $\mathcal{Y}_t$ measurable; taking $f(\mathbf{x}) = a(\mathbf{x}) \wedge n$ for arbitrary n then implies $\tilde{\mathbb{E}}(a^2 \wedge n) = 0$, whence $a = 0$ $\tilde{\mathbb{P}}$ a.s.. $\square$

Remark
As a result if $\phi \in L^1(\Omega, \mathcal{Y}_t, \tilde{\mathbb{P}})$ and we wish to compute $\mathbb{E}\left(\phi|\mathcal{Y}\right)$, it suffices to compute $\mathbb{E}\left(\phi\epsilon_t\right)$ for all $\epsilon_t \in S_t$.

Lemma 18.8.
Let $\{U_t\}_{t\geq 0}$ be a continuous $\mathcal{F}_t$ measurable process such that

$$\tilde{\mathbb{E}}\left(\int_0^T U_t^2 \,\mathrm{d}t \right) < \infty, \quad \forall T \geq 0,$$

then for any $t \geq 0$, and for $j = 1, \ldots, m$ the following holds

$$\tilde{\mathbb{E}}\left(\int_0^t U_s \,\mathrm{d}Y_s^j \,\middle|\, \mathcal{Y}_t \right) = \int_0^t \tilde{\mathbb{E}}\left(U_s|\mathcal{Y}_s\right) \mathrm{d}Y_s^j.$$

Proof
As S_t is a total set in $L^1(\Omega, \mathcal{Y}_t, \tilde{\mathbb{P}})$, it is sufficient to prove that for any $\epsilon_t \in S_t$ that

$$\tilde{\mathbb{E}}\left(\epsilon_t \int_0^t U_s \mathrm{d}Y_s^j \right) = \tilde{\mathbb{E}}\left(\epsilon_t \int_0^t \tilde{\mathbb{E}}\left(U_s|\mathcal{Y}_s\right) \mathrm{d}Y_s^j \right).$$

To this end note that for every $\epsilon_t \in S_t$ the following holds

$$\epsilon_t = 1 + \int_0^t i\epsilon_s \mathbf{r}_s^T \mathrm{d}\mathbf{Y}_s,$$

and hence

$$\begin{aligned}
\tilde{\mathbb{E}}\left(\epsilon_t \int_0^t U_s \mathrm{d}Y_s^j\right) &= \tilde{\mathbb{E}}\left\{\left(1 + \int_0^t i\epsilon_s \mathbf{r}_s^T \mathrm{d}\mathbf{Y}_s\right)\left(\int_0^t U_s \mathrm{d}Y_s^j\right)\right\} \\
&= \tilde{E}\left(\int_0^t U_s \mathrm{d}Y_s^j\right) + \tilde{\mathbb{E}}\left(\int_0^t i\epsilon_s r_s^j U_s \mathrm{d}s\right).
\end{aligned}$$

Here the last term is computed by recalling the definition of the covariation process, in conjunction with the Kunita Watanabe identity. The first term on the right hand side above is a martingale (from the boundedness condition on U_s) and hence its expectation vanishes. Thus using the fact that the lebesgue integral and the expectation $\tilde{\mathbb{E}}(\cdot|\mathcal{Y})$ commute

$$\begin{aligned}
\tilde{\mathbb{E}}\left(\epsilon_t \int_0^t U_s \mathrm{d}Y_s^j\right) &= \tilde{\mathbb{E}}\left(\int_0^t i\epsilon_s r_s^j U_s \mathrm{d}s\right) \\
&= \tilde{\mathbb{E}}\left(\tilde{\mathbb{E}}\left(\int_0^t i\epsilon_s r_s^j U_s \mathrm{d}s \middle| \mathcal{Y}\right)\right) \\
&= \tilde{\mathbb{E}}\left(\int_0^t i\tilde{\mathbb{E}}\left(\epsilon_s r_s^j U_s \middle| \mathcal{Y}\right) \mathrm{d}s\right) \\
&= \tilde{\mathbb{E}}\left(\int_0^t i\epsilon_s r_s^j \tilde{\mathbb{E}}\left(U_s|\mathcal{Y}\right) \mathrm{d}s\right) \\
&= \tilde{\mathbb{E}}\left(\int_0^t i\epsilon_s \mathbf{r}_s^T \mathrm{d}\mathbf{Y}_s \int_0^t \tilde{\mathbb{E}}\left(U_s|\mathcal{Y}\right) \mathrm{d}Y_s^j\right) \\
&= \tilde{\mathbb{E}}\left(\epsilon_t \int_0^t \tilde{\mathbb{E}}\left(U_s|\mathcal{Y}\right) \mathrm{d}Y_s^j\right).
\end{aligned}$$

But from proposition 18.4, it follows that since U_s is $\mathcal{Y}_s$ measurable $\tilde{\mathbb{E}}(U_s|\mathcal{Y}) = \tilde{\mathbb{E}}(U_s|\mathcal{Y}_s)$ whence

$$\tilde{\mathbb{E}}\left(\epsilon_t \int_0^t U_s \mathrm{d}Y_s^j\right) = \tilde{\mathbb{E}}\left(\epsilon_t \int_0^t \tilde{\mathbb{E}}\left(U_s|\mathcal{Y}_s\right) \mathrm{d}Y_s^j\right).$$

As this holds for all ϵ_t in S_t, and the latter is a total set, by the earlier remarks, this establishes the result. □

Lemma 18.9.
Let $\{U_t\}_{t\geq 0}$ be a real valued $\mathcal{F}_t$ adapted continuous process such that

$$\tilde{\mathbb{E}}\left(\int_0^T U_s^2 \,\mathrm{d}s\right) < \infty, \quad \forall T \geq 0, \tag{14}$$

and let $\mathbf{R}_t$ be another continuous $\mathcal{F}_t$ adapted process such that $\langle \mathbf{R}, Y\rangle_t = \mathbf{0}$ for all t. Then

$$\tilde{\mathbb{E}}\left[\int_0^t U_s \mathrm{d}\mathbf{R}_s | \mathcal{Y}_t\right] = \mathbf{0}.$$

Proof
As before use the fact that S_t is a total set, so it suffices to show that for each $\epsilon_t \in S_t$

$$\tilde{\mathbb{E}}\left(\epsilon_t \int_0^t U_s \mathrm{d}\mathbf{R}_s\right) = \mathbf{0}.$$

As in the previous proof we note that each ϵ_t in S_t satisfies

$$\epsilon_t = 1 + \int_0^t i\epsilon_s \mathbf{r}_s^T \mathrm{d}\mathbf{Y}_s$$

Hence substituting this into the above expression yields

$$\begin{aligned}
\tilde{\mathbb{E}}\left(\epsilon_t \int_0^t U_s \mathrm{d}\mathbf{R}_s\right) =& \tilde{\mathbb{E}}\left(\int_0^t U_s \mathrm{d}\mathbf{R}_s\right) + \tilde{\mathbb{E}}\left[\left(\int_0^t i\epsilon_s \mathbf{r}_s^T \mathrm{d}\mathbf{Y}_s\right)\left(\int_0^t U_s \mathrm{d}\mathbf{R}_s\right)\right] \\
=& \tilde{\mathbb{E}}\left(\int_0^t U_s \mathrm{d}\mathbf{R}_s\right) + \tilde{\mathbb{E}}\left(\int_0^t i\epsilon_s U_s \mathbf{r}_s^T \mathrm{d}\langle \mathbf{Y}, \mathbf{R}\rangle_s\right) \\
=& \tilde{\mathbb{E}}\left(\int_0^t U_s \mathrm{d}\mathbf{R}_s\right) \\
=& \mathbf{0}.
\end{aligned}$$

The last term vanishes, because by the condition 14, the stochastic integral is a genuine martingale. The other term vanishes because from the hypotheses $\langle \mathbf{Y}, \mathbf{R}\rangle_t = 0$. □

Remark
In the context of stochastic filtering a natural application of this lemma will be made by setting $\mathbf{R}_t = \mathbf{V}_t$, the stochastic noise process driving the signal process.

We are now in a position to state and prove the Zakai equation. The Zakai equation is important because it is a parabolic stochastic partial differential equation which is satisfied by ρ_t, and indeed it's solution provides a practical means of computing $\rho_t(\phi) = \tilde{\mathbb{E}}\left(\phi(\mathbf{X}_t)\tilde{Z}_t | \mathcal{Y}_t\right)$. Indeed the Zakai equation provides a method by which numerical solution of the non-linear filtering problem may be approached by using recursive algorithms to solve the stochastic differential equation.

Theorem 18.10.
Let A be the infinitesimal generator of the signal process $\mathbf{X}_t$, let the domain of this infinitesimal generator be denoted $\mathcal{D}(A)$. Then subject to the usual conditions on f, σ and h, the un-normalised conditional distribution on X satisfies the Zakai equation which is

$$\rho_t(\phi) = \pi_0(\phi) + \int_0^t \rho_s(A_s\phi)\mathrm{d}s + \int_0^t \rho_s(\mathbf{h}^T\phi)\mathrm{d}\mathbf{Y}_s, \quad \forall t \geq 0, \quad \forall \phi \in \mathcal{D}(A). \tag{15}$$

Proof
We approximate $\tilde{Z}_t$ by

$$\tilde{Z}_t^\epsilon = \frac{\tilde{Z}_t}{1+\epsilon\tilde{Z}_t},$$

noting that since $Z_t \geq 0$, this approximation is bounded by

$$\tilde{Z}_t^\epsilon \leq \frac{4}{27\epsilon^2}.$$

Thus we have approximated the process Z_t by a bounded process. Application of Itô's formula to $f(x) = x/(1+\epsilon x)$, yields

$$\mathrm{d}\tilde{Z}_t^\epsilon = \mathrm{d}f(\tilde{Z}_t) = (1+\epsilon\tilde{Z}_t)^{-2}\tilde{Z}_t\mathbf{h}^T\mathrm{d}\mathbf{Y}_t - \epsilon(1+\epsilon\tilde{Z}_t)^{-3}\tilde{Z}_t^2\|\mathbf{h}(t,X_t)\|^2.$$

The integration by parts formula yields,

$$\mathrm{d}\left(\tilde{Z}_t^\epsilon\phi(\mathbf{X}_t)\right) = \phi(\mathbf{X}_t)\mathrm{d}\tilde{Z}_t^\epsilon + \tilde{Z}_t^\epsilon\mathrm{d}\left(\phi(\mathbf{X}_t)\right) + \left\langle \tilde{Z}_t^\epsilon, \phi(\mathbf{X}_t)\right\rangle.$$

However from the definition of an infinitesimal generator, we know that

$$\phi(\mathbf{X}_t) - \phi(\mathbf{X}_0) = \int_0^t A_s\phi(\mathbf{X}_s)\mathrm{d}s + M_t^\phi,$$

where M_t^ϕ is a martingale given by

$$M_t^\phi = \sum_{j=1}^d \int_0^t \sum_{i=1}^d \frac{\partial\phi}{\partial x^i}(\mathbf{X}_t)\sigma_{ij}(t,\mathbf{X}_t)\,\mathrm{d}\mathbf{W}_t^j.$$

Using the expression for $\mathrm{d}Z_t^\epsilon$ found earlier we may write

$$\begin{aligned}\mathrm{d}\left(\tilde{Z}_t^\epsilon\phi(\mathbf{X}_t)\right) &= \left[\tilde{Z}_t^\epsilon A\phi(\mathbf{X}_t) - \phi(\mathbf{X}_t)\epsilon(1+\epsilon\tilde{Z}_t)^{-3}\tilde{Z}_t^2\|\mathbf{h}(t,\mathbf{X}_t)\|^2\right]\mathrm{d}t \\ &\quad + \tilde{Z}_t^\epsilon\mathrm{d}M_t^\phi + \phi(\mathbf{X}_t)(1+\epsilon\tilde{Z}_t)^{-2}\tilde{Z}_t\mathbf{h}^T(t,\mathbf{X}_t)\mathrm{d}\mathbf{Y}_t\end{aligned}$$

In terms of the functions already defined we can write the infinitesimal generator for $\mathbf{X}_t$ as

$$A_s\phi(\mathbf{x}) = \sum_{i=1}^d \mathbf{f}_i(s,\mathbf{x})\frac{\partial\phi}{\partial x^i}(\mathbf{x}) + \frac{1}{2}\sum_{i=1}^d\sum_{j=1}^d\left(\sum_{k=1}^d \sigma_{ik}(s,\mathbf{x})\sigma_{kj}(s,\mathbf{x})\right)\frac{\partial^2\phi}{\partial x^i\partial x^j}(\mathbf{x}),$$

Now we wish to compute

$$\rho_t(\phi) = \tilde{\mathbb{E}}\left(\tilde{Z}_t^\epsilon\phi(\mathbf{X}_t)|\mathcal{Y}_t\right)$$

we shall do this using the previous expression which on integration gives us

$$\begin{aligned}\tilde{Z}^\epsilon_t\phi(\mathbf{X}_t) =& \tilde{Z}^\epsilon_0\phi(\mathbf{X}_0) + \int_0^t \left[\tilde{Z}^\epsilon_s A\phi(\mathbf{X}_s) - \epsilon\phi(\mathbf{X}_t)(1+\epsilon\tilde{Z}_s)^{-3}\tilde{Z}^2_s\|\mathbf{h}(s,\mathbf{X}_s)\|^2\right]\mathrm{d}s \\ &+ \int_0^t \tilde{Z}^\epsilon_s \mathrm{d}M^\phi_s + \int_0^t \phi(\mathbf{X}_s)(1+\epsilon\tilde{Z}_s)^{-2}\tilde{Z}_s\mathbf{h}^T(s,\mathbf{X}_s)\mathrm{d}\mathbf{Y}_s \qquad (16)\end{aligned}$$

Note that

$$\tilde{\mathbb{E}}\left(\tilde{Z}^\epsilon_0\phi(\mathbf{X}_0)\middle|\mathcal{Y}_t\right) = \tilde{\mathbb{E}}\left(\frac{1}{1+\epsilon}\phi(\mathbf{X}_0)\middle|\mathcal{Y}_0\right) = \pi_0(\phi)\frac{1}{1+\epsilon}.$$

Hence taking conditional expectations of both sides of (16)

$$\begin{aligned}\tilde{\mathbb{E}}\left(\tilde{Z}^\epsilon_t\phi(\mathbf{X}_t)\middle|\mathcal{Y}_t\right) =& \frac{\pi_0(\phi)}{1+\epsilon} + \tilde{\mathbb{E}}\left[\int_0^t \phi(\mathbf{X}_s)(1+\epsilon\tilde{Z}_s)^{-2}\tilde{Z}_s\mathbf{h}^T(s,\mathbf{X}_s)\,\mathrm{d}\mathbf{Y}_s\middle|\mathcal{Y}_t\right] \\ &+ \tilde{\mathbb{E}}\left[\int_0^t \left(\tilde{Z}^\epsilon_s A\phi(\mathbf{X}_s) - \epsilon\phi(\mathbf{X}_s)(1+\epsilon\tilde{Z}_s)^{-3}\tilde{Z}^2_s\|\mathbf{h}(s,\mathbf{X}_s)\|^2\right)\mathrm{d}s\middle|\mathcal{Y}_t\right] \\ &+ \tilde{\mathbb{E}}\left[\int_0^t \tilde{Z}^\epsilon_s\,\mathrm{d}M^\phi_s\middle|\mathcal{Y}_t\right]\end{aligned}$$

Applying lemma 18.8 since $\tilde{Z}^\epsilon_t$ is bounded allows us to interchange the (stochastic) integrals and the conditional expectations to give

$$\begin{aligned}\tilde{\mathbb{E}}\left(\tilde{Z}^\epsilon_t\phi(\mathbf{X}_t)\middle|\mathcal{Y}_t\right) =& \frac{\pi_0(\phi)}{1+\epsilon} + \int_0^t \tilde{\mathbb{E}}\left(\phi(\mathbf{X}_s)(1+\epsilon\tilde{Z}_s)^{-2}\tilde{Z}_s\mathbf{h}^T(s,\mathbf{X}_s)\middle|\mathcal{Y}_t\right)\mathrm{d}\mathbf{Y}_s \\ &+ \int_0^t \tilde{\mathbb{E}}\left[\tilde{Z}^\epsilon_s A\phi(\mathbf{X}_s) - \epsilon\phi(\mathbf{X}_s)(1+\epsilon\tilde{Z}_s)^{-3}\tilde{Z}^2_s\|\mathbf{h}(s,\mathbf{X}_s)\|^2\middle|\mathcal{Y}_t\right]\mathrm{d}s \\ &+ \int_0^t \tilde{\mathbb{E}}\left[\tilde{Z}^\epsilon_s\middle|\mathcal{Y}_t\right]\mathrm{d}M^\phi_s\end{aligned}$$

Now we take the $\epsilon \downarrow 0$ limit. Clearly $\tilde{Z}^\epsilon_t \to \tilde{Z}_t$ pointwise, and thus by the monotone convergence theorem since $Z^\epsilon_t \uparrow Z_t$ as $\epsilon \to 0$,

$$\tilde{\mathbb{E}}\left(\tilde{Z}^\epsilon_t\phi(\mathbf{X}_t)|\mathcal{Y}_t\right) \to \tilde{\mathbb{E}}\left(\tilde{Z}_t\phi(\mathbf{X}_t)|\mathcal{Y}_t\right) = \rho_t(\phi),$$

and also by the monotone convergence theorem

$$\tilde{\mathbb{E}}\left(\tilde{Z}^\epsilon_s A_s\phi(\mathbf{X}_s)|\mathcal{Y}_s\right) \to \rho_s(A_s\phi).$$

We can bound this by

$$\forall\epsilon > 0, \quad \tilde{\mathbb{E}}\left(Z^\epsilon_s A_s\phi(\mathbf{X}_s)|\mathcal{Y}_s\right) \le \|A_s\phi\|\tilde{\mathbb{E}}\left(\tilde{Z}_s|\mathcal{Y}_s\right),$$

and the right hand side is an L^1 bounded quantity. Hence by the dominated convergence theorem we have

$$\int_0^t \mathbb{E}\left(\tilde{Z}_s^\epsilon A_s\phi(\mathbf{X}_s)|\mathcal{Y}_s\right)\,\mathrm{d}s \to \int_0^t \rho_s(A_s\phi)\,\mathrm{d}s \text{ a.s.}$$

Also note that

$$\lim_{\epsilon\to 0} \epsilon\phi(\mathbf{X}_s)\left(\tilde{Z}_s^\epsilon\right)^2\left(1+\epsilon\tilde{Z}_s\right)^{-1}\|\mathbf{h}(s,\mathbf{X}_s)\|^2 = 0.$$

and since the function may be dominated by

$$\begin{aligned}\left|\epsilon\phi(\mathbf{X}_s)\left(\tilde{Z}_s^\epsilon\right)^2\left(1+\epsilon\tilde{Z}_s\right)^{-1}\|\mathbf{h}(s,\mathbf{X}_s)\|^2\right| &\le \epsilon\|\phi\|_\infty^2\frac{\epsilon\tilde{Z}_s}{1+\epsilon\tilde{Z}_s}\frac{Z_s}{1+\epsilon\tilde{Z}_s}\frac{1}{1+\epsilon\tilde{Z}_s}\|\mathbf{h}(s,\mathbf{X}_s)\|^2\\ &\le \tilde{Z}_s\|h(s,\mathbf{X}_s)\|^2\\ &\le \tilde{Z}_s k(1+\|X_s\|^2)\end{aligned}$$

But

$$\begin{aligned}\tilde{\mathbb{E}}(\tilde{Z}_s k(1+\|X_s\|^2)) &= k\mathbb{E}\left(Z_s\tilde{Z}_s\left(1+\|X_s\|^2\right)\right)\\ &= k\left(1+\mathbb{E}(\|X_s\|^2)\right)\\ &\le k\left(1+C\left(1+\mathbb{E}(\|\mathbf{X}_0\|^2)\right)e^{ct}\right) \qquad (17)\end{aligned}$$

Where the final inequality follows by (6). As a consequence of these two results, as the right hand side of the above inequality is integrable, by the conditional form of the Dominated Convergence theorem it follows that

$$\tilde{E}\left(\epsilon\phi(\mathbf{X}_s)\left(\tilde{Z}_s^\epsilon\right)^2\left(1+\epsilon\tilde{Z}_s\right)^{-1}\|\mathbf{h}(s,\mathbf{X}_s)\|^2\middle|\mathcal{Y}\right) \to 0$$

as $\epsilon \to 0$. By proposition 18.4 the sigma field $\mathcal{Y}$ may be replaced by $\mathcal{Y}_t$ which yields that as $\epsilon \to 0$

$$\tilde{E}\left(\epsilon\phi(\mathbf{X}_s)\left(\tilde{Z}_s^\epsilon\right)^2\left(1+\epsilon\tilde{Z}_s\right)^{-1}\|\mathbf{h}(s,\mathbf{X}_s)\|^2\middle|\mathcal{Y}_s\right) \to 0.$$

Now note again that the dominating function (17) is also Lebesgue integrable and bounded thus a second application of the Dominated convergence theorem yields

$$\int_0^t \epsilon\tilde{\mathbb{E}}\left[\phi(\mathbf{X}_s)\left(\tilde{Z}_s^\epsilon\right)^2\left(1+\epsilon\tilde{Z}_s\right)^{-1}\|\mathbf{h}(s,\mathbf{X}_s)\|^2\middle|\mathcal{Y}_s\right]\mathrm{d}s \to 0, \quad \text{as } \epsilon\to 0.$$

It now remains to check that the stochastic integral

$$\int_0^t \tilde{\mathbb{E}}\left[\phi(\mathbf{X}_s)\tilde{Z}_s^\epsilon\left(1+\epsilon\tilde{Z}_s\right)^{-1}\mathbf{h}^T(s,\mathbf{X}_s)\middle|\mathcal{Y}_s\right]\,\mathrm{d}\mathbf{Y}_s \to \int_0^t \rho_s(\mathbf{h}^T\phi(\mathbf{X}_s))\,\mathrm{d}\mathbf{Y}_s.$$

To this end define

$$I_\epsilon(t) := \int_0^t \tilde{\mathbb{E}}\left[\phi(\mathbf{X}_s)\tilde{Z}_s^\epsilon\left(1+\epsilon\tilde{Z}_s\right)^{-1}\mathbf{h}^T(s,\mathbf{X}_s)\middle|\,\mathcal{Y}_s\right]\mathrm{d}\mathbf{Y}_s,$$

and the desired limit

$$I(t) := \int_0^t \rho_s(\mathbf{h}^T\phi(\mathbf{X}_s))\mathrm{d}\mathbf{Y}_s = \int_0^t \tilde{\mathbb{E}}\left(\tilde{Z}_s\mathbf{h}^T(s,\mathbf{X}_s)\phi(\mathbf{X}_s)\middle|\,\mathcal{Y}_t\right)\mathrm{d}\mathbf{Y}_s.$$

Both $I(t)$ and $I_\epsilon(t)$ are continuous bounded square integrable martingales. Consider the difference

$$\begin{aligned} I_\epsilon(t) - I(t) &= \int_0^t \tilde{\mathbb{E}}\left[\left(\frac{\tilde{Z}_s}{\left(1+\epsilon\tilde{Z}_s\right)^2} - \tilde{Z}_s\right)\phi(\mathbf{X}_s)\mathbf{h}^T(s,\mathbf{X}_s)\middle|\,\mathcal{Y}\right]\mathrm{d}\mathbf{Y}_s \\ &= -\int_0^t \tilde{\mathbb{E}}\left[\frac{\epsilon\tilde{Z}_s^2\left(2+\epsilon\tilde{Z}_s\right)}{\left(1+\epsilon\tilde{Z}_s\right)^2}\phi(\mathbf{X}_s)\mathbf{h}^T(s,\mathbf{X}_s)\middle|\,\mathcal{Y}\right]\mathrm{d}\mathbf{Y}_s, \end{aligned}$$

Consider the L^2 norm of this difference and use the fact that Y is a $\tilde{\mathbb{P}}$ Brownian motion

$$\begin{aligned} \tilde{\mathbb{E}}\left(I_\epsilon(t) - I(t)\right)^2 &= \tilde{\mathbb{E}}\left(\int_0^t \tilde{\mathbb{E}}\left[\frac{\epsilon\tilde{Z}_s^2\left(2+\epsilon\tilde{Z}_s\right)}{\left(1+\epsilon\tilde{Z}_s\right)^2}\phi(\mathbf{X}_s)\mathbf{h}^T(s,\mathbf{X}_s)\middle|\,\mathcal{Y}\right]\mathrm{d}\mathbf{Y}_s\right)^2 \\ &= \tilde{\mathbb{E}}\left(\int_0^t \left\|\tilde{\mathbb{E}}\left[\frac{\epsilon\tilde{Z}_s^2\left(2+\epsilon\tilde{Z}_s\right)}{\left(1+\epsilon\tilde{Z}_s\right)^2}\phi(\mathbf{X}_s)\mathbf{h}^T(s,\mathbf{X}_s)\middle|\,\mathcal{Y}\right]\right\|^2\mathrm{d}s\right) \end{aligned}$$

By the conditional form of Jensen's inequality $\|\tilde{\mathbb{E}}(X|\mathcal{Y})\|^2 \le \tilde{\mathbb{E}}\left(\|X\|^2\middle|\,\mathcal{Y}\right)$. Thus

$$\begin{aligned} &\left\|\tilde{\mathbb{E}}\left[\frac{\epsilon\tilde{Z}_s^2\left(2+\epsilon\tilde{Z}_s\right)}{\left(1+\epsilon\tilde{Z}_s\right)^2}\phi(\mathbf{X}_s)\mathbf{h}^T(s,\mathbf{X}_s)\middle|\,\mathcal{Y}\right]\right\|^2 \\ &\qquad\qquad \le \tilde{\mathbb{E}}\left(\left\|\frac{\epsilon\tilde{Z}_s^2\left(2+\epsilon\tilde{Z}_s\right)}{\left(1+\epsilon\tilde{Z}_s\right)^2}\phi(\mathbf{X}_s)\mathbf{h}^T(s,\mathbf{X}_s)\right\|^2\middle|\,\mathcal{Y}\right) \end{aligned}$$

A dominating function can can readily be found for the quantity inside the conditional expectation viz

$$\begin{aligned} \left\|\frac{\epsilon\tilde{Z}_s^2\left(2+\epsilon\tilde{Z}_s\right)}{\left(1+\epsilon\tilde{Z}_s\right)^2}\phi(\mathbf{X}_s)\mathbf{h}^T(s,\mathbf{X}_s)\right\|^2 &\le \|\phi\|_\infty^2\left|\frac{\epsilon\tilde{Z}_s}{1+\epsilon\tilde{Z}_s}\left(1+\frac{1}{1+\epsilon\tilde{Z}_s}\right)\tilde{Z}_s\right|^2\|\mathbf{h}(s,\mathbf{X}_s)\|^2 \\ &\le 4\|\phi\|_\infty^2\tilde{Z}_s^2\|\mathbf{h}(s,\mathbf{X}_s)\|^2 \\ &\le 4k\|\phi\|_\infty^2\tilde{Z}_s^2\left(1+\|\mathbf{X}_s\|^2\right) \end{aligned}$$

Since

$$\tilde{\mathbb{E}}\left(\tilde{Z}_s\left(\tilde{Z}_s+\tilde{Z}_s\|\mathbf{X}_s\|^2\right)\right)=\mathbb{E}\left(\tilde{Z}_s\right)+\mathbb{E}\left(\tilde{Z}_s\|\mathbf{X}_s\|^2\right)$$

Using the dominated convergence theorem as $\epsilon \to 0$

$$\int_0^t \left\| \tilde{\mathbb{E}}\left[\frac{\epsilon \tilde{Z}_s^2\left(2+\epsilon\tilde{Z}_s\right)}{\left(1+\epsilon\tilde{Z}_s\right)^2}\phi(\mathbf{X}_s)\mathbf{h}^T(s,\mathbf{X}_s)\middle| \mathcal{Y}_t\right]\right\|^2 \mathrm{d}s \to 0$$

Now we use the dominated convergence theorem to show that for a suitable subsequence that since $\|\mathbf{h}(s,\mathbf{X}_s)\|^2 \le K(1+\|\mathbf{X}_s\|^2)$ we can write

$$\tilde{\mathbb{E}}\left[\int_0^t \left(\tilde{\mathbb{E}}\left[\phi(\mathbf{X}_s)\frac{\epsilon \tilde{Z}_s^2\left(2+\epsilon\tilde{Z}_s\right)}{\left(1+\epsilon\tilde{Z}_s\right)^2}\mathbf{h}^T(s,\mathbf{X}_s)|\mathcal{Y}_s\right]\right)^2 \mathrm{d}s\right] \to 0. \ a.s.$$

Thus we have shown that

$$\tilde{\mathbb{E}}\left[(I_\epsilon(t)-I(t))^2\right] \to 0.$$

By a standard theorem of convergence, this means that there is a suitable subsequence ϵ_n such that

$$I_{\epsilon_n}(t) \to I(t) \ a.s.$$

which is sufficient to establish the claim. □

18.7. Kushner-Stratonowich Equation

The Zakai equation which has been derived in the previous section is an SDE which is satisfied by the unnormalised conditional distribution of $\mathbf{X}_t$, i.e. $\rho_t(\phi)$. It seems natural therefore to derive a similar equation satisfied by the normalised conditional distribution, $\pi_t(\phi)$. Recall that the unnormalised conditional distribution is given by

$$\rho_t(\phi)=\tilde{\mathbb{E}}\left[\phi(\mathbf{X}_t)\tilde{Z}_t|\mathcal{Y}_t\right],$$

and the normalised conditional law is given by

$$\pi_t(\phi):=\mathbb{E}\left[\phi(\mathbf{X}_t)|\mathcal{Y}_t\right].$$

As a consequence of the proposition 17.5

$$\pi_t(\phi)=\frac{\rho_t(\phi)}{\rho_t(1)},$$

Now using this result together with the Zakai equation, we can derive the Kushner-Stratonovitch equation.

Theorem (Kushner-Stratonovich) 18.11.
The normalised conditional law of the process $\mathbf{X}_t$ subject to the usual conditions on the processes $\mathbf{X}_t$ and $\mathbf{Y}_t$ satisfies the Kushner-Stratonovich equation

$$\pi_t(\phi)=\pi_0(\phi)+\int_0^t \pi_s(A_s\phi)\mathrm{d}s+\int_0^t\left(\pi_s(\mathbf{h}^T\phi)-\pi_s(\mathbf{h}^T)\pi_s(\phi)\right)(\mathrm{d}\mathbf{Y}_s-\pi_s(\mathbf{h})\mathrm{d}s). \quad (18)$$

The process $\mathbf{Y}_t-\int_0^t \pi_s(\mathbf{h}(s,\mathbf{X}_s))\,\mathrm{d}s$ is called the *innovations process.*

19. Gronwall's Inequality

An important and frequently used result in the theory of stochastic differential equations is Gronwall's lemma. There are various versions of Gronwall's theorem (see [Ethier and Kurtz, 1986], [Karatzas and Shreve, 1987]) but the following general form and proof follows that given in [Mandelbaum et al., 1998].

Theorem (Gronwall).
Let x, y and z be measurable non-negative functions on the real numbers. If y is bounded and z is integrable on $[0,T]$, and for all $0 \leq t \leq T$ then

$$x(t) \leq z(t) + \int_0^t x(s)y(s)\mathrm{d}s, \tag{19}$$

then

$$x(t) \leq z(t) + \int_0^t z(s)y(s)\exp\left(\int_s^t y(r)\mathrm{d}r\right)\mathrm{d}s.$$

Proof
Multiplying both sides of the inequality (19) by $y(t)\exp\left(-\int_0^t y(s)\mathrm{d}s\right)$ yields

$$x(t)y(t)\exp\left(-\int_0^t y(s)\mathrm{d}s\right) - \left(\int_0^t x(s)y(s)\mathrm{d}s\right)y(t)\exp\left(-\int_0^t y(s)\mathrm{d}s\right) \\ \leq z(t)y(t)\exp\left(-\int_0^t y(s)\mathrm{d}s\right).$$

The left hand side can be written as the derivative of a product,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\left(\int_0^t x(s)y(s)\mathrm{d}s\right)\exp\left(-\int_0^t y(s)\mathrm{d}s\right)\right] \leq z(t)y(t)\exp\left(-\int_0^t y(s)\mathrm{d}s\right).$$

This can now be integrated to give

$$\left(\int_0^t x(s)y(s)\mathrm{d}s\right)\exp\left(-\int_0^t y(s)\mathrm{d}s\right) \leq \int_0^t z(s)y(s)\exp\left(-\int_0^s y(r)\mathrm{d}r\right)\mathrm{d}s,$$

or equivalently

$$\int_0^t x(s)y(s)\mathrm{d}s \leq \int_0^t z(s)y(s)\exp\left(\int_s^t y(r)\mathrm{d}r\right)\mathrm{d}s.$$

Combining this with the original equation (19) gives the desired result. □

.

Corollary 19.1.
If x, y, and z satisfy the conditions for Gronwall's theorem then

$$\sup_{0\le t\le T} x(t) \le \left(\sup_{0\le t\le T} z(t)\right)\exp\left(\int_0^T y(s)\mathrm{d}s\right).$$

Proof
From the conclusion of Gronwall's theorem we see that for all $0 \le t \le T$

$$x(t) \le \sup_{0\le t\le T} z(t) + \int_0^T z(s)y(s)\exp\left(\int_s^T y(r)\mathrm{d}r\right)\mathrm{d}s,$$

which yields

$$\begin{aligned}
\sup_{0\le t\le T} x(t) &\le \sup_{0\le t\le T} z(t) + \int_0^T z(s)y(s)\exp\left(\int_s^T y(r)\mathrm{d}r\right)\mathrm{d}s,\\
&\le \sup_{0\le t\le T} z(t) + \left(\sup_{0\le t\le T} z(t)\right)\int_0^T y(s)\exp\left(\int_s^T y(r)\mathrm{d}r\right)\mathrm{d}s\\
&\le \sup_{0\le t\le T} z(t) + \left(\sup_{0\le t\le T} z(t)\right)\left(\exp\left(\int_0^T y(s)\mathrm{d}s\right) - 1\right)\\
&\le \left(\sup_{0\le t\le T} z(t)\right)\exp\left(\int_0^T y(s)\mathrm{d}s\right).
\end{aligned}$$

$\square$

20. Kalman Filter

It is clear that in the general case the SDEs given by the Kushner-Stratonivich equation are infinite dimensional. A special case in which they are of finite dimension is the most famous example of stochastic filtering is undoubtedly the *Kalman-Bucy filter*. The relevant equations can be derived without using the general machinery of stochastic filtering theory. The first practical use of the Kalman filter was in the flight guidance computer for the Apollo missions, long before the general non-linear stochastic filtering theory had been developed. However by examining this special linear case, it is hoped that the operation of the Kushner-Stratonowich equation may be made clearer.

We consider the system $(\mathbf{X}_t, \mathbf{Y}_t) \in \mathbb{R}^N \times \mathbb{R}^M$ where $\mathbf{X}_t$ is the signal process and $\mathbf{Y}_t$ is the observation process, described by the following SDEs,

$$\begin{aligned}\mathrm{d}\mathbf{X}_t &= (B_t \mathbf{X}_t + \mathbf{b}_t)\,\mathrm{d}t + F_t \mathrm{d}\mathbf{V}_t \\ \mathrm{d}\mathbf{Y}_t &= H_t \mathbf{X}_t \mathrm{d}t + \mathrm{d}\mathbf{W}_t\end{aligned}$$

where B_t and F_t are a $N \times N$ matrix valued processes, $\mathbf{V}_t$ is a N dimensional Brownian motion, H_t is an $M \times N$ matrix valued process and $\mathbf{W}_t$ is an M dimensional Brownian Motion, which is independent of $\mathbf{V}_t$.

Thus in terms of the notation used earlier

$$\begin{aligned}f(t, \mathbf{X}_t) &= \mathbf{b}_t + B_t \mathbf{X}_t \\ \sigma(t, \mathbf{X}_t) &= F_t \\ \mathbf{h}(t, \mathbf{X}_t) &= H_t \mathbf{X}_t\end{aligned}$$

The infinitesimal generator of the signal process $\mathbf{X}_t$ is given by

$$A_t u = \sum_{i=1}^{N} b_t^{(i)}(\mathbf{X}_t) \frac{\partial u}{\partial x_i} + \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} (FF^T)_{ij} \frac{\partial^2 u}{\partial x_i \partial x_j}$$

which may readily be verified by applying Itô's formula to $f(\mathbf{X}_t)$.

Since the system of equations is linear it is simple to show that the conditional distribution of $\mathbf{X}_t$ conditional upon $\mathcal{Y}_t$ is a multivariate normal distribution. Thus it may be characterized by its first and second moments alone. It is therefore sufficient to consider $\phi(\mathbf{X}) = x_i$ and $\phi(\mathbf{X}) = x_i x_j$ as test functions in the Kushner-Stratonovich equation (18). The notation in what follows will become somewhat messy, we write $X_t^{(i)}$ for the ith component of the signal process.

Since we shall only be concerned with these two moments, it will simplify the notation to define the conditional mean process

$$\hat{X}_t^{(i)} := \pi_t(x_i) = \mathbb{E}\left(X_t^{(i)} \middle| \mathcal{Y}_t \right)$$

and the conditional covariance process

$$R_t^{ij} = \mathbb{E}\left(\left(X_t^{(i)} - \hat{X}_t^{(i)} \right) \left(X_t^{(j)} - \hat{X}_t^{(j)} \right) \right) = \pi_t(x_i x_j) - \pi_t(x_i)\pi_t(x_j)$$

Thus at time t our best estimate of the signal will be a normal distribution $N(\hat{\mathbf{X}}_t, R_t)$.

20.1. Conditional Mean

To derive an equation for the evolution of the conditional mean process, we apply the Kushner-Stratonowich equation to $\phi(\mathbf{X}) = X^{(i)}$, from theorem 18.11 this satisfies

$$\hat{X}_t^{(i)} = \hat{X}_0^{(i)} + \int_0^t \mathbb{E}\left((B_s\mathbf{X}_s)^{(i)} + b_t^{(i)} \middle| \mathcal{Y}_t \right) \mathrm{d}t + \int_0^t \left(\mathbb{E}\left(X_s^{(i)} (H_s\mathbf{X}_s)^T | \mathcal{Y}_s \right) - \hat{X}_s^{(i)} (H_s\hat{\mathbf{X}}_s)^T \right) \mathrm{d}\mathbf{N}_s$$

where $\mathbf{N}_t$, the innovations process, is

$$\mathbf{N}_t = \mathbf{Y}_t - \int_0^t H_s\hat{\mathbf{X}}_s \,\mathrm{d}s$$

Whence writing the equation for the evolution of the conditional mean in vector form, we obtain (since $R^T = R$)

$$\mathrm{d}\hat{\mathbf{X}}_t = \left(B_t\hat{\mathbf{X}}_t + \mathbf{b}_s \right) \mathrm{d}t + R_t H_t^T (\mathrm{d}\mathbf{Y}_t - H_t\hat{\mathbf{X}}_t \mathrm{d}t) \tag{20}$$

20.2. Conditional Covariance

We now derive an equation for the evolution of the covariance matrix R, but first we shall need a result about the third moment of a multivariate normal distribution.

If $\mathbf{X} \sim N(\hat{\mathbf{X}}, R)$ then

$$\begin{aligned} \mathbb{E}\left(X_t^{(i)} X_t^{(j)} X_t^{(k)} \right) &= \mathbb{E}\left((\hat{X}_t^{(i)} + Z^{(i)})(\hat{X}_t^{(j)} + Z^{(j)})(\hat{X}_t^{(k)} + Z_t^{(k)}) \right) \\ &= \hat{X}_t^{(i)}\hat{X}_t^{(j)}\hat{X}_t^{(k)} + \hat{X}_t^{(i)} R_t^{jk} + \hat{X}_t^{j} R_t^{ik} + \hat{X}_t^{(k)} R_t^{ij} \end{aligned}$$

Applying the Kushner-Stratonovich equation to $\phi(\mathbf{X}) = x_i x_j$ yields

$$\begin{aligned} \mathbb{E}\left(X_t^{(i)} X_t^{(j)} \middle| \mathcal{Y}_t \right) &= \mathbb{E}\left(X_0^{(i)} X_0^{(j)} \middle| \mathcal{Y}_0 \right) \\ &\quad + \int_0^t \mathbb{E}\left((F_t F_t^T)^{ij} + (B_t\mathbf{X}_t + \mathbf{b})^{(i)} X_t^{(j)} + (B_t\mathbf{X}_t + \mathbf{b})^{(j)} X_t^{(i)} \middle| \mathcal{Y}_t \right) \mathrm{d}t \\ &\quad + \int_0^t \left\{ \mathbb{E}(X_t^{(i)} X_t^{(j)} H_t\mathbf{X}_t | \mathcal{Y}_t) - \mathbb{E}(X_t^{(i)} X_t^{(j)} | \mathcal{Y}_t) \mathbb{E}(H_t\mathbf{X}_t | \mathcal{Y}_t) \right\} \cdot \mathrm{d}\mathbf{N}_s \end{aligned} \tag{21}$$

It is clear that

$$\mathrm{d}R^{ij} = \mathrm{d}\left\{ \mathbb{E}(X_t^{(i)} X_t^{(j)} | \mathcal{Y}_t) \right\} - \mathrm{d}\left(\hat{X}_t^{(i)} \hat{X}_t^{(j)} \right) \tag{22}$$

the first term is given by (21), and using Itô's form of the product rule to expand out the second term

$$\mathrm{d}\left(\hat{X}_t^{(i)} \hat{X}_t^{(j)} \right) = \hat{X}^{(i)} \mathrm{d}\hat{X}^{(j)} + \hat{X}^{(j)} \mathrm{d}\hat{X}^{(i)} + \mathrm{d}[\hat{X}^{(i)}, \hat{X}^{(j)}]_t$$

Therefore using (20) we can evaluate this as

$$\mathrm{d}\left(\hat{X}_t^{(i)}\hat{X}_t^{(j)}\right) = \hat{X}_t^{(i)}(B_t\mathbf{X}_t + \mathbf{b}_t)^{(j)}\mathrm{d}t + \hat{X}_t^{(i)}(R_tH_t^T\cdot\mathrm{d}\mathbf{N}_t)^{(j)} + \hat{X}_t^{(j)}(B_t\mathbf{X}_t + \mathbf{b}_t)^{(i)}\mathrm{d}t$$
$$+ \hat{X}_t^{(j)}(R_tH_t^T\cdot\mathrm{d}\mathbf{N}_t)^{(i)} + \left[(R_tH_t^T\cdot\mathrm{d}N_t)^{(i)}, (R_tH_t^T\cdot\mathrm{d}N_t)^{(j)}\right] \tag{23}$$

For evaluating the quadratic covariation term in this expression it is simplest to work componentwise using the Einstein summation convention and the fact that the innovation process $\mathbf{N}_t$ is readily shown by Lévy's characterization to be a $\mathbb{P}$ Brownian motion

$$\begin{aligned}\left[(R_tH_t^T\cdot\mathrm{d}N_t)^{(i)}, (R_tH_t^T\cdot\mathrm{d}N_t)^{(j)}\right] &= \left[R_t^{il}H^{kl}\mathrm{d}N^k, R^{jm}H_{nm}\mathrm{d}N^n\right]\\ &= R^{il}H^{kl}R^{jm}H_{nm}\delta_{kn}\mathrm{d}t\\ &= R^{il}H^{kl}H^{km}R^{jm}\mathrm{d}t\\ &= (RH^THR^T)^{ij}\mathrm{d}t\\ &= (RH^THR)^{ij}\mathrm{d}t \end{aligned}\tag{24}$$

where the last equality follows since $R^T = R$. Substituting (24) into (23) and then using this and (21) in (22) yields the following equation for the evolution of the ijth element of the covariance matrix

$$\mathrm{d}R_t^{ij} = \mathrm{d}t\left(\frac{1}{2}(F_tF_t^T)^{ij} + (B_tR_t)^{ij} + (R_t^TB_t^T)^{ij} - (RH^THR)^{ij}\right)$$
$$+ (\hat{X}_t^{(i)}R_t^{jm} + \hat{X}_t^{(j)}R_t^{jm})H_{lm}\mathrm{d}N^l - (\hat{X}_t^iR_t^{jm} + \hat{X}_t^jR_t^{jm})H^{lm}\mathrm{d}N^l$$

Thus we obtain the final differential equation for the evolution of the conditional covariance matrix; notice that all of the stochastic terms have cancelled

$$\frac{\mathrm{d}R_t}{\mathrm{d}t} = F_tF_t^T + B_tR_t + R_tB_t^T - R_tH_t^TH_tR_t \tag{25}$$

The two equations (20) and (25) therefore provide a complete description of the best prediction of the system state $\mathbf{X}_t$, conditional on the observations up to time t. This is the *Kalman-Bucy* filter.

It is most frequently encountered in a discrete time version, since in this form it is practical for use in some of the problems described in the introduction to the stochastic filtering section of these notes.

21. Discontinuous Stochastic Calculus

So far we have discussed integrals with repect to continuous semimartingales. It might seem that this includes a great deal of the cases which are of real interest, but a very important class of processes have been neglected.

The standard Poisson process on the line should be familiar. This has trajectories $p(t)$ which are $\mathcal{F}_t$-adapted and piecewise continuous. These have the properties that for any $t \geq 0$, $p(t)$ is a.s. finite and for any $t, s \geq 0$, $p(t+s) - p(t)$ is independent of $\mathcal{F}_t$. The process is also stationary – the distribution of $p(t+s) - p(t)$ does not depend upon t. If the probability of a jump in an infinitessimal time interval is proportional to the length of the time interval, the constant of proportionality is called the intensity λ and it can be shown that the inter jump times have exponential distribution with parameter λ (i.e. mean $1/\lambda$).

Such processes can also be described by their jump measures. The jump measure $N(\sigma, B)$ is the number of jumps within the time interval σ in the (Borel) set B. From the above definition for any $B \in \mathcal{B}$,

$$\mathbb{E}\left(N([0, t+s], B) - N([0, t], B) \,|\mathcal{F}_t\,\right) = \lambda \Pi(B),$$

where Π is a probability measure. More formally, we make the following definition

Definition 1. **Poisson Process**
A Poisson process with intensity measure $h(\mathrm{d}t, \mathrm{d}y) = \lambda\, \mathrm{d}t \times \mathrm{d}y$ as a measurable map N from $(\Omega, \mathcal{F}, \mathbb{P})$ into the space of positive integer valued counting measures on $(\mathbb{R}^+ \times \Gamma, \mathcal{B}(\mathbb{R}^+ \times \Gamma))$ with the properties

(i) For every $t \geq 0$ and every Borel subset A of $[0, t] \times \Gamma$, $N(A)$ is $\mathcal{F}_{t-}$-measurable.
(ii) For every $t \geq 0$ and every collection of Borel subsets A_i of $[t, \infty) \times \Gamma$, $\{N(A_i)\}$ is independent of $\mathcal{F}_{t-}$.
(iii) $E(N(A)) = h(A)$ for every Borel subset A of $\mathbb{R}^+ \times \Gamma$.
(iv) For each $s > 0$ and Borel $B \subset \Gamma$ the distribution of $N([t, t+s], B)$ does not depend upon t.

This measure can be related back to the path $p(\cdot)$ by the following integral for any $t \geq 0$,

$$p(t) = \int_0^t \int_\Gamma \gamma\, N(\mathrm{d}s, \mathrm{d}\gamma).$$

The above integral can be defined without difficulty!

21.1. Compensators

Let $p(t)$ be an $\mathcal{F}_t$-adapted jump process with bounded jumps, such that $p(0) = 0$ and $\mathbb{E}(p(t)) < \infty$ for all $t \geq 0$. Let $\lambda(t)$ be a bounded non-negative $\mathcal{F}_t$-adapted process with RCLL sample paths. If

$$p(t) - \int_0^t \lambda(s)\, \mathrm{d}s$$

is an $\mathcal{F}_t$-martingale, then $\lambda(\cdot)$ is the *intensity* of the jump process $p(\cdot)$ and the process $\int_0^t \lambda(s)\, \mathrm{d}s$ is the *compensator* of the process $p(\cdot)$.

Why are compensators interesting? Recall from the theory of stochastic integration with respect to continuous semimartingales that one of the most useful properties was that a stochastic integral with respect to a continuous local is a local martingale. We would like the same sort of property to hold when we integrate with respect to a suitable class of discontinuous process.

Definition 21.2.
For A a non-decreasing locally integrable process, there is a non-decreasing previsible process, called the compensator of A, denoted A^p which is a.s. unique characterised by one of the following
(i) $A - A^p$ *is a local martinagle.*
(ii) $\mathbb{E}(A^p_T) = \mathbb{E}(A_T)$ *for all stopping times T.*
(iii) $\mathbb{E}\left[(H \cdot A^p)_\infty\right] = \mathbb{E}\left[(H \cdot A)_\infty\right]$ *for all non-negative previsible processes H.*

Corollary 21.3.
For A a process of locally finite variation there exists a previsible locally finite variation compensator process A^p unique a.s. such that $A - A^p$ is a local martingale.

Example
Consider a Poisson process N_t with mean measure $m_t(\cdot)$. This process has a compensator of m_t. To see this note that $\mathbb{E}(N_t - N_s|\mathcal{F}_s) = m_t - m_s$, for any $s \leq t$. So $N_t - m_t$ is a martingale, and m_t is a previsible (since it is deterministic) process of locally finite variation.

Definition 21.4.
Two local martingales N and M are orthogonal if their product NM is a local martingale.

Definition 21.5.
A local martingale X is purely discontinuous if $X_0 = 0$ and it is orthogonal to all continuous local martingales.

The above definition, although important is somewhat counter intuitive, since for example if we consider a Poisson process N_t which consists soley of jumps, then we can show that $N_t - m_t$ is purely discontinuous despite the fact that this process does not consist solely of jumps!

Theorem 21.6.
Any local martingale M has a unique (up to indistinguishability) decomposition

$$M = M_0 + M^c + M^d,$$

where $M^d_0 = M^c_0 = 0$, M^c is a continuous local martingale and M^d is a purely discontinuous local martingale.

Theorem 21.7.
If f is a real valued C^2 function with domain $\mathbb{R}^n$ and X is an n-dimensional semi-martingale with decomposition $X = M^c + M^d + Y$ where $M^c + M^d$ is a local martingale and M^c is

a continuous local martingale then the following generalization of Itô's formula holds

$$\begin{aligned} f(X_t) =& f(X_0) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(X_{s-})\,\mathrm{d}X_s \\ &+ \frac{1}{2}\sum_{i,j=1}^{n} \int_0^t \frac{\partial^2 f}{\partial x_i \partial x_j}(X_{s-})\,\mathrm{d}[(M^c)^i, (M^c)^j]_s \\ &+ \sum_{0<s\leq t} \left(f(X_s) - f(X_{s-}) - \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(X_{s-})\Delta X_s^i \right). \end{aligned}$$

The rather ugly notation $(M^c)^i$ refers to the ith component of the continuous local martingale M^c and $\Delta X_s = X_s - X_{s-}$.

A rough heuristic explanation for this generalized form of Itô and an aid to memory of the formula is obtained by noting that the first three terms on the right hand side are identical to those in the multidimensional form of Itô for continuous processes.

Also terms 1,2 and 4 are exactly what you would have expected by writing down an analogue of Itô for discontinuous processes of finite variation. In other words if the process X has a jump at time s, the value of $f(X)$ changes by $f(X_s) - f(X_{s-})$. The jump in X_s will give rise to a discontinuity in

$$\sum_{i=1}^{n} \int_0^t \frac{\partial f}{\partial x_i}(X_{s-})\,\mathrm{d}X_s$$

with the jump having size

$$\sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(X_{s-})(X_s - X_{s-})$$

To get the correct jump of $f(X_s) - f(X_{s-})$ then the corrective term is

$$\sum_{0<s\leq t} \left(f(X_s) - f(X_{s-}) - \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(X_{s-})\Delta X_s^i \right).$$

21.2. RCLL processes revisited

We shall be working extensively in the space of right continuous left limits processes (CADLAG) and we shall need a metric on the function space. First let us consider the space D_T of CADLAG functions from $[0, T]$ to $\mathbb{R}$. The usual approach of using the sup norm on $[0, T]$ isn't going to be very useful. Consider two functions defined on $[0, 10]$ by

$$f(x) = \begin{cases} 0 & \text{for } x < 1, \\ 1 & \text{for } x \geq 1, \end{cases}$$

and

$$g(x) = \begin{cases} 0 & \text{for } x < 1 + \epsilon, \\ 1 & \text{for } x \geq 1 + \epsilon. \end{cases}$$

For arbitrary small ϵ these functions are always one unit apart in the sup metric. We attempt to surmount this problem by considering the class Λ_T of increasing maps from $[0,T]$ to $[0,T]$. Now define a metric

$$\mathrm{d}_T(f,g) := \inf\left\{\epsilon : \sup_{0\le s\le T} |s-\lambda(s)| \le \epsilon,\ \sup_{0\le s\le T} |f(s)-g(\lambda(s))| \le \epsilon \text{ for some } \lambda \in \Lambda_T\right\}.$$

This metric would appear to nicely have solved the problem described above. Unfortunately it is not complete! This is a major problem because Prohorov's theorem (among others) requires a complete separable metric space! However the problem can be fixed via a simple trick; for a time transform $\lambda \in \Lambda_T$ define a norm

$$|\lambda| := \sup_{0\le s\le T}\left|\log\left\{\frac{\lambda(t)-\lambda(s)}{t-s}\right\}\right|.$$

Thus we replace the condition $|\lambda(s)-s| \le \epsilon$ by $|\lambda| \le \epsilon$ and we obtain a new metric on D_T which this time is complete and separable;

$$d'_T(f,g) := \inf\left\{\epsilon : |\lambda| \le \epsilon,\ \sup_{0\le s\le T} |f(s)-g(\lambda(s))| \le \epsilon,\ \text{ for some } \lambda \in \Lambda_T\right\}.$$

This is defined on the space of CADLAG functions on the compact time interval $[0,T]$ and can be exteneded via another standard trick to D_∞,

$$d'(f,g) := \int_0^\infty e^{-t} \min\left[1, d'_t(f,g)\right] \mathrm{d}t.$$

22. References

Bensoussan, A. (1982). *Stochastic control of partially observable systems.* CUP.

Bollobas, B. (1990). *Linear Analysis.* Cambridge.

Dellacherie, C. and Meyer, P.-A. (1975). *Probabilités et Potentiel A.* Hermann, Paris.

Dellacherie, C. and Meyer, P.-A. (1980). *Probabilités et Potentiel B.* Hermann, Paris.

Durrett, R. (1996). *Probability: Theory and Examples.* Duxbury.

Ethier, S. and Kurtz, T. (1986). *Markov Processes – Characterization and Convergence.* Wiley.

Karatzas, I. and Shreve, S. E. (1987). *Brownian Motion and Stochastic Calculus.* Springer.

Mandelbaum, A., Massey, W. A., and Reiman, M. I. (1998). Strong approximations for markovian service networks. *Queueing Systems – Theory and Applications*, 30:149–201.

Musiela, M. and Rutkowski, M. (2005). *Martingale Methods in Financial Modelling.* Springer, 2nd edition.

Pardoux, E. (1989). *Filtrage non linéaire et équations aux dérivées partielles stochastiques associées*, pages 67–163. Number 1464 in Lecture notes in Mathematics. Springer Verlag.

Protter, P. (1990). *Stochastic Integration and Differential Equations.* Springer.

Rogers, L. C. G. and Williams, D. (1994). *Diffusions, Markov Processes and Martingales: Volume One: Foundations.* Wiley.

Rogers, L. C. G. and Williams, D. (2000). *Diffusions, Markov Processes and Martingales: Volume Two: Itô Calculus.* Cambridge University Press, 2nd edition.