Stochastic Calculus

Michael R. Tehranchi

Contents

CHAPTER 1

A possible motivation: diffusions

Why study stochastic calculus? In this chapter we discuss one possible motivation.

1. Markov chains

Let $(X_n)_{n\geq 0}$ be a (time-homogeneous) Markov chain on a finite state space S. As you know, Markov chains arise naturally in the context of a variety of model of physics, biology, economics, etc. In order to do any calculation with the chain, all you need to know are two basic ingredients: the initial distribution of the chain $\lambda = (\lambda_j)_{j\in S}$ defined by

$$\lambda_i = \mathbb{P}(X_0 = i) \text{ for all } i \in S$$

and the one-step transition probabilities $P = (p_{ij})_{i,j\in S}$ defined by

$$p_{ij} = \mathbb{P}(X_{n+1} = j | X_n = i) \text{ for all } i, j \in S,\ n \geq 0.$$

We now ask an admittedly pure mathematical question: given the ingredients, can we build the corresponding Markov chain? That is, given a λ such that

$$\lambda_i \geq 0 \text{ for all } i \in S, \text{ and } \sum_{i\in S} \lambda_i = 1$$

and P such that

$$\lambda_{ij} \geq 0 \text{ for all } i, j \in S, \text{ and } \sum_{j\in S} p_{ij} = 1$$

does there exist a Markov chain with initial distribution λ and transition matrix P? More precisely, does there exist a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a collection of measurable maps $X_n : \Omega \to S$ such that

$$\mathbb{P}(X_0 \in i) = \lambda_i \text{ for all } i \in S$$

and satisfying the Markov property

$$\mathbb{P}(X_{n+1} = i_{n+1} | X_n = i_n, \ldots, X_0 = i_0) = p_{i_n,i_{n+1}} \text{ for all } i_0, \ldots, i_{n+1} \in S \text{ and } n \geq 0?$$

The answer is yes, of course.

Here is one construction, involving a random dynamical system. For each i, let $I_{ij} \subseteq [0,1]$ be an interval of length p_{ij} such that the intervals $(I_{ij})_{j\in S}$ are disjoint.[1] Define a function $G : S \times [0,1] \to S$ by

$$G(i, u) = j \text{ when } u \in I_{ij}.$$

[1] For instance, if S is identified with $\{1, 2, 3, \ldots\}$ then let

$$I_{ij} = \left[\sum_{k=1}^{j-1} p_{ik}, \sum_{k=1}^{j} q_{ik}\right)$$

Now let X_0 be a random taking values in S with law λ, and let $U_1, U_2, \dots$ be independent random variables uniformly distributed on the interval $[0,1]$ and independent of X_0. Finally define $(X_n)_{n\geq 0}$ recursively by

$$X_{n+1} = G(X_n, U_{n+1}) \text{ for } n \geq 0.$$

Then it is easy to check that X is a Markov chain with the correct transition probabilities.

The lesson is that we can construct a Markov chain X given the basic ingredients λ and P. What can be said about the continuous time, continuous space case?

2. Continuous-time Markov processes

In this section, we consider a continuous-time Markov process $(X_t)_{t\geq 0}$ on the state space $S = \mathbb{R}$. What are the relevant ingredients now? Obviously, we need an initial distribution λ, a probability measure on $\mathbb{R}$ such that

$$\lambda(A) = \mathbb{P}(X_0 \in A) \text{ for all Borel } A \subseteq \mathbb{R}.$$

In this course, we are interested in *continuous* Markov processes (also known as *diffusions*), so we expect that over a short time horizon, the process has not moved too far. Heuristically we would expect that when $\Delta t > 0$ is very small, we have

$$\text{‘ } \mathbb{E}(X_{t+\Delta t} | X_t = x) \approx x + b(x)\Delta t \text{ ’}$$
$$\text{‘ } \mathrm{Var}(X_{t+\Delta t} | X_t = x) \approx \sigma(x)^2 \Delta t \text{ ’}$$

for some functions b and σ.

Now given the function b and σ can we actually build a Markov process X whose infinitesimal increments have the right mean and variance? One way to answer this question is to study stochastic differential equations.

3. Stochastic differential equations

Inspired by our construction of a Markov chain via a discrete-time random dynamical system, we try to build the diffusion X introduced in the last section via the continuous time analogue. Consider the stochastic differential equation

$$\dot{X} = b(X) + \sigma(X) N$$

where the dot denotes differentiation with respect to the time variable t and where N is some sort of noise process, filling the role of the sequence $(U_n)_{n\geq 1}$ of i.i.d. uniform random variables appearing the Markov chain construction. In particular, we will assume that $(N_t)_{t\geq 0}$ is a Gaussian white noise, but roughly speaking the important property is that N_s and N_t are independent for $s \neq t$.

3.1. Case: $\sigma = 0$**.** In this case there is no noise, so we are actually studying an ordinary differential equation

$$\dot{X} = b(X) \tag{*}$$

where the dot denotes differentiation with respect to the time variable t. When studying this system, we are lead to a variety of questions: For what types of b does the ordinary differential equation (*) have a solution? For what types of b does the ODE have a unique solution? Here is a sample answer:

THEOREM. *If b is Lipschitz, i.e. there is a constant $C > 0$ such that*

$$\|b(x) - b(y)\| \leq C\|x - y\| \text{ for all } x, y$$

then there exists a unique solution to equation (*).

3.2. The general case. Actually the above theorem holds in the more general SDE

$$(**) \qquad \dot{X} = b(X) + \sigma(X)N$$

This generalisation is due to Itô:

THEOREM. *If b and σ are Lipschitz, then the SDE (**) has a unique solution.*

Now notice that in the Markov chains example, we had a *causality principle* in the sense that the solution $(X_n)_{n\geq 0}$ of the random dynamical system

$$X_{n+1} = G(X_n, U_{n+1})$$

can be written $X_n = F_n(X_0, U_1, \ldots, U_n)$ for a deterministic function $F_n : S \times [0,1]^n \to S$. In particular, given the initial position X_0 and the driving noise $U_1, \ldots, U_n$ it is possible to reconstruct X_n. Does the same principle hold in the continuous case?

Mind-bendingly, the answer is sometimes no! And this is not because of some technicality about what is meant by measurability: we will see an explicit example of a continuous time Markov process X driven by a white noise process N in the sense that it satisfies the SDE (**), yet somehow the X is *not* completely determined by X_0 and $(N_s)_{0\leq s\leq t}$.

However, it turns out things are nice in the Lipschitz case:

THEOREM. *If b and σ are Lipschitz, then the unique solution of the SDE (**) is a measurable function of X_0 and $(N_s)_{0\leq s\leq t}$.*

4. Markov calculations

The Markov property can be exploited to do calculations. For instance, let $(X_n)_{n\geq 0}$ be a Markov chain, and define

$$u(n, i) = \mathbb{E}(f(X_n)|X_0 = i)$$

for some fixed function $f : S \to \mathbb{R}$. Then clearly

$$u(0, i) = f(i) \text{ for all } i \in S$$

and

$$\begin{aligned} u(n+1, i) &= \sum_{j\in S} \mathbb{P}(X_1 = j|X_0 = i)\mathbb{E}(f(X_{n+1})|X_1 = j) \\ &= \sum_{j\in S} p_{ij} u(n, j) \end{aligned}$$

and hence

$$u(n, i) = (P^n f)(i)$$

where P^n denotes the n-th power of the one-step transition matrix. We now explore the analogous computation in continuous time.

Let $(X_t)_{t\geq 0}$ be the diffusion with infinitesimal characteristics b and σ. Define the transition kernel $P(t, x, dy)$ by

$$P(t, x, A) = \mathbb{P}(X_t \in A|X_0 = x)$$

for any Borel set $A \subseteq \mathbb{R}$. The transition kernels satisfy the Chapman–Kolmogorov equation:

$$P(s+t,x,A) = \int P(s,x,dy)P(t,y,A).$$

Associated to the transition kernels we can associate a family $(P_t)_{t\geq 0}$ of operators on suitably integrable measurable functions defined by

$$(P_t f)(x) = \int P(t,x,dy)f(y) = \mathbb{E}[f(X_t)|X_0 = x].$$

Note that the Chapman–Kolmogorov equations then read

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

Since $P_0 = I$ is the identity operator, the family $(P_t)_{t\geq 0}$ is called the transition semigroup of the Markov process.

Suppose that there exists an operator $\mathcal{L}$ such that

$$\frac{P_t - I}{t} \to \mathcal{L}$$

as $t \downarrow 0$ in some suitable sense. Then $\mathcal{L}$ is called the infinitesimal generator of the Markov process. Notice that

$$\begin{aligned}\frac{d}{dt}P_t &= \lim_{s\downarrow 0}\frac{P_{s+t} - P_t}{s}\\ &= \lim_{s\downarrow 0}\frac{(P_s - I)P_t}{s}\\ &= \mathcal{L}P_t.\end{aligned}$$

This is called the Kolmogorov backward equation. More concretely, let

$$u(t,x) = \mathbb{E}[f(X_t)|X_0 = x]$$

for some given $f : \mathbb{R} \to \mathbb{R}$. Note that

$$\begin{aligned}u(t,x) &= \int P(t,x,dy)f(y)\\ &= (P_t f)(x)\end{aligned}$$

and so we expect u to satisfy the equation

$$\frac{\partial u}{\partial t} = \mathcal{L}u$$

with initial condition $u(0,x) = f(x)$.

The operator $\mathcal{L}$ can be calculated explicitly in the case of the solution of an SDE. Note that for small $t > 0$ we have by Taylor's theorem

$$\begin{aligned}(P_t g)(x) &= \mathbb{E}[g(X_t)|X_0 = x]\\ &\approx \mathbb{E}[g(x) + g'(x)(X_t - x) + \frac{1}{2}g''(x)(X_t - x)^2|X_0 = x]\\ &\approx g(x) + \left(b(x)g'(x) + \frac{1}{2}\sigma(x)^2 g''(x)\right)t\end{aligned}$$

and hence if there is any justice in the world, the generator is given by the differential operator

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

4.1. The case $b = 0$ and $\sigma = 1$. From the discussion above, we are lead to solve the PDE

$$u(0, x) = f(x)$$
$$\frac{\partial u}{\partial t} = \frac{1}{2}\frac{\partial^2}{\partial x^2}.$$

This is a classical equation of mathematical physics, called the heat equation, studied by Fourier among others. There is a well-known solution given by

$$u(t, x) = \int \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-y)^2}{2t}} f(y) dy.$$

On the other hand, we know that

$$u(t, x) = \mathbb{E}[f(X_t)|X_0 = x]$$

where $(X_t)_{t\geq 0}$ is a Markov process with infinitesimal drift $b = 0$ and infinitesimal variance $\sigma^2 = 1$. Comparing these to formulas, we discover that conditional on $X_0 = x$, the law of X_t is $N(x, t)$, the normal distribution with mean x and variance t. Given the central role played by the normal distribution, this should not come as a big surprise.

The goal of these lecture notes is to fill in many of the details of the above discussion. And along the way, we will learn other interesting things about Brownian motion and other continuous-time martingales! Please send all comments and corrections (including small typos and major blunders) to me at `m.tehranchi@statslab.cam.ac.uk`.

CHAPTER 2

Brownian motion

1. Motivation

Recall that we are interested in making sense of stochastic differential equations of the form

$$\dot{X} = b(X) + \sigma(X)N$$

where N is some suitable noise process. We now formally manipulate this equation to see what properties we would want from this process N. The next two sections are devoted to proving that such a process exists.

Note that we can integrate the SDE to get

$$\begin{aligned} X_{t+\Delta t} &= X_t + \int_t^{t+\Delta t} b(X_s) ds + \int_t^{t+\Delta t} \sigma(X_s) N_s \, ds \\ &\approx X_t + b(X_t)\Delta t + \sigma(X_t) W(t, t+\Delta t] \end{aligned}$$

where we think of N is as the density of a 'random (signed) measure' W so that formally

$$W(A) = \int_A N_s \, ds$$

Since we want

$$\text{` } \mathbb{E}(X_{t+\Delta t} | X_t = x) \approx x + b(x)\Delta t \text{ '}$$
$$\text{` } \mathrm{Var}(X_{t+\Delta t} | X_t = x) \approx \sigma(x)^2 \Delta t \text{ '}$$

we are looking to construct a 'white noise' W that has these properties:

- $W(A \cup B) = W(A) + W(B)$ when A and B are disjoint,
- $\mathbb{E}[W(A)] = 0$, and
- $\mathbb{E}[W(A)^2] = \mathrm{Leb}(A)$ for suitable sets A.

2. Isonormal process and white noise

In this section we introduce the isonormal processes in more generality than we need to build the white noise to drive our SDEs. It turns out that the generality comes at no real cost, and actually simplifies the arguments.

Let H be a real, separable Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$.

Definition. An isonormal process on H is a process $\{W(h), h \in H\}$ such that

- (linearity) $W(ag + bh) = aW(g) + bW(h)$ a.s. for all $a, b \in \mathbb{R}$ and $g, h \in H$,
- (normality) $W(h) \sim N(0, \|h\|^2)$ for all $h \in H$.

Assuming for the moment that we can construct an isonormal process, we now record a useful fact for future reference.

PROPOSITION. *Suppose that* $\{W(h) : h \in H\}$ *is an isonormal process.*

- *For all* $h_1, \ldots, h_n \in H$ *the random variables* $W(h_1), \ldots, W(h_n)$ *are jointly normal.*
- $\mathrm{Cov}(W(g), W(h)) = \langle g, h \rangle$.

In particular, if $\langle g, h \rangle = 0$ *then* $W(g)$ *and* $W(h)$ *are independent.*

PROOF. Fix $\theta_1, \ldots, \theta_n \in \mathbb{R}$. Using linearity, the joint characteristic function is

$$\mathbb{E}[e^{\mathrm{i}(\theta_1 W(h_1) + \ldots + \theta_n W(h_n))}] = \mathbb{E}[e^{\mathrm{i}W(\theta_1 h_1 + \ldots + \theta_n h_n)}]$$
$$\mathbb{E}[e^{-\frac{1}{2}\|\theta_1 h_1 + \ldots + \theta_n h_n\|^2}]$$

which the the joint characteristic function of mean-zero jointly normal random variables with $\mathrm{Cov}(W(h_i), W(h_j)) = \langle h_i, h_j \rangle$. □

Now we show that we construct an isonormal process.

THEOREM. *For any real separable Hilbert space, there exists a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *and a collection of measurable maps* $W(h) : \Omega \to \mathbb{R}$ *for each* $h \in H$ *such that* $\{W(h), h \in H\}$ *is an isonormal process.*

PROOF. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on which an i.i.d. sequence $(\xi_k)_k$ of $N(0,1)$ random variables is defined. Let $(g_k)_k$ be an orthonormal basis of H, and for any $h \in H$ let

$$W_n(h) = \sum_{k=1}^{n} \xi_k \langle h, g_k \rangle.$$

As the sum of independent normals, we have $W_n(h) \sim N\big(0, \sum_{k=1}^{n} \langle h, g_k \rangle^2\big)$.

With respect to the filtration $\sigma(\xi_1, \ldots, \xi_n)$, the sequence $W_n(h)$ defines a martingale. Since by Parseval's formula

$$\sup_n \mathbb{E}[W_n(h)^2] = \sum_{k=1}^{\infty} \langle h, g_k \rangle^2 = \|h\|^2$$

the martingale is bounded in $L^2(\mathbb{P})$ and hence converges to a random variable $W(h)$ on an almost sure set Ω_h. Therefore,

$$W(ag + bh) = \lim_n W_n(ag + bh) = \lim_n aW_n(g) + bW_n(h) = aW(g) + bW(h)$$

on the almost sure set $\Omega_{ag+bh} \cap \Omega_g \cap \Omega_h$. Finally, for any $\theta \in \mathbb{R}$, we have by the dominated convergence theorem

$$\mathbb{E}[e^{\mathrm{i}\theta W(h)}] = \lim_n \mathbb{E}[e^{\mathrm{i}\theta W_n(h)}]$$
$$= \lim_n e^{-\frac{1}{2}\theta^2 \sum_{k=1}^{n} \langle h, g_k \rangle^2} = e^{-\frac{1}{2}\theta^2 \|h\|^2}$$

so by the uniqueness of the characteristic functions $W(h) \sim N(0, \|h\|^2)$. □

Now we can introduce the process was motivated in the previous section.

DEFINITION. Let $(E, \mathcal{E}, \mu)$ be a measure space. A Gaussian white noise is a process $\{W(A) : A \in \mathcal{E}, \mu(A) < \infty\}$ satisfying

- (finite additivity) $W(A \cup B) = W(A) + W(B)$ a.s. for disjoint measurable sets A and B,

- (normality) $W(A) \sim N(0, \mu(A))$ for all Borel A.

THEOREM. *Suppose $(E, \mathcal{E}, \mu)$ is a measure space such that $L^2(E, \mathcal{E}, \mu)$ is separable.*[1] *Then there exists a Gaussian white noise on $\{W(A) : A \in \mathcal{E}, \mu(A) < \infty\}$.*

PROOF. Let $\{W(h) : h \in L^2(\mu)\}$ be an isonormal process. Without too much confusion, we use the same notation (a standard practice in measure theory) to define

$$W(A) = W(\mathbb{1}_A).$$

Note that if A and B are disjoint we have

$$\begin{aligned} W(A \cup B) &= W(\mathbb{1}_{A \cup B}) \\ &= W(\mathbb{1}_A + \mathbb{1}_B) \\ &= W(\mathbb{1}_A) + W(\mathbb{1}_B) \\ &= W(A) + W(B). \end{aligned}$$

Also, the first two moments can be calculated as

$$\mathbb{E}[W(A)] = \mathbb{E}[W(\mathbb{1}_A)] = 0$$

and

$$\begin{aligned} \mathbb{E}[W(A)^2] &= \mathbb{E}[W(\mathbb{1}_A)^2] \\ &= \|\mathbb{1}_A\|_{L^2}^2 \\ &= \int \mathbb{1}_A d\mu \\ &= \mu(A). \end{aligned}$$

$\square$

As a concluding remark, note that if A and B are disjoint then

$$\langle \mathbb{1}_A, \mathbb{1}_B \rangle_{L^2} = \int \mathbb{1}_{A \cap B} d\mu = 0$$

and hence $W(A)$ and $W(B)$ are independent.

3. Wiener's theorem

The previous section shows that there is some hope in interpreting the SDE

$$\dot{X} = b(X) + \sigma(X) N$$

as the integral eqution

$$X_t = X_0 + \int_0^t b(X_s) ds + \int_0^t \sigma(X_s) W(ds)$$

[1] Let $\mathcal{E}_f = \{A \in \mathcal{E} : \mu(A) < \infty\}$ be the measurable sets of finite measure. Define a pseudo-metric d on $\mathcal{E}_f$ by $d(A, B) = \mu(A \Delta B)$ where the symmetric difference is defined by $A \Delta B = (A \cap B^c) \cup (A^c \cap B)$. Say A is equivalent to A' iff $d(A, A') = 0$, and let $\tilde{\mathcal{E}}_f$ be collection of equivalence classes of $\mathcal{E}_f$. Then the Hilbert space $L^2(E, \mathcal{E}, \mu)$ is separable if and only if the metric space $(\tilde{\mathcal{E}}_f, d)$ is separable.

where W is the white noise on $([0,\infty),\mathcal{B},\mathrm{Leb})$. There still remains a problem: we have only shown that $W(\cdot,\omega)$ is *finitely* additive for almost all $\omega\in\Omega$. It turns out that $W(\cdot,\omega)$ is *not* countably additive, so the usual Lebesgue integration theory will not help us!

So, let's lower our ambitions and consider the case where $\sigma(x)=\sigma_0$ is constant. In this case we should have the integral equation

$$X_t = X_0 + \int_0^t b(X_s)ds + \sigma_0 W_t$$

where $W_t = W(0,t]$ is the 'distribution function' of white noise set function W. This seems easier than the general case, except that a nagging technicality remains: how do we make sense of the remaining integral? Why should this be a problem? Recall that for each set A, the random variable $W(A)=W(\mathbb{1}_A)$ was constructed via an infinite series that converged on an almost sure event Ω_A. Unfortunately, Ω_A depends on A in a complicated way, so it is not at all clear that we can define $W_t = W(0,t]$ simultaneously for all *uncountable* $t\geq 0$ since it might be the case that $\cap_{t\geq 0}\Omega_{(0,t]}$ is not measurable or even empty. This is a problem, since we need to check that $t\mapsto b(X_t(\omega))$ is measurable for almost all $\omega\in\Omega$ to use the Lebesgue integration theory.... Fortunately, its possible to do the construction is such a way that $t\mapsto W_t(\omega)$ is continuous for almost all ω. This we call a Brownian motion.

DEFINITION. A (standard) Brownian motion is a stochastic process $W=(W_t)_{t\geq 0}$ such that

- $W_0 = 0$ a.s.
- (stationary increments) $W_t - W_s \sim N(0,t-s)$ for all $0\leq s<t$,
- (independent increments) For any $0\leq t_0<t_1<\cdots<t_n$, the random variables $W_{t_1}-W_{t_0},\ldots,W_{t_n}-W_{t_{n-1}}$ are independent.
- (continuity) W is continuous, meaning $\mathbb{P}(\omega\in\Omega : t\mapsto W_t(\omega)\text{ is continuous })=1$.

THEOREM (Wiener 1923). *There exists a probability space $(\Omega,\mathcal{F},\mathbb{P})$ and a collection of random variables $W_t:\Omega\to\mathbb{R}$ for each $t\geq 0$ such that the process $(W_t)_{t\geq 0}$ is a Brownian motion.*

As we seen already, we can take the Gaussian white noise $\{W(A): A\subset[0,\infty)\}$, and set $W_t = W(0,t]$. Then we will automatically have $W_0=0$, and since

$$W_t - W_s = W(s,t] \sim N(0,t-s)$$

for $0\leq s<t$, the increments are stationary. Furthermore, as discussed in the last section, the increments over disjoint intervals are uncorrelated and hence independent. Therefore, we need only show continuity. The main idea is to revisit the construction of the isonormal process, and now make a clever choice of orthonormal basis of L^2.

First we need a lemma that says we need only build a Brownian motion on the interval $[0,1]$ because we can concatenate independent copies to build a Brownian motion on the interval $[0,\infty)$. The proof is an easy exercise.

LEMMA. *Let $(W_t^{(n)})_{0\leq t\leq 1}$, $n=1,2,\ldots$ be independent Brownian motions and let*

$$W_t = W_1^{(1)} + \cdots + W_1^{(n)} + W_{t-n}^{(n+1)} \text{ for } n\leq t<n+1$$

Then $(W_t)_{t\geq 0}$ is a Brownian motion.

Now the clever choice of basis is this:

Definition. The set of functions

$$\{1; \quad h_0^0; \quad h_1^0, h_1^1; \quad \ldots, ; h_n^0, \ldots, h_n^{2^n-1}; \ldots\}$$

where

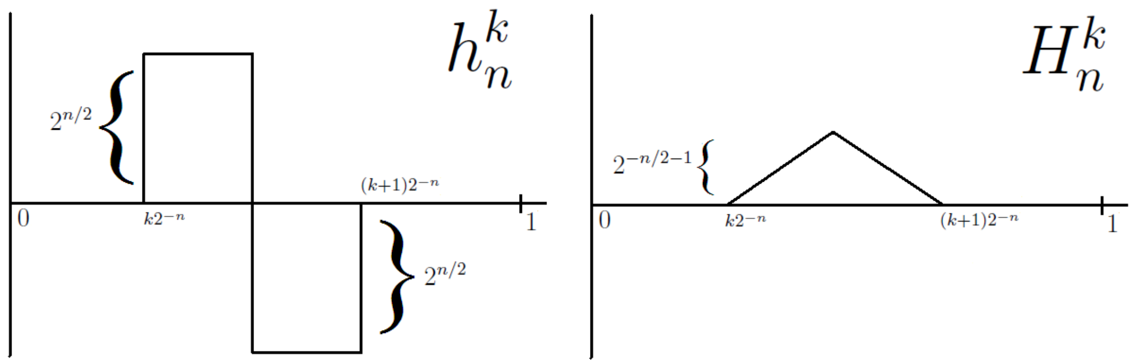
$$h_n^k = 2^{n/2}\left(\mathbb{1}_{[k2^{-n},(k+1/2)2^{-n})} - \mathbb{1}_{[(k+1/2)2^{-n},(k+1)2^{-n})}\right)$$

is called the *Haar* functions.

Lemma. *The collection of Haar functions is an orthonormal basis of $L^2[0,1]$.*

We will come back to prove this lemma at the of the section. Finally, we introduce some useful notation:

$$\begin{aligned} H_n^k(t) &= \langle h_n^k, \mathbb{1}_{(0,t]}\rangle_{L^2} \\ &= \int_0^t h_n^k(s)ds. \end{aligned}$$



Proof of Wiener's theorem. Let ξ^* and $(\xi_n^k)_{n,k}$ be a collection of independent $N(0,1)$ random variables, and consider the event

$$\Omega_0 = \left\{\omega : \sum_{n=0}^{\infty} \sup_{0\le t\le 1}\left|\sum_{k=0}^{2^n-1} \xi_n^k(\omega) H_n^k(t)\right| < \infty\right\}.$$

Note that for every $\omega \in \Omega_0$, the partial sums $W_t^{(N)}(\omega)$ converge uniformly in t where

$$W_t^{(N)}(\omega) = t\xi^*(\omega) + \sum_{n=0}^{N}\sum_{k=0}^{2^n-1} \xi_n^k(\omega)H_n^k(t).$$

Since $t \mapsto W_t^{(N)}(\omega)$ is continuous, and the uniform limit of continuous functions is continuous, we need only show that Ω_0 is almost sure.

Now note that, for fixed n, the supports of the functions $H_n^0, \ldots, H_n^{2^n-1}$ are disjoint. Hence

$$\sup_{0\le t\le 1}\left|\sum_{k=0}^{2^n-1} \xi_n^k H_n^k(t)\right| = 2^{-n/2-1} \max_{0\le k\le 2^n-1} |\xi_n^k|.$$

Now, for any $p > 1$ we have

$$\begin{aligned}
\mathbb{E} \max_{0 \le k \le 2^n - 1} |\xi_n^k| &\le \left(\mathbb{E} \max_{0 \le k \le 2^n-1} |\xi_n^k|^p \right)^{1/p} \quad \text{by Jensen's inequality} \\
&\le \left(\mathbb{E} \sum_{k=0}^{2^n-1} |\xi_n^k|^p \right)^{1/p} \\
&= (2^n c_p)^{1/p}
\end{aligned}$$

where

$$\begin{aligned}
c_p &= \mathbb{E}(|\xi|^p) \\
&= \pi^{-1/2} 2^{p/2} \Gamma(\frac{p+1}{2}).
\end{aligned}$$

Now, choosing any $p > 2$ yields

$$\begin{aligned}
\mathbb{E} \sum_{n=0}^{\infty} \sup_{0 \le t \le 1} \left| \sum_{k=0}^{2^n-1} \xi_n^k H_n^k(t) \right| &= \sum_{n=0}^{\infty} 2^{-n/2-1} \mathbb{E}(\max_{0 \le k \le 2^n-1} |\xi_n^k|) \\
&\le c_p^{1/p} 2^{-1} \sum_{n=0}^{\infty} 2^{-n(1/2-1/p)} < \infty.
\end{aligned}$$

This shows that $\mathbb{P}(\Omega_0) = 1$. □

And to fill in a missing detail:

Proof that the Haar functions are a basis. The orthogonality and normalisation of the Haar functions is easy to establish.

First we will show that the set of functions which are constant over intervals of the form $[k2^{-n}, (k+1)2^{-n})$ are dense in $L^2[0,1]$. There are many ways of seeing this. Here is a probabilistic argument: for any $f \in L^2[0,1]$, let

$$f_n = \sum_{k=0}^{2^n-1} f_n^k I_n^k$$

where

$$I_n^k = \mathbb{1}_{[k2^{-n},(k+1)2^{-n})}$$

and

$$f_n^k = 2^n \int_{k2^{-n}}^{(k+1)2^{-n}} f(x) dx.$$

Then $f_n \to f$ in L^2 (and almost everywhere). To see this, consider the family of sigma-fields

$$\mathcal{F}_n = \sigma([k2^{-n}, (k+1)2^{-n}) : 0 \le k \le 2^n - 1)$$

on $[0,1]$. For any $f \in L^2[0,1]$ let

$$f_n = \mathbb{E}(f | \mathcal{F}_n)$$

where the expectation is with respect to the probability measure $\mathbb{P} = \text{Leb}$. Since $\sigma(\cup_{n \ge 0} \mathcal{F}_n)$ is the Borel sigma-field on $[0,1]$, the result follows from the martingale convergence theorem.

Now we will show that every indicator of a dyadic interval is a linear combination of Haar functions. This is done inductively on the level n. Indeed for $k = 2j$ we have

$$I_{n+1}^{2j} = \frac{1}{2}(I_n^j + 2^{-n/2} h_n^j)$$

and for $k = 2j + 1$ we have

$$I_{n+1}^{2j+1} = \frac{1}{2}(I_n^j - 2^{-n/2} h_n^j).$$

□

4. Some sample path properties

Recall from our motivating discussion that would like to think of the Brownian motion as the integral

$$W_t = \int_0^t N_s ds$$

for some noise process N. Is there any chance that this notation can be anything but formal? No:

THEOREM (Paley, Wiener and Zygmund 1933). *Let W be a scalar Brownian motion defined on a complete probability space. Then*

$$\mathbb{P}\{\omega : t \mapsto W_t(\omega) \text{ is differentiable somewhere } \} = 0.$$

PROOF. First, we will find a necessary condition that a function $f : [0, \infty) \to \mathbb{R}$ is differentiable at a point. The idea is to express things in such a way that computations of probabilities are possible when applied to the Brownian sample path.

Recall that to say a function f is differentiable at a point $s \geq 0$ means that for any $\delta > 0$ there exists an $\varepsilon > 0$ such that for all $t \geq 0$ in the interval $(s - \varepsilon, s + \varepsilon)$ the inequality

$$|f(t) - f(s) - (t - s)f'(s)| \leq \delta |t - s|$$

holds.

So suppose f is differentiable at s. By setting $\delta = 1$, we know there exists an $\varepsilon_0 > 0$ such that for all $0 \leq \varepsilon < \varepsilon_0$ and $t_1, t_2 \in [s, s + \varepsilon]$ we have

$$\begin{aligned}
|f(t_1) - f(t_2)| &\leq |f(t_1) - f(s) - (t_1 - s)f'(s)| + |f(t_2) - f(s) - (t_2 - s)f'(s)| + |(t_2 - t_1)f'(s)| \\
&\leq |t_1 - s| + |t_2 - s| + |t_2 - t_1||f'(s)| \\
&\leq (2 + |f'(s)|)\varepsilon
\end{aligned}$$

by the triangle inequality. In particular, there exists an integer $M \geq 4(2 + |f'(s)|)$ and integer $N \geq 4/\varepsilon_0$ such that

$$|f(t_1) - f(t_2)| \leq M/n$$

whenever $n > N$ and $t_1, t_2 \in [s, s + 4/n]$. (The reason for the mysterious factor of 4 will become apparent in a moment.)

Note that any $n > N$, there exists an integer i such that $s \leq i/n < s + 1/n$. For this integer i, the four points $i/n, (i + 1)/n, (i + 2)/n, (i + 3)/n$ are all contained in the interval $[s, s + 4/n]$. And in particular,

$$|f(\tfrac{j+1}{n}) - f(\tfrac{j}{n})| \leq M/n$$

for $j = i, i+1, i+2$. Finally, there exists an integer $K \geq s$, and therefore $i < nK + 1$.

Therefore, we have shown that

$$\{\omega : t \mapsto W_t(\omega) \text{ is differentiable somewhere } \} \subseteq \mathcal{N}$$

where

$$\mathcal{N} = \bigcup_M \bigcup_K \bigcup_N \bigcap_{n>N} \bigcup_{i \leq nK} \bigcap_{j=i}^{i+2} A_{j,n,M}$$

and

$$A_{j,n,M} = \{|W_{(j+1)/n} - W_{j/n}| \leq M/n\}.$$

Note that the sets $A_{j,n,M}$ are increasing in M, in the sense that $A_{j,n,M} \subseteq A_{j,n,M+1}$. Similarly, the sets $\bigcup_K \bigcup_{i \leq nK} A_{j,n,M}$ are increasing in K. Hence from basic measure theory and the definition of Brownian motion we have the following estimates:

$$\begin{aligned}
\mathbb{P}(\mathcal{N}) &= \sup_{M,K} \mathbb{P}\left(\bigcup_N \bigcap_{n>N} \bigcup_{i \leq nK} \bigcap_{j=i}^{i+2} A_{j,n,M} \} \right) \text{ continuity of } \mathbb{P} \\
&\leq \sup_{M,K} \liminf_{n\to\infty} \mathbb{P}\left(\bigcup_{i \leq nK} \bigcap_{j=i}^{i+2} A_{j,n,M} \right) \text{ Fatou's lemma} \\
&\leq \sup_{M,K} \liminf_{n\to\infty} \sum_{i=0}^{nK} \mathbb{P}\left(\bigcap_{j=i}^{i+2} A_{j,n,M} \right) \text{ subadditivity of } \mathbb{P} \\
&= \sup_{M,K} \liminf_{n\to\infty} \left[(nK+1) \mathbb{P}\left(A_{0,n,M}\right)^3 \right]
\end{aligned}$$

The last step follows from the fact that the increments $W_{(j+1)/n} - W_{j/n}$ are independent and each have the $N(0, 1/n)$ distribution.

Note that

$$\begin{aligned}
\mathbb{P}(|W_t| \leq r) &= \int_{-r/\sqrt{t}}^{r/\sqrt{t}} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \\
&\leq C \frac{r}{\sqrt{t}}
\end{aligned}$$

where $C = \sqrt{2/\pi}$ and hence

$$\mathbb{P}(A_{0,n,M}) = \mathbb{P}(|W_{1/n}| \leq M/n) \leq C \frac{M}{\sqrt{n}}.$$

Putting this together yields

$$\mathbb{P}(\mathcal{N}) \leq \sup_{M,K} \liminf_{n\to\infty} \left[(nK+1) \left(C \frac{M}{\sqrt{n}} \right)^3 \right] = 0.$$

$\square$

REMARK. It should be clear from the argument that differentiability of a function at a point means that we can control the size of increments of the function over p disjoints intervals, where $p \geq 1$ is any integer, and so

$$\{\omega : t \mapsto W_t(\omega) \text{ is differentiable somewhere }\} \subseteq \mathcal{N}_p$$

where

$$\mathcal{N}_p = \bigcup_M \bigcup_K \bigcup_N \bigcap_{n>N} \bigcup_{i \leq nK} \bigcap_{j=i}^{i+p-1} \{|W_{(j+1)/n} - W_{j/n}| \leq M/n\}$$

The somewhat arbitrary choice of $p = 3$ in the proof above is to ensure that $n^{1-p/2} \to 0$.

So, sample paths of Brownian motion are continuous but nowhere differentiable. It is possible to peer even deeper into the fine structure of Brownian sample paths. The following results are stated without proof since we will not use them in future lectures.

THEOREM. *Let W be a scalar Brownian motion.*

- *The law of iterated logarithms. (Khinchin 1933) For fixed $t \geq 0$,*

$$\limsup_{\delta \downarrow 0} \frac{W_{t+\delta} - W_t}{\sqrt{2\delta \log \log 1/\delta}} = 1 \ a.s.$$

- *The Brownian modulus of continuity. (Lévy 1937)*

$$\limsup_{\delta \downarrow 0} \frac{\max_{s,t \in [0,1], |t-s|<\delta} |W_t - W_s|}{\sqrt{2\delta \log 1/\delta}} = 1 \ a.s.$$

5. Filtrations and martingales

Recall that we are also interested in Markov processes with infinitesimal characteristics

$$\mathbb{E}[X_{t+\Delta t}|X_t = x] \approx x + b(x)\Delta t \text{ and } \mathrm{Var}(X_{t+\Delta t}|X_t = x) \approx \sigma(x)^2 \Delta t$$

with corresponding formal differential equation

$$\dot{X} = f(X) + \sigma(X)N.$$

If $\sigma(x) = \sigma_0$ and b is measurable, we can *define* the solution of the differential equation to be a continuous process $(X_t)_{t \geq 0}$ such that

$$X_t = X_0 + \int_0^t b(X_s)ds + \sigma_0 W_t$$

where W is a Brownian motion. We still haven't proven[2] that this equation has a solution, but at least we know how to interpret all of the terms.

But we are also interested in Markov processes whose increments have a state-dependent variance, where the function σ is not necessarily constant. It seems that the next step in our programme is to give meaning to the integral equation

$$X_t = X_0 + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)W(ds)$$

[2] When b is Lipschitz, one strategy is to use the Picard–Lindelöf theorem as suggested on the first example sheet. On the fourth example sheet you are asked to construct a solution of the SDE, only assuming that b is measurable and of linear growth and that $\sigma_0 \neq 0$.

where W is the Gaussian white noise on $[0,\infty)$. Unfortunately, this does not work:

DEFINITION. A signed measure μ on a measurable space $(E,\mathcal{E})$ is a set function of the form $\mu(A)=\mu_+(A)-\mu_-(A)$ for all $A\in\mathcal{E}$, where $\mu_\pm$ are measures (such that at least one of $\mu_+(A)$ or $\mu_-(A)$ is finite for any A to avoid $\infty-\infty$). For convenience, we will also insist that both measures $\mu_\pm$ are sigma-finite.

THEOREM. *Let* $\{W(A) : A\subseteq[0,\infty)\}$ *be a Gaussian white noise with* $\mathrm{Var}[W(A)] = \mathrm{Leb}(A)$ *defined on a complete probability space. Then*

$$\mathbb{P}(\omega : A\mapsto W(A,\omega) \textit{ is a signed measure }) = 0$$

The following proof depends on this result of real analysis:

THEOREM (Lebesgue). *If* $f:[0,\infty)\to\mathbb{R}$ *is monotone, then* f *is differentiable (Lebesgue)-almost everywhere.*

PROOF. Suppose that $W(\cdot,\omega)$ is a signed measure for some outcome ω. Then $t\mapsto W((0,t],\omega)=\mu_+(0,t]-\mu_-(0,t]$ is the difference of two monotone functions and hence is differentiable almost everywhere. But we have proven that the probability that this map is differentiable *some where* is zero! □

Nevertheless, all hope is not lost. The key insights that will allow us to give meaning to the integral $\int_0^t a_s dW_s$ are that

- the Brownian motion W is a martingale, and
- we need not define the integral for all processes $(a_t)_{t\geq 0}$, since it is sufficient for our application to consider only processes that do not anticipate the future in some sense.

We now discuss the martingale properties of the Brownian motion. Recall some definitions:

DEFINITION. A *filtration* $\mathbb{F}=(\mathcal{F}_t)_{t\geq0}$ is an increasing family of sigma-fields, i.e. $\mathcal{F}_s\subseteq\mathcal{F}_t$ for all $0\leq s\leq t$.

DEFINITION. A process $(X_t)_{t\geq0}$ is *adapted* to a filtration $\mathbb{F}$ iff X_t is $\mathcal{F}_t$-measurable for all $t\geq 0$.

DEFINITION. An adapted process $(X_t)_{t\geq0}$ is a *martingale* with respect to the filtration $\mathbb{F}$ iff

(1) $\mathbb{E}(|X_t|)<\infty$ for all $t\geq0$,
(2) $\mathbb{E}(X_t|\mathcal{F}_s)=X_s$ for all $0\leq s\leq t$.

Our aim is to show a Brownian motion is a martingale. But in what filtration?

DEFINITION. A Brownian motion W is a Brownian motion in a filtration $\mathbb{F}$ (or, equivalently, the filtration $\mathbb{F}$ is compatible with the Brownian motion W) iff W is adapted to $\mathbb{F}$ and the increments $(W_u-W_t)_{u\in[t,\infty)}$ are independent of $\mathcal{F}_t$ for all $t\geq0$.

EXAMPLE. The natural filtration

$$\mathcal{F}_t^W=\sigma(W_s, 0\leq s\leq t)$$

is compatible with a Brownian motion W. This a direct consequence of the definition.

The reason that we have introduced this definition is that we will soon find it useful sometimes to work with filtrations bigger than a Brownian motion's natural filtration.

THEOREM. *Let W be a Brownian motion in a filtration $\mathbb{F}$. Then the following processes are martingales:*

(1) *The Brownian motion itself W.*
(2) $W_t^2 - t$.
(3) $e^{\mathrm{i}\theta W_t + \theta^2 t/2}$ *for any complex θ, where $\mathrm{i} = \sqrt{-1}$*

PROOF. This is an easy exercise in using the independence of the Brownian increments and computing with the normal distribution. □

The above theorem has an important converse.

THEOREM. *Let X be a continuous process adapted to a filtration $\mathbb{F}$ such that $X_0 = 0$ and*

$$e^{\mathrm{i}\theta X_t + \theta^2 t/2}$$

is a martingale. Then X is a Brownian motion in $\mathbb{F}$.

PROOF. By assumption

$$\mathbb{E}(e^{\mathrm{i}\ \theta(X_t - X_s)}|\mathcal{F}_s) = e^{-\theta^2(t-s)/2}.$$

for all $\theta \in \mathbb{R}$ and $0 \le s \le t$.

So fix $0 \le s \le t_0 < t_1 < \cdots < t_n$ and $\theta_1, \ldots, \theta_n \in \mathbb{R}^d$. By iterating expectations we have

$$\begin{aligned}
\mathbb{E}(e^{\mathrm{i}\sum_{k=1}^n \theta_k (X_{t_k} - X_{t_{k-1}})}|\mathcal{F}_s) &= \mathbb{E}[\mathbb{E}(e^{\mathrm{i}\sum_{k=1}^n \theta_k (X_{t_k} - X_{t_{k-1}})}|\mathcal{F}_{t_{n-1}})|\mathcal{F}_s] \\
&= \mathbb{E}[e^{\mathrm{i}\sum_{k=1}^{n-1} \theta_k (X_{t_k} - X_{t_{k-1}})}\mathbb{E}(e^{\mathrm{i}\theta_n \cdot (X_{t_n} - X_{t_{n-1}})}|\mathcal{F}_{t_{n-1}})|\mathcal{F}_s] \\
&= \mathbb{E}[e^{\mathrm{i}\sum_{k=1}^{n-1} \theta_k (X_{t_k} - X_{t_{k-1}})}|\mathcal{F}_s]e^{-\theta_n^2(t_n - t_{n-1})/2} \\
&= \cdots \\
&= e^{-\frac{1}{2}\sum_{k=1}^n \theta_k^2 (t_k - t_{k-1})}
\end{aligned}$$

Since the conditional characteristic function of the increments $(X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}})$ given $\mathcal{F}_s$ is not random, we can conclude that $(X_u - X_s)_{u \ge s}$ is independent of $\mathcal{F}_s$. Details are in the footnote[3]. Furthermore, by inspecting this characteristic function, we can conclude

[3] Fix a set $A \in \mathcal{F}_s$. Define two measures on $\mathbb{R}^n$ by

$$\mu(B) = \mathbb{P}(\{(X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}}) \in B\} \cap A)$$

and

$$\nu(B) = \mathbb{P}((X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}}) \in B)\mathbb{P}(A).$$

Note that for any $\theta \in \mathbb{R}^n$, we have

$$\begin{aligned}
\int e^{\mathrm{i}\theta \cdot y}\mu(dy) &= \mathbb{E}(e^{\mathrm{i}\sum_{k=1}^n \theta_k (X_{t_k} - X_{t_{k-1}})}\mathbb{1}_A) \\
&= \mathbb{E}(e^{\mathrm{i}\sum_{k=1}^n \theta_k (X_{t_k} - X_{t_{k-1}})})\mathbb{P}(A) \\
&= \int e^{\mathrm{i}\theta \cdot y}\nu(dy).
\end{aligned}$$

Since characteristic functions characterise measures, we have $\mu(B) = \nu(B)$ for all measurable $B \subseteq \mathbb{R}^n$, and hence the random vector $(X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}})$ is independent of the event A. Since this independence holds for all $0 \le s \le t_0 < t_1 < \cdots < t_n$, the sigma-field $\sigma(X_u - X_s : u \ge s)$ is independent of $\mathcal{F}_s$.

that the increments are independent with distribution $X_t - X_s \sim N(0, |t-s|)$. Finally, the process is assumed to be continuous, so it must be a Brownian motion.

□

6. The usual conditions

We now discuss some details that have been glossed over up to now. For instance, we have used the assumption that our probability space is complete a few times. Recall what that means:

Definition. A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is complete iff $\mathcal{F}$ contains all the $\mathbb{P}$-null sets. That is to say, if $A \subseteq B$ and $B \in \mathcal{F}$ and $\mathbb{P}(B) = 0$, then $A \in \mathcal{F}$ and $\mathbb{P}(A) = 0$.

For instance, in our discussion of the nowhere differentiability of Brownian sample paths, we only proved that the set

$$\{\omega : t \mapsto W_t(\omega) \text{ is somewhere differentiable }\}$$

is contained in an event $\mathcal{N}$ of probability zero. But is the set above measurable? In general, it may fail to be measurable since the condition involves estimating the behaviour of the sample paths at an uncountable many points $t \geq 0$. But, if we assume completeness, this technical issue disappears.

And as we assume completeness of our probability space, we will also employ a convenient assumption on our filtrations.

Definition. A filtration $\mathbb{F}$ satisfies the *usual conditions* iff

- $\mathcal{F}_0$ contains all the $\mathbb{P}$-null sets, and
- $\mathbb{F}$ is right-continuous in that $\mathcal{F}_t = \bigcap_{\epsilon > 0} \mathcal{F}_{t+\epsilon}$ for all $t \geq 0$.

Why do we need the usual conditions? It happens to make life easier in that the usual conditions guarantee that there are plenty of stopping times...

Definition. A *stopping time* is a random variable $T : \Omega \to [0, \infty]$ such that $\{T \leq t\} \in \mathcal{F}_t$ for all $t \geq 0$.

Theorem. *Suppose the filtration $\mathbb{F}$ satisfies the usual conditions. Then a random time T is a stopping time iff $\{T < t\} \in \mathcal{F}_t$ for all $t \geq 0$.*

Proof. First suppose that $\{T < t\} \in \mathcal{F}_t$ for all $t \geq 0$. Since for any $N \geq 1$ we have

$$\{T \leq t\} = \bigcap_{n \geq N} \{T < t + 1/n\}$$

and each event $\{T < t + 1/n\}$ is in $\mathcal{F}_{t+1/n} \subseteq \mathcal{F}_{t+1/N}$ by assumption, then

$$\{T \leq t\} \in \bigcap_{N} \mathcal{F}_{t+1/N} = \mathcal{F}_t$$

by the assumption of right-continuity. Hence T is a stopping time.

Conversely, now suppose that $\{T \leq t\} \in \mathcal{F}_t$ for all $t \geq 0$. Since

$$\{T < t\} = \bigcup_{n} \{T \leq t - 1/n\}$$

and each event $\{T \le t - 1/n\}$ is in $\mathcal{F}_{t-1/n} \subseteq \mathcal{F}_t$ by the definition of stopping time, then by the definition of sigma-field the union

$$\{T < t\} \in \mathcal{F}_t$$

is measurable as claimed. $\square$

THEOREM. *Suppose X is a right-continuous n-dimensional process adapted to a filtration satisfying the usual conditions. Fix an open set $A \subseteq \mathbb{R}^n$ and let*

$$T = \inf\{t \ge 0 : X_t \in A\}.$$

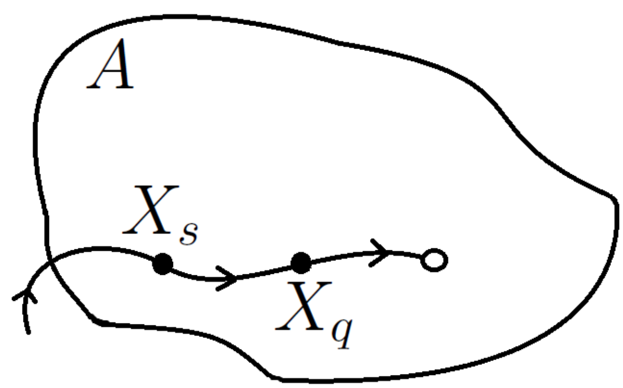
Then T is a stopping time.

PROOF. Let $f : [0, \infty) \to \mathbb{R}^n$ be right-continuous. Suppose that for some $s \ge 0$ we have $f(s) \in A$. Since A is open, there exists a $\delta > 0$ such that

$$f(s) + y \in A$$

for all $\|y\| < \delta$. Since f is right-continuous, there exists $\epsilon > 0$ such that $\|f(s) - f(u)\| < \delta$ for all $u \in [s, s + \epsilon)$. In particular, for all rational $q \in [s, s + \epsilon)$ we have

$$f(q) \in A.$$



Now, by the previous theorem is enough to check that $\{T < t\} \in \mathcal{F}_t$ for each $t \ge 0$. Note that

$$\begin{aligned}\{T < t\} &= \{X_s \in A \text{ for some } s \in [0, t)\} \\ &= \bigcup_{q \in [0,t) \cap Q} \{X_q \in A\}\end{aligned}$$

where Q is the set of rationals. Since X_q is $\mathcal{F}_q$-measurable by the assumption that X is adapted, then $\{X_q \in A\} \in \mathcal{F}_q \subseteq \mathcal{F}_t$. And since the set of rationals is countable, the union is in $\mathcal{F}_t$ also. $\square$

REMARK. Note that by the definition of infimum, the event $\{T = t\}$ involves the behaviour of X shortly after time t. The usual conditions help smooth over these arguments.

One of the main reasons why the usual conditions were recognised as a natural assumption to make of a filtration is that they appear in the following famous theorem. The proof is omitted since this is done in Advanced Probability.

Theorem (Doob's regularisation.). *If X is a martingale with respect to a filtration satisfying the usual conditions, then X has a right-continuous*[4]*, modification, i.e. a right-continuous process X^* such that*

$$\mathbb{P}(X_t = X_t^*) = 1 \textit{ for all } t \geq 0.$$

In the previous lectures, we have discussed filtrations without checking that the usual conditions are satisfied. Do we have to go back and reprove everything? Fortunately, the answer is no since we have been dealing with continuous processes:

Theorem. *Let X be a right-continuous martingale with respect to a filtration $\mathbb{F}$. Then X is also martingale for $\mathbb{F}^*$ where*

$$\mathcal{F}_t^* = \sigma\left(\bigcap_{\epsilon>0} \mathcal{F}_{t+\epsilon} \cup \{\mathbb{P} - \textit{null sets}\}\right)$$

Proof. We need to show that for any $0 \leq s \leq t$ that

$$\mathbb{E}[X_t | \mathcal{F}_s^*] = X_s$$

which means that for any $A \in \mathcal{F}_s^*$, the equality

$$\mathbb{E}[(X_t - X_s)\mathbb{1}_A] = 0$$

holds.

First, by the definition of null set, for any $A \in \mathcal{F}_s^*$ we can find a set $B \in \mathcal{F}_{s+}$ such that

$$\mathbb{1}_A = \mathbb{1}_B \text{ a.s.}$$

so it is enough to show that

$$\mathbb{E}[(X_t - X_s)\mathbb{1}_B] = 0.$$

Now, since X is an $\mathbb{F}$ martingale, we have

$$\mathbb{E}[X_t | \mathcal{F}_{s+\epsilon}] = X_{s+\epsilon}$$

for any $s + \epsilon \leq t$. Hence, since $B \in \mathcal{F}_{s+} \subseteq \mathcal{F}_{s+\epsilon}$ we have

$$\mathbb{E}[(X_t - X_{s+\epsilon})\mathbb{1}_B] = 0.$$

Now note that $X_{s+\epsilon} \to X_s$ a.s. as $\epsilon \downarrow 0$ by the assumption of right-continuity, and that the collection of random variables $(X_{s+\epsilon})_{\epsilon \in [0,t-s]}$ is uniformly integrable by the martingale property. Therefore,

$$\mathbb{E}[(X_t - X_s)\mathbb{1}_A] = \mathbb{E}[(X_{s+\epsilon} - X_s)\mathbb{1}_B] \to 0.$$

□

With these considerations in mind, we now state an assumption that will be in effect throughout the course:

Assumption. All filtrations are assumed to satisfy the usual conditions, unless otherwise explicitly stated.

We are now ready to move on to the main topic of the course.

[4]... in fact the sample paths are càdlàg, which is an abbreviation for *continue à droite, limite à gauche*

CHAPTER 3

Stochastic integration

1. Overview

The goal of this chapter is to construct the stochastic integral. Recall that from our motivating discussion, we would like to give meaning to integrals of the form

$$X_t = \int_0^t K_s dM_s.$$

where M is a Brownian motion and K is non-anticipating in some sense.

It turns out that a convenient class of integrands K are the predictable processes, and a convenient class of integrators M are the continuous local martingales. In particular, if K is compatible with M is a certain sense, then the integral X can be defined, and it is also a local martingale.

What we will see is that for each continuous local martingale M there exists a continuous non-decreasing process $\langle M\rangle$, called its quadratic variation, which somehow measures the 'speed' of M. The compatibility condition to define the integral is that

$$\int_0^t K_s^2 d\langle M\rangle_s < \infty \text{ a.s. for all } t \geq 0.$$

Note that since $\langle M\rangle$ is non-decreasing, we can interpret the above integral as a Lebesgue–Stieltjes integral for each outcome $\omega \in \Omega$, i.e. we don't need a new integration theory to understand it. Finally, to start to appreciate where this compatibility condition comes from, it turns out that the quadratic variation of the local martingale X is given by

$$\langle X\rangle_t = \int_0^t K_s^2 d\langle M\rangle_s.$$

That is the outline of the programme. In particular, we need to define some new vocabulary, including what is meant by a local martingale. This is the topic of the next section.

2. Local martingales and uniform integrability

As indicated in the last section, a starring role in this story is played by local martingales:

DEFINITION. A *local martingale* is a right-continuous adapted process X such that there exists a non-decreasing sequence of stopping times $T_n \uparrow \infty$ such that the stopped process

$$X^{T_n} - X_0 = (X_{t\wedge T_n} - X_0)_{t\geq 0}$$

is a martingale for each n. The localising sequence $(T_n)_{n\geq 1}$ is said to reduce X to martingale.

Remark. Note that the definition can be simplified under the additional assumption that X_0 is integrable: a local martingale is an adapted process X such that there exists stopping times $T_n \uparrow \infty$ for which the stopped process

$$X^{T_n} = (X_{t\wedge T_n})_{t\geq 0}$$

is a martingale for each n.

This simplifying assumption holds in several common contexts. For instance, it is often the case that $\mathcal{F}_0$ is trivial, in which case X_0 is constant and hence integrable. Also, the local martingales defined by stochastic integration are such that $X_0 = 0$, so the integrability condition holds automatically.

This definition might seem mysterious, so we will now review the notion of uniform integrability:

Definition. A family $\mathcal{X}$ of random variables is *uniformly integrable* if and only if $\mathcal{X}$ is bounded in L^1 and

$$\sup_{\substack{X\in\mathcal{X}\\ A:\mathbb{P}(A)\leq\delta}} \mathbb{E}(|X|\mathbb{1}_A) \to 0$$

as $\delta \downarrow 0$.

Remark. Recall that to say $\mathcal{X}$ is bounded in L^p for some $p \geq 1$ means

$$\sup_{X\in\mathcal{X}} \mathbb{E}(|X|^p) < \infty.$$

There is an equivalent definition of uniform integrability that is sometimes easier to use:

Proposition. *A family $\mathcal{X}$ is uniformly integrable if and only if*

$$\sup_{X\in\mathcal{X}} \mathbb{E}(|X|\mathbb{1}_{\{|X|\geq k\}}) \to 0$$

as $k \uparrow \infty$.

. The reason to define the notion of uniform integrability is because it is precicely the condition that upgrades convergence in probability to convergence in L^1:

Theorem (Vitali). *The following are equivalent:*

(1) $X_n \to X$ *in* $L^1(\Omega, \mathcal{F}, \mathbb{P})$.
(2) $X_n \to X$ *in probability and* $(X_n)_{n\geq 1}$ *is uniformly integrable.*

Here are some sufficient conditions for uniform integrability:

Proposition. *Suppose $\mathcal{X}$ is bounded in L^p for some $p > 1$. Then $\mathcal{X}$ is uniformly integrable.*

Proof. By assumption, there is a $C > 0$ such that $\mathbb{E}(|X|^p) \leq C$ for all $X \in \mathcal{X}$. So if $\mathbb{P}(A) \leq \delta$ then by Hölder's inequality

$$\begin{aligned}\sup_{X\in\mathcal{X}} \mathbb{E}(|X|\mathbb{1}_A) &\leq \sup_{X\in\mathcal{X}} \mathbb{E}(|X|^p)^{1/p}\mathbb{P}(A)^{1-1/p}\\ &\leq C^{1/p}\delta^{1-1/p} \to 0.\end{aligned}$$

□

Proposition. *Let X be an integrable random variable on $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\mathbb{G}$ be a collection of sub-sigma-fields of $\mathcal{F}$; i.e. if $\mathcal{G} \in \mathbb{G}$, then $\mathcal{G}$ is a sigma-field and $\mathcal{G} \subseteq \mathcal{F}$. Let*

$$\mathcal{Y} = \{\mathbb{E}(X|\mathcal{G}) : \mathcal{G} \in \mathbb{G}\}$$

Then $\mathcal{Y}$ is uniformly integrable.

Proof. For all $Y \in \mathcal{Y}$ we have

$$\begin{aligned} \mathbb{P}(|Y| > k) &\leq \frac{\mathbb{E}(|Y|)}{k} \\ &\leq \frac{\mathbb{E}(|X|)}{k} \to 0 \end{aligned}$$

as $k \uparrow \infty$ by Markov's inequality and the conditional Jensen inequality. So let

$$\delta(k) = \sup_{Y \in \mathcal{Y}} \mathbb{P}(|Y| > k).$$

Now, if $Y = \mathbb{E}(X|\mathcal{G})$ then Y is $\mathcal{G}$-measurable, and particular, the event $\{|Y| > k\}$ is in $\mathcal{G}$. Then

$$\begin{aligned} \mathbb{E}(|Y| \mathbb{1}_{\{|Y|>k\}}) &\leq \mathbb{E}(|X| \mathbb{1}_{\{|Y|>k\}}) \\ &\leq \sup_{A : \mathbb{P}(A) \leq \delta(k)} \mathbb{E}(|X| \mathbb{1}_A), \end{aligned}$$

where the first line follows from the conditional Jensen inequality and iterating conditional expectation. Since the bound does not depend on which $Y \in \mathcal{Y}$ was chosen, and vanishes as $k \uparrow \infty$, the theorem is proven. □

An application of the preceding result is this:

Corollary. *Let M be a martingale. Then for any finite (non-random) $T > 0$, $(M_t)_{0 \leq t \leq T}$ is uniformly integrable.*

Proof. Note that $M_t = \mathbb{E}(M_T | \mathcal{F}_t)$ for all $0 \leq t \leq T$. □

The lesson is that on a *finite* time horizon, a martingale is well-behaved. Things are more complicated over infinite time horizons.

Theorem (Martingale convergence theorem). *Let M be a martingale (in either discrete time or in continuous time, in which case we also assume that M is right-continuous) which is bounded in L^1. Then there exists an integrable random variable M_∞ such that*

$$M_t \to M_\infty \text{ a.s. as } t \uparrow \infty.$$

The convergence is in L^1 if and only if $(M_t)_{t \geq 0}$ is uniformly integrable, in which case

$$M_t = \mathbb{E}(M_\infty | \mathcal{F}_t) \text{ for all } t \geq 0.$$

Corollary. *Non-negative martingales converge.*

Proof. We need only check L^1 boundedness:

$$\begin{aligned} \sup_{t \geq 0} \mathbb{E}(|M_t|) &= \sup_{t \geq 0} \mathbb{E}(M_t) \\ &= \mathbb{E}(M_0) < \infty. \end{aligned}$$

□

Here is an example of a martingale that converges, but that is not uniformly integrable:

Example. Let

$$X_t = e^{W_t - t/2}$$

where W is a Brownian motion in its own filtration $\mathbb{F}$. As we have discussed, X is a martingale. Since it is non-negative, it must converge. Indeed, by the Brownian law of large numbers $W_t/t \to 0$ as $t \uparrow \infty$. In particular, there exists a $T > 0$ such that $W/t < 1/4$ for all $t \geq T$. Hence, for $t \geq T$ we have

$$X_t = (e^{W_t/t - 1/2})^t \leq e^{-t/4} \to 0.$$

In this case $X_\infty = 0$. Since $X_t \neq \mathbb{E}(X_\infty|\mathcal{F}_t)$, the martingale X is *not* uniformly integrable. (This also provides an example where the inequality in Fatou's lemma is strict.)

Let's dwell on this example. Let $T_n = \inf\{t \geq 0 : X_t > n\}$, where $\inf \emptyset = +\infty$, and note these are stopping time for $\mathbb{F}$. Since $t \mapsto X_t$ is continuous and convergent as $t \to \infty$, the random variable $\sup_t X_t$ is finite-valued. In particular, for almost all ω, there is an N such that $T_n(\omega) = +\infty$ for $n \geq N$. Now X_{T_n} is well-defined, with $X_{T_n} = 0$ when $T_n = +\infty$. Also, since the stopped martingale $(X_{t \wedge T_n})_{t \geq 0}$ is bounded (and hence uniformly integrable), we have

$$X_{t \wedge T_n} = \mathbb{E}(X_{T_n}|\mathcal{F}_t)$$

The intuition is that the rare large values of X_{T_n} some how balance out the event where $X_{T_n} = 0$.

Now, to get some insight into the phenomenon that causes local martingales to fail to be true martingales, let's introduce a new process Y defined as follows: for $0 \leq u < 1$, let

$$Y_u = X_{u/(1-u)}$$

and

$$\mathcal{G}_u = \mathcal{F}_{u/(1-u)}.$$

Note that Y is just X, but running at a much faster clock. In particular, it is easy to see that $(Y_u)_{0 \leq u < 1}$ is a martingale for the filtration $(\mathcal{G}_u)_{0 \leq u < 1}$. But let's now extend for $u \geq 1$ the process by

$$Y_u = 0$$

and the filtration by

$$\mathcal{G}_u = \mathcal{F}_\infty = \sigma\left(\cup_{t \geq 0} \mathcal{F}_t\right).$$

The process $(Y_u)_{u \geq 0}$ is not a martingale over the infinite horizon, since, for instance $1 = Y_0 \neq \mathbb{E}(Y_1) = 0$. Indeed, the process $(Y_u)_{0 \leq u \leq 1}$ is not even a martingale.

The process Y actually is a *local* martingale. Indeed, let

$$U_n = \inf\{u \geq 0 : Y_u > n\}$$

where $\inf \emptyset = +\infty$ as always. We will show that $Y_{u \wedge U_n} = \mathbb{E}(Y_{U_n}|\mathcal{G}_u)$ for all $u \geq 0$ and hence the stopped process Y^{U_n} is a martingale. We will look at the two cases $u \geq 1$ and $u < 1$ separately.

Note that for $u \geq 1$, we have $\mathcal{G}_u = \mathcal{F}_\infty$ and in particular Y_{U_n} is $\mathcal{G}_u$ measurable. Hence

$$\begin{aligned}\mathbb{E}(Y_{U_n}|\mathcal{G}_u) &= Y_{U_n} \\ &= Y_{U_n} \mathbb{1}_{\{U_n < u\}} + Y_u \mathbb{1}_{\{U_n = +\infty\}} = Y_{u \wedge U_n}\end{aligned}$$

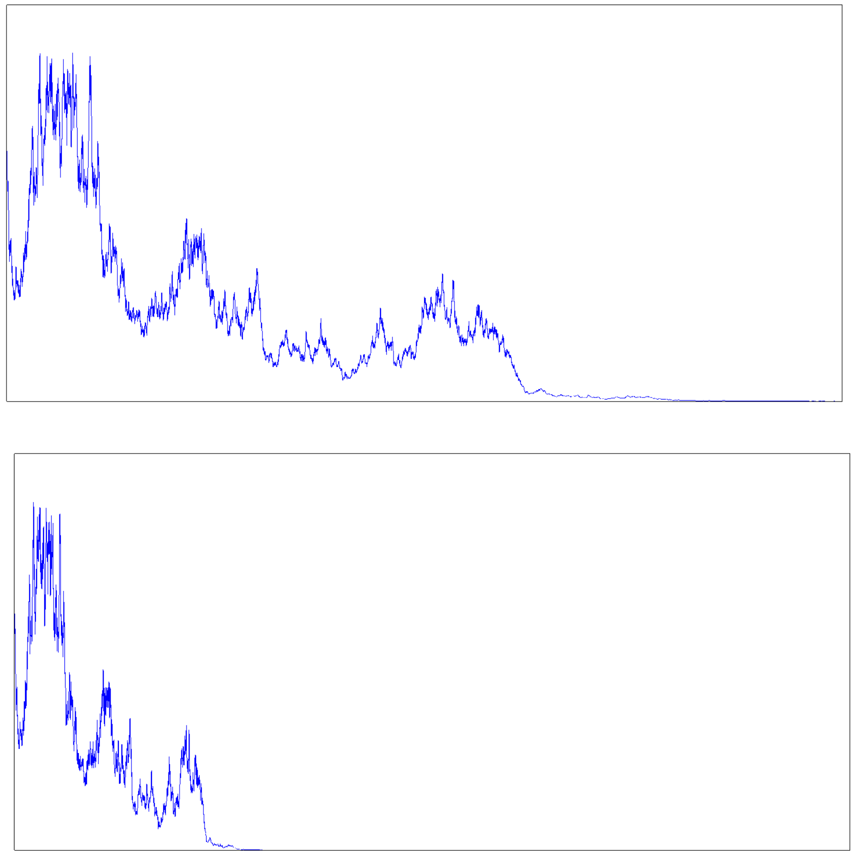


FIGURE 1. The typical graphs of $t \mapsto X_t(\omega)$, above, and $u \mapsto Y_u(\omega)$, below, for the same outcome $\omega \in \Omega$.

since U_n takes values in $[0,1) \cup \{+\infty\}$ and $Y_u = 0 = Y_\infty$ when $u \geq 1$.

For $0 \leq u < 1$, let $t = u/(1-u)$ and note

$$T_n = \begin{cases} U_n/(1-U_n) \text{ on } \{U_n < 1\} \\ +\infty \text{ on } \{U_n = +\infty\} \end{cases}$$

In particular, we have

$$\begin{aligned} \mathbb{E}(Y_{U_n} | \mathcal{G}_u) &= \mathbb{E}(X_{T_n} | \mathcal{F}_t) \\ &= X_{T_n \wedge t} \\ &= Y_{U_n \wedge u} \end{aligned}$$

as claimed.

Now we see that local martingales are martingales 'locally', and what can prevent them from being true martingales is a lack of uniform integrability. To formalise this, we introduce a useful concept:

DEFINITION. A right-continuous adapted process X is in class D (or Doob's class) iff the family of random variables

$$\{X_T : \ T \text{ a finite stopping time }\}$$

is uniformly integrable. The process is in class DL (or locally in Doob's class) iff for all $t \geq 0$ the family

$$\{X_{t\wedge T} : \ T \text{ a stopping time }\}$$

is uniformly integrable.

PROPOSITION. *A martingale is in class DL. A uniformly integrable martingale is in class D.*

PROOF. Let X be a martingale and T a stopping time. Note that by the optional sampling theorem

$$\mathbb{E}(X_t|\mathcal{F}_T) = X_{t\wedge T}.$$

In particular, we have expressed $X_{t\wedge T}$ as the conditional expectation of an integrable random variable X_t, and hence the family is uniformly integrable.

If X is uniformly integrable, we have $X_t \to X_\infty$ in L^1, so we can take the limit $t \to \infty$ to get

$$\mathbb{E}(X_\infty|\mathcal{F}_T) = X_T.$$

Hence $\{X_T : T \text{ stopping time }\}$ is uniformly integrable. □

PROPOSITION. *A local martingale X is a true martingale if X is in class DL.*

PROOF. Without loss, suppose $X_0 = 0$. Since X is a local martingale, there exists a family of stopping times $T_n \uparrow \infty$ such that X^{T_n} is a martingale. Note that $X_{t\wedge T_n} \to X_t$ a.s. for all $t \geq 0$. Now if X is in class DL, we have uniform integrability and hence we can upgrade: $X_{t\wedge T_n} \to X_t$ in L^1. In particular X_t is integrable and

$$\begin{aligned}
\mathbb{E}(X_t|\mathcal{F}_s) &= \mathbb{E}(\lim_n X_{t\wedge T_n}|\mathcal{F}_s) \\
&= \lim_n \mathbb{E}(X_{t\wedge T_n}|\mathcal{F}_s) \\
&= \lim_n X_{s\wedge T_n} \\
&= X_s.
\end{aligned}$$

□

The last thing to say about *continuous* local martingales is that we can find an explicit localising sequence of stopping times.

PROPOSITION. *Let X be a continuous local martingale, and let*

$$S_n = \inf\{t \geq 0 : |X_t - X_0| > n\}.$$

Then $(S_n)_n$ reduces $X - X_0$ to a bounded martingale.

PROOF. Again, assume $X_0 = 0$ without loss. By definition, there exists a sequence of stopping times $(T_N)_N$ increasing to ∞ such that the process X^{T_N} is a martingale for each

N. Since S_n is a stopping time, the stopped process $(X^{T_N})^{S_n} = X^{T_N \wedge S_n}$ is a martingale. Furthermore, it is bounded and hence uniformly bounded, so we have

$$\begin{aligned}\mathbb{E}[X_{S_n}|\mathcal{F}_t] &= \mathbb{E}[\lim_N X_{T_N \wedge S_n}|\mathcal{F}_t] \\ &= \lim_N \mathbb{E}[X_{T_N \wedge S_n}|\mathcal{F}_t] \\ &= \lim_N X_{T_N \wedge S_n \wedge t} \\ &= X_t^{S_n}.\end{aligned}$$

□

3. Square-integrable martingales and quadratic variation

As previewed at the beginning of the chapter, for each continuous local martingale M there is an adapted, non-decreasing process $\langle M \rangle$, called its quadratic variation, that measures the 'speed' of M in some sense. In order to construct this process, we need to first take step back and consider the space of square-integrable martingales.

We will use the notation

$$\mathcal{M}^2 = \{X = (X_t)_{t \geq 0} \text{ continuous martingale with } \sup_{t \geq 0} \mathbb{E}(X_t^2) < \infty\}$$

Note that if $X \in \mathcal{M}^2$ then by the martingale convergence theorem

$$X_t \to X_\infty \text{ a.s. and in } L^2$$

Since $(X_t^2)_{t \geq 0}$ is a submartingale, the map $t \mapsto \mathbb{E}(X_t^2)$ is increasing, and hence

$$\sup_{t \geq 0} \mathbb{E}(X_t^2) = \mathbb{E}(X_\infty^2)$$

Also, recall Doob's L^2 inequality: for $X \in \mathcal{M}^2$ we have

$$\mathbb{E}(\sup_{t \geq 0} X_t^2) \leq 4\mathbb{E}(X_\infty^2).$$

So, every element $X \in \mathcal{M}^2$ can be associated with an element $X_\infty \in L^2(\Omega, \mathcal{F}, \mathbb{P})$. Since L^2 is a Hilbert space, i.e. a complete inner product space, it is natural to ask if $\mathcal{M}^2$ has the same structure. The answer is yes. Completeness is important because we will want to take limits soon.

Theorem. *The vector space $\mathcal{M}^2$ is complete with respect to the norm*

$$\|X\|_{\mathcal{M}^2} = (\mathbb{E}[X_\infty^2])^{1/2}.$$

Proof. Let $(X^n)_n$ be a Cauchy sequence in $\mathcal{M}^2$, which means

$$\mathbb{E}[(X_\infty^n - X_\infty^m)^2] \to 0$$

as $m, n \to \infty$.

We now find a subsequence $(n_k)_k$ such that

$$\mathbb{E}[(X_\infty^{n_k} - X_\infty^{n_{k-1}})^2] \leq 2^{-k}.$$

Note that by Jensen's inequality and Doob's inequality, that

$$
\begin{aligned}
\mathbb{E}\left(\sum_k \sup_{t\geq 0}|X_t^{n_k}-X_t^{n_{k-1}}|\right) &\leq \sum_k \mathbb{E}\left(\sup_{t\geq 0}|X_t^{n_k}-X_t^{n_{k-1}}|^2\right)^{1/2} \\
&\leq \sum_k 2\mathbb{E}\left(|X_\infty^{n_k}-X_\infty^{n_{k-1}}|^2\right)^{1/2} \\
&\leq \sum_k 2^{1-k/2} < \infty.
\end{aligned}
$$

Hence, there exists an almost sure set on which the sum

$$X_t^{n_k} = X_t^{n_0} + \sum_{i=1}^k X_t^{n_i} - X_t^{n_{i-1}}$$

converges uniformly in $t \geq 0$. Since each X^{n_k} is continuous, so is the limit process X^*.

Now since the family of random variables $(X_\infty^n)_n$ is bounded in L^2, and hence uniformly integrable, we have

$$
\begin{aligned}
\mathbb{E}(X_\infty^*|\mathcal{F}_t) &= \lim_n \mathbb{E}(X_\infty^n|\mathcal{F}_t) \\
&= \lim_n X_t^n \\
&= X_t^*
\end{aligned}
$$

so X^* is a martingale. $\square$

We now introduce a notion of convergence that is well-suited to our needs:

Definition. A sequence of processes Z^n converges uniformly on compacts in probability (written u.c.p.) iff there is a process Z such that

$$\mathbb{P}(\sup_{s\in[0,t]}|Z_s^n - Z_s| > \varepsilon) \to 0$$

for all $t > 0$ and $\varepsilon > 0$.

With this, we are ready to construct the quadratic variation:

Theorem. *Let X be a continuous local martingale, and let*

$$\langle X\rangle_t^{(n)} = \sum_{k\geq 1}(X_{t\wedge t_k^n} - X_{t\wedge t_{k-1}^n})^2$$

where we will use the notation $t_k^n = k2^{-n}$. There there exists a continuous, adapted, non-decreasing process $\langle X\rangle$ such that

$$\langle X\rangle^{(n)} \to \langle X\rangle \quad u.c.p.$$

as $n \to \infty$. The process $\langle X\rangle$ is called the quadratic variation of X.

The proof will make repeated use of this elementary observation:

LEMMA (Martingale Pythagorean theorem). *Let $X \in \mathcal{M}^2$ and $(t_n)_n$ an increasing sequence such that $t_n \uparrow \infty$. Then*

$$\mathbb{E}(X_\infty^2) = \mathbb{E}(X_{t_0}^2) + \sum_{n=1}^{\infty} \mathbb{E}[(X_{t_n} - X_{t_{n-1}})^2].$$

PROOF OF LEMMA. Note that martingales have uncorrelated increments: if $i \le j-1$ then

$$\begin{aligned}\mathbb{E}[(X_{t_i} - X_{t_{i-1}})(X_{t_j} - X_{t_{j-1}})] &= \mathbb{E}[(X_{t_i} - X_{t_{i-1}})\mathbb{E}(X_{t_j} - X_{t_{j-1}} | \mathcal{F}_{t_{j-1}})] \\ &= 0\end{aligned}$$

by the tower and slot properties of conditional expectation. Hence

$$\begin{aligned}\mathbb{E}(X_{t_N}^2) &= \mathbb{E}\left[\left(X_{t_0} + \sum_{n=1}^{N} X_{t_n} - X_{t_{n-1}}\right)^2\right] \\ &= \mathbb{E}(X_{t_0}^2) + \sum_{n=1}^{N} \mathbb{E}[(X_{t_n} - X_{t_{n-1}})^2].\end{aligned}$$

Now take the supremum over N of both sides. □

PROOF OF EXISTENCE OF QUADRATIC VARIATION. There is no loss to suppose that $X_0 = 0$. For now, we will also suppose that X is uniformly bounded, so that there exists a constant $C > 0$ such that $|X_t(\omega)| \le C$ for all $(t, \omega) \in [0, \infty) \times \Omega$. This is a big assumption that will have to be removed later.

Since $X \in \mathcal{M}^2$, the limit X_∞ exists. In particular,

$$\langle X \rangle_t^{(n)} - \langle X \rangle_{2^{-n}\lfloor 2^n t \rfloor}^{(n)} = (X_t - X_{2^{-n}\lfloor 2^n t \rfloor})^2 \to 0 \text{ as } t \uparrow \infty,$$

the limit

$$\langle X \rangle_\infty^{(n)} = \sup_k \langle X \rangle_{t_k^n}^{(n)}$$

is unambiguous. By the Pythagorean theorem

$$\mathbb{E}(\langle X \rangle_\infty^{(n)}) = \mathbb{E}(X_\infty^2) \le C^2$$

so $\langle X \rangle_\infty^{(n)}$ is finite almost surely.

Now, define a new process $M^{(n)}$ by

$$M_t^{(n)} = \frac{1}{2}(X_t^2 - \langle X \rangle_t^{(n)}).$$

Note that since X is continuous, so is $M^{(n)}$. Also, by telescoping the sum, we have

$$M_t^{(n)} = \sum_{k=1}^{\infty} X_{t_{k-1}^n}(X_{t \wedge t_k^n} - X_{t \wedge t_{k-1}^n}).$$

In particular, $M^{(n)}$ is a martingale, since each term[1] in the sum is. By the Pythagorean theorem

$$\begin{aligned}\mathbb{E}[(M^{(n)}_\infty)^2] &= \mathbb{E}\sum_{k=1}^{\infty} X^2_{t^n_{k-1}}(X_{t^n_k} - X_{t^n_{k-1}})^2 \\ &\le C^2\mathbb{E}\sum_{k=1}^{\infty}(X_{t^n_k} - X_{t^n_{k-1}})^2 \\ &= C^2\mathbb{E}(X^2_\infty) \\ &\le C^4\end{aligned}$$

since X is bounded by C.

For future use, note that

$$\begin{aligned}\mathbb{E}[(\langle X\rangle^{(n)}_\infty)^2] &= \mathbb{E}[(X^2_\infty - 2M^{(n)}_\infty)^2] \\ &\le 2\mathbb{E}[X^4_\infty] + 8\mathbb{E}[(M^{(n)}_\infty)^2] \\ &\le 10C^4\end{aligned}$$

from the boundedness of X and the previous estimate.

By some more rearranging of terms, we have the formula

$$M^{(n)}_\infty - M^{(m)}_\infty = \sum_{j=1}^{\infty}(X_{j2^{-n}} - X_{\lfloor j2^{m-n}\rfloor 2^{-m}})(X_{(j+1)2^{-n}} - X_{j2^{-n}}),$$

for $n > m$, and once more by the Pythagorean theorem, letting $Z_m = \sup_{|s-t|\le 2^{-m}} |X_s - X_t|$

$$\begin{aligned}\mathbb{E}[(M^{(n)}_\infty - M^{(m)}_\infty)^2] &= \mathbb{E}\left[\sum_{j=1}^{\infty}(X_{j2^{-n}} - X_{\lfloor j2^{m-n}\rfloor 2^{-m}})^2(X_{(j+1)2^{-n}} - X_{j2^{-n}})^2\right] \\ &\le \mathbb{E}\left[\sup_{|s-t|\le 2^{-m}}(X_s - X_t)^2\sum_{j=1}^{\infty}(X_{(j+1)2^{-n}} - X_{j2^{-n}})^2\right] \\ &= \mathbb{E}\left[Z^2_m\langle X\rangle^{(n)}_\infty\right] \\ &\le \mathbb{E}\left[Z^4_m\right]^{1/2}\mathbb{E}\left[(\langle X\rangle^{(n)}_\infty)^2\right]^{1/2} \\ &\le \sqrt{10}C^2\mathbb{E}(Z^4_m)^{1/2}\end{aligned}$$

by the Cauchy–Schwarz inequality and the previous estimate. Now

$$Z_m = \sup_{|s-t|\le 2^{-m}}|X_s - X_t| \to 0 \text{ a.s.}$$

by the uniform continuity[2] X, and that $|Z_m| \le 2C$, so that $\mathbb{E}(Z^4_m) \to 0$ by the dominated convergence theorem. This shows that the sequence $(M^{(n)})_n$ is Cauchy.

---

[1]If Y is a uniformly integrable martingale and K is bounded and $\mathcal{F}_{t_0}$-measurable then $K(Y_t - Y_{t\wedge t_0})$ defines a uniformly integrable martingale. We need only show $\mathbb{E}[K(Y_\infty - Y_{t_0})|\mathcal{F}_t] = K(Y_t - Y_{t\wedge t_0})$ which can be checked by considering the cases $t\in[0,t_0]$ and $t\in(t_0,\infty)$ separately.

[2]The map $t\mapsto X_t$ is continuous by assumption, and it is uniformly continuous since $X_t \to X_\infty$.

By the completeness of $\mathcal{M}^2$, there exists a limit continuous martingale M^*. Let

$$\langle X \rangle = X^2 - 2M^*.$$

Note that $\langle X \rangle$ is continuous and adapted, since the right-hand side is. By Doob's maximal inequality

$$\begin{aligned}\mathbb{E}[\sup_{t\geq 0}(\langle X \rangle_t^{(n)} - \langle X \rangle_t)^2] &= 4\mathbb{E}[\sup_{t\geq 0}(M_t^{(n)} - M_t^*)^2] \\ &\leq 16\mathbb{E}[(M_\infty^{(n)} - M_\infty^*)^2] \to 0\end{aligned}$$

so that

$$\langle X \rangle^{(n)} \to \langle X \rangle \text{ uniformly in } L^2.$$

By passing to a subsequence, we can assert uniform almost sure convergence. Since

$$\langle X \rangle_t^{(n)} - \langle X \rangle_{2^{-n}\lfloor 2^n t \rfloor}^{(n)} = (X_t - X_{2^{-n}\lfloor 2^n t \rfloor})^2 \to 0 \text{ as } n \uparrow \infty,$$

by the continuity of X, and $t \mapsto \langle X \rangle_{2^{-n}\lfloor 2^n t \rfloor}^{(n)}$ is obviously non-decreasing, the almost sure limit $\langle X \rangle$ is also non-decreasing.

Now we will remove the assumption that X is bounded. For every $N \geq 1$ let

$$T_N = \inf\{t \geq 0 : |X_t| > N\}.$$

Note that for every N, the process X^{T_N} is a bounded martingale. Hence the process $\langle X^{T_N} \rangle$ is well-defined by the above construction. But since

$$\langle X^{T_{N+1}} \rangle_t^{(n)} - \langle X^{T_N} \rangle_t^{(n)} \begin{cases} = 0 & \text{if } t \leq T_N \\ \geq 0 & \text{if } t > T_N \end{cases}$$

for each n, just by inspecting the definition, we also have

$$\langle X^{T_{N+1}} \rangle_t - \langle X^{T_N} \rangle_t \begin{cases} = 0 & \text{if } t \leq T_N \\ \geq 0 & \text{if } t > T_N \end{cases}$$

for the limit. In particular, the monotonicity allows us to define $\langle X \rangle$ by

$$\begin{aligned}\langle X \rangle_t &= \sup_N \langle X^{T_N} \rangle_t \\ &= \lim_{N\to\infty} \langle X^{T_N} \rangle_t.\end{aligned}$$

That $\langle X \rangle$ is adapted and non-decreasing follows from the above definition. And since $\langle X \rangle_t = \langle X^{T_N} \rangle_t$ on $\{t \leq T_N\}$, and the stopping times $T_N \to \infty$ as $N \to \infty$, we can conclude that $\langle X \rangle$ is continuous also.

Finally, for any $t > 0$ and $\epsilon > 0$, we have

$$\begin{aligned}\mathbb{P}(\sup_{s\in[0,t]} |\langle X \rangle_s - \langle X \rangle_s^{(n)}| > \varepsilon) &\leq \mathbb{P}(\sup_{s\in[0,t]} |\langle X \rangle_s - \langle X \rangle_s^{(n)}| > \epsilon, \text{ and } T_N \geq t) + \mathbb{P}(T_N < t) \\ &\leq \mathbb{P}(\sup_{s\geq 0} |\langle X^{T_N} \rangle_s - \langle X^{T_N} \rangle_s^{(n)}| > \varepsilon) + \mathbb{P}(T_N < t).\end{aligned}$$

By first sending $n \to \infty$ then $N \to \infty$, we establish the u.c.p. convergence. □

Now that we have constructed the quadratic variation process, we explore some of its properties.

Proposition. *For every $X \in \mathcal{M}^2$, the process*

$$M = X^2 - \langle X \rangle$$

is a continuous, uniformly integrable martingale. In particular,

$$\mathbb{E}[\langle X \rangle_\infty] = \mathbb{E}[(X_\infty - X_0)^2] = \mathbb{E}[X_\infty^2] - \mathbb{E}[X_0^2].$$

Proof. Assume without loss that $X_0 = 0$. Let

$$T_N = \inf\{t \geq 0 : |X_t| > N\}$$

The proof from last time shows that the process

$$(X^{T_N})^2 - \langle X^{T_N} \rangle$$

is a continuous square-integrable martingale for all N, and hence

$$\mathbb{E}(X_{T_N}^2 - \langle X \rangle_{T_N} | \mathcal{F}_t) = X_{T_N \wedge t}^2 - \langle X \rangle_{T_N \wedge t}$$

Since $X_{T_N}^2 \leq \sup_{t \geq 0} X_t^2$ for all N, which is assumed integrable, the first term above converges by the dominated convergence theorem. Also, since $N \mapsto \langle X \rangle_{T_N}$ is non-negative and non-decreasing, second term above converges as $N \to \infty$ by the monotone convergence theorem. Hence

$$\mathbb{E}(M_\infty | \mathcal{F}_t) = M_t$$

This shows M is a uniformly integrable martingale. □

Assuming that we can calculate the quadratic variation of a given local martingale, the following theorem gives a useful sufficient condition to check whether the local martingale is actually a true martingale.

Proposition. *Suppose X is a continuous local martingale with $X_0 = 0$. If $\mathbb{E}(\langle X \rangle_\infty) < \infty$ then $X \in \mathcal{M}^2$ and*

$$\mathbb{E}[\langle X \rangle_\infty] = \mathbb{E}[X_\infty^2].$$

Proof. Suppose $\mathbb{E}[\langle X \rangle_\infty] < \infty$. Then, by the monotone convergence theorem

$$\begin{aligned}
\mathbb{E}[\sup_{t \geq 0} X_t^2] &= \lim_N \mathbb{E}[\sup_{0 \leq t \leq T_N} X_t^2] \\
&= \lim_N \mathbb{E}[\sup_{t \geq 0} (X_t^{T_N})^2] \\
&\leq \lim_N 4\ \mathbb{E}[(X_{T_N}^2] \\
&= \lim_N 4\ \mathbb{E}[\langle X \rangle_{T_N}] \\
&= 4\ \mathbb{E}[\langle X \rangle_\infty] < \infty
\end{aligned}$$

where Doob's inequality was used in the third line. In particular, since for any stopping time X_T is dominated by the integrable random variable $\sup_{u \geq 0} |X_u|$, the local martingale X is class D and hence is a true martingale. □

Corollary. *If $\mathbb{E}(\langle X \rangle_t) < \infty$ for all $t \geq 0$, then X is a true martingale and*

$$\mathbb{E}[X_t^2] = \mathbb{E}(\langle X \rangle_t).$$

Proof. Apply the previous proposition to the stopped process $X^t = (X_s)_{0 \leq s \leq t}$. □

REMARK. Be careful with the wording of the preceding results! In particular, we will see an example of a continuous local martingale X such that $\sup_{t \geq 0} \mathbb{E}(X_t^2) < \infty$ but is *not* a true martingale. Indeed, for this example $\mathbb{E}(\langle X \rangle_t) = \infty$ for all $t > 0$.

The following corollary is easy, but is surprisingly useful:

COROLLARY. *If X is a continuous local martingale such that $\langle X \rangle_t = 0$ for all t, then $X_t = X_0$ a.s. for all $t \geq 0$.*

PROOF. By the previous proposition, since $\mathbb{E}(\langle X \rangle_\infty) < \infty$ then

$$\sup_{t \geq 0} \mathbb{E}[(X_t - X_0)^2] = \mathbb{E}(\langle X \rangle_\infty) = 0.$$

□

The next result needs a definition. Recall that we are interested in developing an integration theory for (random) functions on $[0, \infty)$. The following definition is intimately related to the Lebesgue–Stieltjes integration theory.

DEFINITION. For a function $f : [0, \infty) \to \mathbb{R}$, define $\|f\|_{t,\mathrm{var}}$ by

$$\|f\|_{t,\mathrm{var}} = \sup_{0 \leq t_0 < \ldots < t_n \leq t} \sum_{k=1}^{n} |f(t_k) - f(t_{k-1})|.$$

A function f is of finite variation iff $\|f\|_{t,\mathrm{var}} < \infty$ for all $t \geq 0$.

EXAMPLE. If f is monotone, then f is of finite variation. Indeed, in this case

$$\|f\|_{t,\mathrm{var}} = |f(t) - f(0)|.$$

The above example essentially classifies all finite variation functions:

PROPOSITION. *If f is of finite variation, then $f = g - h$ where g and h are non-decreasing.*

PROOF. For a partition $0 \leq t_0 < t_1 < \ldots < t_n \leq s < t$ we have

$$\sum_{k=1}^{n} |f(t_k) - f(t_{k-1})| + |f(t) - f(s)| \leq \|f\|_{t,\mathrm{var}}$$

by definition. Taking the supremum over the sub-partition of $[0, s]$ yields

$$(*) \qquad \|f\|_{s,\mathrm{var}} + |f(t) - f(s)| \leq \|f\|_{t,\mathrm{var}}$$

Hence

$$\begin{aligned} \|f\|_{s,\mathrm{var}} - f(s) &\leq \|f\|_{s,\mathrm{var}} + |f(t) - f(s)| - f(t) \\ (**) \qquad &\leq \|f\|_{t,\mathrm{var}} - f(t). \end{aligned}$$

Let $g(t) = \|f\|_{t,\mathrm{var}}$ and $h(t) = g(t) - f(t)$. By (*) g is non-decreasing. and by (**) so is h. □

Now back to stochastic processes...

PROPOSITION. *If X is a continuous local martingale of finite variation (i.e. $t \mapsto X_t(\omega)$ is of finite-variation for almost all ω), then X is almost surely constant.*

Proof. Since the quadratic variation process $\langle X \rangle$ is the u.c.p. limit of processes $\langle X \rangle^{(n)}$, for any fixed $t \geq 0$, we can find a subsequence such that $\langle X \rangle_t^{(n)} \rightarrow \rangle X \rangle_t$ a.s Now

$$\begin{aligned} \langle X \rangle_t &= \lim_n \sum_{k=1}^{\infty} (X_{t \wedge t_k^n} - X_{t \wedge t_{k-1}^n})^2 \\ &\leq \|X\|_{t,\text{var}} \limsup_n \sup_{\substack{s_1, s_2 \in [0,t] \\ |s_1 - s_2| \leq 2^{-n}}} |X_{s_1} - X_{s_2}| = 0 \end{aligned}$$

by the uniform continuity of $t \mapsto X_t(\omega)$ on the compact $[0,t]$ for almost all ω and the assumption that $\|X(\omega)\|_{t,\text{var}} < \infty$ for almost all ω.

The result now follows from the above corollary. □

Finally, an important application of this proposition.

Theorem. *(Characterisation of quadratic variation) Suppose X is a continuous local martingale and A is a continuous adapted process of finite variation such that $A_0 = 0$. The process $X^2 - A$ is a local martingale if and only if $A = \langle X \rangle$.*

Example. It is an easy exercise to see that $W_t^2 - t$ is a martingale, where W is a Brownian motion. Hence $\langle W \rangle_t = t$. We will shortly see a striking converse of this result due to Lévy.

Proof. Suppose that $X_0 = 0$. Let $(T_n)_n$ reduce X to a square-integrable martingale. Then Since $(X^{T_n})^2 - \langle X^{T_n} \rangle = (X^2 - \langle X \rangle)^{T_n}$ is a martingale, then $X^2 - \langle X \rangle$ is a local martingale. For general X_0, note that $X^2 - \langle X \rangle = (X - X_0)^2 - \langle X \rangle + 2X_0(X - X_0) + X_0^2$. The term $2X_0(X - X_0) + X_0^2$ is also local martingale, proving the 'if' direction.

Now suppose that $X^2 - A$ is a local martingale. Then $(X^2 - \langle X \rangle) - (X^2 - A) = A - \langle X \rangle$ is a local martingale. But $A - \langle X \rangle$ is of finite variation. Hence, the difference $A - \langle X \rangle$ is almost surely constant. □

4. The stochastic integral

Now with our preparatory remarks about local martingales and quadratic variation out of the way, we are ready to construct the stochastic integral, the main topic of this course.

We first introduce our first space of integrands.

Definition. A simple predictable process $Y = (Y_t)_{t>0}$ is of the form

$$Y_t(\omega) = \sum_{k=1}^{n} H_k(\omega) \mathbb{1}_{(t_{k-1}, t_k]}(t)$$

where $0 \leq t_0 < \ldots < t_n$ are not random, and for each k the random variable H_k is bounded and $\mathcal{F}_{t_{k-1}}$ measurable. That is to say, Y is an adapted process with left-continuous, piece-wise constant sample paths.

Notation. The collection of all simple predictable processes will be denoted $\mathcal{S}$. The collection of continuous local martingales will be denoted $\mathcal{M}_{\text{loc}}$.

And here is the first definition of the integral:

Definition. If $X \in \mathcal{M}_{\text{loc}}$ and $Y \in \mathcal{S}$, then the stochastic integral is defined by

$$\int_0^t Y_s dX_s = \sum_{k=1}^n H_k (X_{t\wedge t_k} - X_{t\wedge t_{k-1}})$$

Proposition. *If $X \in \mathcal{M}_{\text{loc}}$ and $Y \in \mathcal{S}$ then $\int Y dX$ is a continuous local martingale and*

$$\begin{aligned}\left\langle \int Y dX \right\rangle_t &= \int_0^t Y_s^2 d\langle X\rangle_s \\ &= \sum_{k=1}^n H_k^2 (\langle X\rangle_{t\wedge t_k} - \langle X\rangle_{t\wedge t_{k-1}}).\end{aligned}$$

Proof. For of all, note that there is no loss assuming $X_0 = 0$. Now, to show that $\int Y dX$ is a local martingale, we need to find a sequence of stopping times $(T_n)_n$ such that the $(\int Y dX)^{T_n}$ is a martingale. The natural choice is $T_n = \inf\{t \geq 0 : |X_t| > n\}$. Therefore, we need only show that if X is a bounded martingale then so is $\int Y dX$. This can be confirmed term by term: if $s \leq t_{k-1}$ then

$$\begin{aligned}\mathbb{E}[H_k(X_{t_k} - X_{t_{k-1}})|\mathcal{F}_s] &= \mathbb{E}[H_k \mathbb{E}(X_{t_k} - X_{t_{k-1}}|\mathcal{F}_{t_{k-1}})|\mathcal{F}_s] \\ &= 0 \\ &= H_k(X_{s\wedge t_k} - X_{s\wedge t_{k-1}})\end{aligned}$$

and if $s \geq t_{k-1}$ then

$$\begin{aligned}\mathbb{E}[H_k(X_{t_k} - X_{t_{k-1}})|\mathcal{F}_s] &= H_k \mathbb{E}(X_{t_k} - X_{t_{k-1}}|\mathcal{F}_s) \\ &= H_k(X_{s\wedge t_k} - X_{s\wedge t_{k-1}})\end{aligned}$$

where in both cases the assumption that the random variable H_k is bounded and $\mathcal{F}_{t_{k-1}}$-measurable is used to justify pulling it out of the conditional expectations.

The verification of the quadratic variation formula is left as an exercise. One method is to use the characterisation of quadratic variation, and show that

$$\left(\int Y dX\right)^2 - \int Y^2 d\langle X\rangle$$

is a local martingale. By localisation, we can assume X is bounded as before. Now apply the Pythagorean theorem. Another method is to show that the sum of the squares of the increments of the integral converge to the correct expression. □

Corollary (Itô's isometry). *If $Y \in \mathcal{S}$ and $X \in \mathcal{M}^2$ then $\int Y dX \in \mathcal{M}^2$ and*

$$\mathbb{E}\left[\left(\int_0^\infty Y_s dX_s\right)^2\right] = \mathbb{E}\left[\int_0^t Y_s^2 d\langle X\rangle_s\right]$$

Proof. Since Y is bounded and has a bounded number of jumps and since X is square integrable, one checks that

$$\mathbb{E}\left[\left(\sup_{t\geq 0}\int_0^t Y_s dX_s\right)^2\right] < \infty$$

by the triangle inequality, for instance. Hence $\left(\int Y dX\right)^2 - \int Y^2 d\langle X\rangle$ is a uniformly integrable martingale and hence the claim. □

The next stage in the programme is to enlarge the space of integrands. The idea is that we should be able define the stochastic integral of certain limits of simple predictable processes by using the completeness of $\mathcal{M}^2$. What limits should we consider?

There are several ways to view a stochastic process Y. For instance, either as a collection of random variables $\omega \to Y_t(\omega)$ for each $t \geq 0$, or as a collection of sample paths $t \to Y_t(\omega)$ for each $\omega \in \Omega$. There is a third view that will prove fruitful: consider $(t, \omega) \mapsto Y_t(\omega)$ as a function on $[0, \infty) \times \Omega$. Since we are interested in probability, we need to define a sigma-field:

DEFINITION. The predictable sigma-field, denoted $\mathcal{P}$ is the sigma-field generated by sets of the form $(s, t] \times A$ where $0 \leq s \leq t$ and $A \in \mathcal{F}_s$. Equivalently, the predictable sigma-field is the smallest sigma-field for which each simple predictable process $(t, \omega) \mapsto Y_t(\omega)$ is measurable.

Now motivated from the discussion above, for fixed $X \in \mathcal{M}^2$, we construct a measure μ_X on the predictable sigma-field $\mathcal{P}$ as follows: For almost every $\omega \in \Omega$, the map $t \mapsto \langle X\rangle_t(\omega)$ is continuous and non-decreasing. Hence, we can associate with it the Lebesgue-Stieltjes measure

$$\langle X\rangle(\cdot, \omega)$$

on the Borel subsets of $[0, \infty)$. This measure has the property that

$$\langle X\rangle((s, t], \omega) = \langle X\rangle_t(\omega) - \langle X\rangle_s(\omega).$$

We can use this measure as a kernel, and define μ_X on $\mathcal{P}$ by

$$\mu_X(dt \times d\omega) = \langle X\rangle(dt, \omega)\mathbb{P}(d\omega).$$

In particular, we have the following equality

$$\mu_X((s, t] \times A) = \mathbb{E}[(\langle X\rangle_t - \langle X\rangle_s)\mathbb{1}_A]$$

for $A \in \mathcal{F}_s$ and $0 \leq s \leq t$. Finally, we let $L^2(X) = L^2([0, \infty) \times \Omega, \mathcal{P}, \mu_X)$ denote the Hilbert space of square-integrable predictable processes with norm

$$\begin{aligned}\|Y\|^2_{L^2(X)} &= \int_{[0,\infty)\times\Omega} Y^2 d\mu_X \\ &= \mathbb{E}\int_0^\infty Y_s^2 d\langle X\rangle_s.\end{aligned}$$

In particular, if we want to extend the definition of the stochastic integral to include integrands more general then simple predictable processes, a natural place to look is the space $L^2(X)$.

We will use the following notation:

$$\text{if } X \in \mathcal{M}^2 \text{ then } \|X\|_{\mathcal{M}^2} = \mathbb{E}(X_\infty^2)^{1/2}.$$

PROPOSITION. *For simple predictable $Y \in \mathcal{S}$ and $X \in \mathcal{M}^2$ we have*

$$\|Y\|_{L^2(X)} = \left\| \int Y dX \right\|_{\mathcal{M}^2}.$$

PROOF. This is just Itô's isometry in different notation. □

Proposition. *Given a sequence of simple predictable processes $Y_n \in \mathcal{S}$ which converge $Y_n \to Y$ in $L^2(X)$, there exists a martingale $M \in \mathcal{M}^2$ such that*

$$\int Y^n dX \to M \text{ in } \mathcal{M}^2.$$

Furthermore, if $\tilde{Y}_n \to Y$ in $L^2(X)$ and $\int \tilde{Y} dX \to \tilde{M}$ in $\mathcal{M}^2$, then $M = \tilde{M}$.

Proof. Note that the sequence of integrands Y_n is Cauchy in $L^2(X)$. By Itô's isometry, the sequence of integrals $\int Y_n dX$ is Cauchy in $\mathcal{M}^2$. And since we know that the space $\mathcal{M}^2$ of continuous square-integrable martingales is complete, we can assert the existence of a limit M to the Cauchy sequence.

For the second part, we just use a standard argument:

$$\begin{aligned}
\|M - \tilde{M}\|_{\mathcal{M}^2} &\le \|M - M_n\|_{\mathcal{M}^2} + \|\tilde{M} - \tilde{M}_n\|_{\mathcal{M}^2} + \|M_n - \tilde{M}_n\|_{\mathcal{M}^2} \\
&= \|M - M_n\|_{\mathcal{M}^2} + \|\tilde{M} - \tilde{M}_n\|_{\mathcal{M}^2} + \|Y_n - \tilde{Y}_n\|_{L^2(X)} \\
&\le \|M - M_n\|_{\mathcal{M}^2} + \|\tilde{M} - \tilde{M}_n\|_{\mathcal{M}^2} + \|Y - Y_n\|_{L^2(X)} + \|Y - \tilde{Y}_n\|_{L^2(X)} \\
&\to 0
\end{aligned}$$

where $M_n = \int Y_n dX$ and $\tilde{M}_n = \int \tilde{Y}_n dX$ and hence

$$M_n - \tilde{M}_n = \int (Y_n - \tilde{Y}_n) dX.$$

□

Now we state a standard result from measure theory:

Proposition. *Let μ be a finite measure on a measurable space $(E, \mathcal{E})$, and let $\mathcal{A}$ be a π-system of subsets of E generating the sigma-field $\mathcal{E}$. Then the set of functions of the form*

$$\sum_{i=1}^{n} a_i \mathbb{1}_{A_i}, \quad a_i \in \mathbb{R} \text{ and } A_i \in \mathcal{A},$$

is dense in $L^p(E, \mathcal{E}, \mu)$ for any $p \ge 1$.

Corollary. *When $X \in \mathcal{M}^2$, the simple predictable process $\mathcal{S}$ are dense in $L^2(X)$. That is to say, for any predictable process $Y \in L^2(X)$, we can find a sequence $(Y_n)_n$ of simple predictable processes such that*

$$\|Y - Y_n\|_{L^2(X)} \to 0.$$

The above proposition justifies this definition:

Definition. For a martingale $X \in \mathcal{M}^2$ and a square-integrable predictable process Y which is the $L^2(X)$ limit of a sequence $Y_n \in S$, the stochastic integral $\int Y dX$ is defined to be the (unique) $\mathcal{M}^2$ limit of the sequence $\int Y_n dX$.

The next result says that the stochastic integral behaves nicely with respect to stopping:

Theorem. *Fix $X \in \mathcal{M}^2$ and $Y \in L^2(X)$, and a stopping time T. Then $X^T \in \mathcal{M}^2$ and $Y \mathbb{1}_{(0,T]} \in L^2(\mu_X)$ and*

(1) $\langle X^T \rangle = \langle X \rangle^T$

(2) $\int Y \mathbb{1}_{(0,T]} dX = \int Y dX^T = \left(\int Y dX\right)^T$

Proof. By the optional sampling theorem, the stopped martingale X^T is still a martingale, and by Doob's inequality

$$\begin{aligned}\|X^T\|^2_{\mathcal{M}^2} = \mathbb{E}(X_T^2) &\le \mathbb{E}(\sup_{t\ge 0} X_t^2)\\ &\le 4\mathbb{E}(X_\infty^2) = 4\|X\|_{\mathcal{M}^2} < \infty.\end{aligned}$$

and hence $X^T \in \mathcal{M}^2$ as claimed.

Also the process $\mathbb{1}_{(0,T]}$ is adapted and left-continuous, so it is predictable. And since Y is predictable by assumption, so is the product $Y\mathbb{1}_{(0,T]}$. We also have the calculation

$$\mathbb{E}\int_0^\infty (Y_s \mathbb{1}_{(0,T]}(s))^2 d\langle X\rangle_s \le \mathbb{E}\int_0^\infty Y_s^2 d\langle X\rangle_s < \infty$$

so $Y\mathbb{1}_{(0,T]} \in L^2(X)$ as claimed.

(1) This is an exercise.

(2) First we show that $\boxed{\int Y dX^T = \left(\int Y dX\right)^T}$.

Case: Simple predictable $Y \in \mathcal{S}$. This is obvious. Indeed, let Y have the representation

$$Y = \sum_{k=1}^n H_k \mathbb{1}_{(t_{k-1},t_k]}.$$

Then

$$\begin{aligned}\int_0^t Y_s dX_s^T &= \sum_{k=1}^n H_k (X^T_{t\wedge t_k} - X^T_{t\wedge t_{k-1}})\\ &= \sum_{k=1}^n H_k (X_{T\wedge t\wedge t_k} - X^T_{T\wedge t\wedge t_{k-1}})\\ &= \left(\int Y dX\right)^T_t\end{aligned}$$

Case: General $Y \in L^2(X)$. Let $(Y^n)_n$ be a sequence in $\mathcal{S}$ converging to Y in $L^2(X)$. We have already shown that $\int Y^n dX^T = \left(\int Y^n dX\right)^T$ for each n.

- $\int Y^n dX^T \to \int Y dX^T$ in $\mathcal{M}^2$. Proof: $Y^n \to Y$ in $L^2(X^T)$ also, since

$$\begin{aligned}\mathbb{E}\int_0^\infty (Y_s^n - Y_s)^2 d\langle X^T\rangle_s &= \mathbb{E}\int_0^T (Y_s^n - Y_s)^2 d\langle X\rangle_s\\ &\le \mathbb{E}\int_0^\infty (Y_s^n - Y_s)^2 d\langle X\rangle_s \to 0.\end{aligned}$$

  And by the definition of stochastic integral, $\int Y^n dX^T \to \int Y dX^T$.
- $\left(\int Y^n dX\right)^T \to \left(\int Y dX\right)^T$ in $\mathcal{M}^2$. Proof: note that if $M^n \to M$ in $\mathcal{M}^2$, then $(M^n)^T \to M^T$ also, since

$$\|(M^n)^T - M^T\|_{\mathcal{M}^2} = \|(M^n - M)^T\|_{\mathcal{M}^2} \le 2\|M^n - M\|_{\mathcal{M}^2}.$$

And since $\int Y^n dX \to \int Y dX$ by the definition of stochastic integral, we are done.

Now show that $\boxed{\int Y \mathbb{1}_{(0,T]} dX = \left(\int Y dX\right)^T}$.

Case: Simple predictable $Y \in \mathcal{S}$ and T takes only a finite number of values $0 \leq s_1 < \cdots < s_N$. The process $\mathbb{1}_{(0,T]}$ is a simple and predictable since

$$\begin{aligned}\mathbb{1}_{(0,T]} &= \sum_{k=1}^{N} \mathbb{1}_{\{T=s_k\}} \mathbb{1}_{(0,s_k]} \\ &= \sum_{k=1}^{N} \mathbb{1}_{\{T>s_{k-1}\}} \mathbb{1}_{(s_{k-1},s_k]}\end{aligned}$$

where $s_0 = 0$. Note that $\{T > s_{k-1}\} = \{T \leq s_{k-1}\}^c \in \mathcal{F}_{s_{k-1}}$ since T is a stopping time. Consequently, the product $Y \mathbb{1}_{(0,T]} \in \mathcal{S}$. As before, we have the identity

$$\int_0^t Y_s \mathbb{1}_{(0,T]}(s) dX_s = \left(\int Y dX\right)_t^T$$

by expanding both sides into finite sums and routine book-keeping.
Case: Simple predictable $Y \in \mathcal{S}$ and general stopping time T. The process $Y \mathbb{1}_{(0,T]}$ is generally not simple. So we approximate it by $T_n = (2^{-n} \lceil 2^n T \rceil) \wedge n$. Note that T_n is a stopping time taking only a finite number of values since

$$\{T_n \leq t\} = \begin{cases} \{T \leq k2^{-n}\} & \text{if } k2^{-n} \leq t < (k+1)2^{-n}, k < n2^n \\ \Omega & \text{if } t \geq n \end{cases}$$

- $\int Y \mathbb{1}_{(0,T_n]} dX \to \int Y \mathbb{1}_{(0,T]} dX$ in $\mathcal{M}^2$. Proof. Note that $Y \mathbb{1}_{(0,T_n]} \to Y \mathbb{1}_{(0,T]}$ in $L^2(X)$ since $T_n \to T$ a.s. and
$$\mathbb{E} \int_0^\infty (Y_s \mathbb{1}_{(0,T_n]}(s) - Y_s \mathbb{1}_{(0,T]}(s))^2 d\langle X \rangle_s \to 0$$
by the dominated convergence theorem. Now apply the definition of stochastic integral.
- $\left(\int Y dX\right)^{T_n} \to \left(\int Y dX\right)^T$ in $\mathcal{M}^2$. Proof. Let $M = \int Y dX$. Then $|M_{T_n} - M_T|^2$ converges to 0 a.s. and is dominated by the integrable random variable $2 \sup_{t \geq 0} M_t^2$, so the result follows from the dominated convergence theorem.

Case: General $Y \in L^2(X)$ and general stopping time T. Let $(Y^n)_n$ be a sequence in $\mathcal{S}$ converging to Y in $L^2(X)$. We have already shown $(\int Y^n dX)^T \to (\int Y dX)^T$. So we now show $\int Y^n \mathbb{1}_{(0,T]} dX \to \int Y \mathbb{1}_{(0,T]} dX$. Proof. We need only show that $Y^n \mathbb{1}_{(0,T]} \to Y \mathbb{1}_{(0,T]}$ in $L^2(X)$. But

$$\mathbb{E} \int_0^\infty (Y_s^n \mathbb{1}_{(0,T]}(s) - Y_s \mathbb{1}_{(0,T]}(s))^2 d\langle X \rangle_s \leq \mathbb{E} \int_0^\infty (Y_s^n - Y_s)^2 d\langle X \rangle_s \to 0.$$

□

The above proof is a bit tedious, but the result allows us to extend the definition of stochastic integral in considerably. Before making this extension, we make an easy observation:

Proposition. *Let $X \in \mathcal{M}^2$ and $Y \in L^2(X)$. If S and T are stopping times such that $S \le T$ a.s., then*

$$\int_0^t Y_s \mathbb{1}_{(0,S]}(s) dX_s^S = \int_0^t Y_s \mathbb{1}_{(0,T]}(s) dX_s^T$$

on the event $\{t \le S\}$.

Proof. The left integral is $\left(\int Y dX\right)_{t \wedge S}$ and the right is $\left(\int Y dX\right)_{t \wedge T}$ so they agree on $\{t \le S\}$. □

Proposition. *Suppose X is a continuous local martingale with $X_0 = 0$ and Y is a predictable process such that*

$$\int_0^t Y_s^2 d\langle X\rangle_s < \infty \text{ a.s. for all } t \ge 0.$$

Let

$$T_n = \inf\left\{t \ge 0 : |X_t| > n \text{ or } \int_0^t Y_s^2 d\langle X\rangle_s > n\right\}.$$

Then $(T_n)_n$ is an increasing sequence of stopping times which have the property that $X^{T_n} \in \mathcal{M}^2$ and $Y\mathbb{1}_{(0,T_n]} \in L^2(X^{T_n})$. Let

$$M^{(n)} = \int Y \mathbb{1}_{(0,T_n]} dX^{T_n}.$$

Then there is a continuous local martingale M such that $M^{(n)} \to M$ u.c.p.

Proof. By the preceding proposition on the event $\{T_n \ge t\}$ we have that $M_t^{(n)} = M_t^{(N)}$ for all $N \ge n$. Hence for every $t \ge 0$ the sequence $M_t^{(n)}$ converges almost surely to a random variable M_t. In particular, we have

$$\mathbb{P}\big(\sup_{s\in[0,t]} |M_s^{(n)} - M_s| > \varepsilon\big) \le \mathbb{P}(T_n > t) \to 0.$$

for every $\varepsilon > 0$. Note that $M^{T_n} = M^{(n)}$, and since M^{T_n} is a continuous martingale for each n, the limit process M is a continuous local martingale. □

Definition. Suppose X is a continuous local martingale with $X_0 = 0$ and Y is a predictable process such that

$$\int_0^t Y_s^2 d\langle X\rangle_s < \infty \text{ a.s. for all } t \ge 0.$$

Then the stochastic integral

$$\int Y dX$$

is defined to be the u.c.p. limit of $\int Y \mathbb{1}_{(0,T_n]} dX^{T_n}$ where

$$T_n = \inf\left\{t \ge 0 : |X_t| > n \text{ or } \int_0^t Y_s^2 d\langle X\rangle_s > n\right\}.$$

Remark. Note that this definition of the stochastic integral actually does not depend on the particular localising sequence of stopping times. For instance, let $(U_n)_n$ be another increasing sequence of stopping times $U_n \to \infty$ with the properties that $X^{U_n} \in \mathcal{M}^2$ and $Y\mathbb{1}_{(0,U_n]} \in L^2(X^{U_n})$. Then

$$\int Y\mathbb{1}_{(0,U_n]} dX^{U_n} \to \int Y dX.$$

Indeed, note that

$$\left(\int Y\mathbb{1}_{(0,U_n]} dX^{U_n}\right)^{T_m} = \left(\int Y\mathbb{1}_{(0,T_m]} dX^{T_m}\right)^{U_n} = \left(\int Y dX\right)^{T_m \wedge U_n}.$$

Example. Here is the typical example of a local martingale that we shall encounter. Let $a = (a_t)_{t\geq 0}$ be an adapted, continuous process and let $W = (W_t)_{t\geq 0}$ be a Brownian motion. Note that a is predictable and locally bounded, and that W is a martingale. Hence the stochastic integral $M_t = \int_0^t a_s dW_s$ defines a continuous local martingale M. It is a true martingale if

$$\mathbb{E}\int_0^t a_s^2 ds < \infty \text{ for all } t \geq 0.$$

5. Semimartingales

We have now seen three mutually consistent definitions of stochastic integral as we have generalised, step by step, the class of integrands and integrators. We will now give a final one. First a definition:

Definition. A continuous semimartingale X is a process of the form

$$X_t = X_0 + A_t + M_t$$

where A is a continuous adapted process of finite variation and M is a continuous local martingale and $A_0 = M_0 = 0$.

Proposition. *The decomposition of a continuous semimartingale is unique.*

Proof. Suppose

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

Then $A - A' = M' - M$ is a continuous martingale of finite variation. Hence, it is almost surely constant. □

We will define integrals with respect to semimartingales in the obvious way

$$\int Y dX = \int Y dM + \int Y dA$$

where the second integral on the right is a Lebesgue–Stieltjes integral. For completeness, we now recall some facts about such integrals.

5.1. An aside on Lebesgue–Stieltjes integration. Recall, that if g is non-decreasing and right-continuous, there exists a unique measure μ_g such that $\mu_g(a,b] = g(b) - g(a)$. We can then define the Lebesgue–Stieltjes integral via

$$\int_A \phi(s) dg(s) = \int \mathbb{1}_A \phi \ \mu_g$$

for a Borel set A and measurable function ϕ such that the function $\mathbb{1}_A \phi$ is μ_g-integrable.

PROPOSITION. *Suppose ϕ is locally-μ_g integrable, and let*

$$\Phi(t) = \int_{(0,t]} \phi(s) \ dg(s).$$

Then Φ is right-continuous. If g is continuous, then Φ is also continuous.

PROOF. Write

$$\Phi(t+h) - \Phi(t) = \int \mathbb{1}_{(t,t+h]} \phi \ d\mu_g$$

and note that $\mathbb{1}_{(t,t+h]}\phi \to 0$ everywhere as $h \downarrow 0$, and is uniformly dominated by $\mathbb{1}_{(t,t+1]}|\phi|$. Right-continuity then follows from the dominated convergence theorem.

Similary, write

$$\Phi(t) - \Phi(t-h) = \int \mathbb{1}_{(t-h,t]} \phi \ d\mu_g.$$

If g is continuous, then $\mathbb{1}_{(t-h,t]}\phi \to 0$ almost everywhere, since $\mu_g\{t\} = 0$. Apply the dominated convergence theorem again for left-continuity. $\square$

When g is continuouss, there is no confusion then using the notation

$$\int_{(0,t]} \phi(s) dg(s) = \int_0^t \phi(s) \ dg(s).$$

Of course, we have already been using this type of integral to give sense to expressions such as $\int Y^2 d\langle X \rangle$.

Now, if f is of finite variation, we can write $f = g - h$ where g and h are non-decreasing. If f is continuous, more is true:

THEOREM. *Suppose f is continuous and of finite variation. Then $f = g - h$ where g and h are non-decreasing and continuous.*

PROOF. From last time, we need only show that $t \mapsto \|f\|_{t,\mathrm{var}}$ is continuous.

First we begin with some vocabulary and two easy observations. A partition of the interval $[a,b]$ is a finite set $P = \{p_1, \ldots, p_N\}$ such that $a \leq p_1 < \cdots < p_N \leq b$. First, observe that if P and P' are partitions of $[0,T]$ such that $P \subseteq P'$, then the inequality

$$\sum_P |f(p_k) - f(p_{k-1})| \leq \sum_{P'} |f(p_k) - f(p_{k-1})|$$

holds by the triangle inequality. Second, observe that if $P[0,s]$ is a partition of $[0,s]$ and $P[s,T]$ is a partition of $[s,T]$ we have

$$\|f\|_{T,\mathrm{var}} \geq \sum_{P[0,s]} |f(p_k) - f(p_{k-1})| + \sum_{P[s,T]} |f(p_k) - f(p_{k-1})|$$

by the definition of the left-hand side. By taking the supremum over all such partitions $P[0,s]$ we have

$$\sum_{P[s,T]} |f(p_k) - f(p_{k-1})| \leq \|f\|_{T,\text{var}} - \|f\|_{t,\text{var}}$$

for any partition $P[s,T]$ of $[s,T]$.

Now, fix a $T > 0$ and an $\epsilon > 0$. Pick any $0 < t < T$. By continuity, there exists a $\delta > 0$ such that $|f(s) - f(t)| < \epsilon$ for $t - \delta \leq s \leq t + \delta$. Now, by the definition of the variation norm, there exists a partition $0 \leq p_1 < \cdots < p_N \leq T$ such that

$$(*) \qquad \|f\|_{T,\text{var}} \leq \epsilon + \sum_{k=1}^{N} |f(p_k) - f(p_{k-1})|$$

By the first observation above, we can find (perhaps by refining the partition) a number M such that $p_{M-1} \geq t - \delta$, $p_M = t$ and $p_{M+1} \leq t + \delta$. Now expanding the inequality (*) yields

$$\begin{aligned} \|f\|_{T,\text{var}} \leq \epsilon + & \sum_{P[0,t-\delta]} |f(p_k) - f(p_{k-1})| + |f(t) - f(p_{M-1})| \\ & + |f(p_M) - f(t)| + \sum_{P[t+\delta,T]} |f(p_k) - f(p_{k-1})| \\ & \leq 3\epsilon + \|f\|_{t-\delta,\text{var}} + \|f\|_{T,\text{var}} - \|f\|_{t+\delta,\text{var}} \end{aligned}$$

where $P[0,t-\delta] = \{p_1, \ldots, p_{M-1}\}$ and $P[t+\delta,T] = \{p_{M+1}, \ldots, p_N\}$ and we have used the second observation. Hence, by the monotonicity of $t \mapsto \|f\|_{t,\text{var}}$ we have

$$\|f\|_{t+\delta,\text{var}} - 3\epsilon \leq \|f\|_{t-\delta,\text{var}} \leq \|f\|_{t,\text{var}} \leq \|f\|_{t+\delta,\text{var}} \leq \|f\|_{t-\delta,\text{var}} + 3\epsilon.$$

Taking $\epsilon \downarrow 0$ shows that $t \mapsto \|f\|_{t,\text{var}}$ is continuous. □

Remark. If f is only assumed right-continuous, the above proof can be modified to show that g is right-continuous also. Similarly, if f is only left-continuous, we can show g is left-continuous by a similar argument.

Now if f is continuous and of finite variation, we can define the Lebesgue–Stieltjes integral as

$$\int_0^t \phi(s) df(s) = \int_0^t \phi(s) dg(s) - \int_0^t \phi(s) dh(s)$$

where g, h are non-decreasing continuous functions such that $f = g - h$. Note that this integral does not depend on decomposition. Indeed, if $f = g - h = g' - h'$ then $g + h' = g' + h$ and hence

$$\begin{aligned} \int_0^t \phi(s) d[g(s) + h'(s)] = \int_0^t \phi(s) d[g'(s) + h(s)] \Rightarrow \\ \int_0^t \phi(s) dg(s) - \int_0^t \phi(s) dh(s) = \int_0^t \phi(s) dg'(s) - \int_0^t \phi(s) dh'(s) \end{aligned}$$

The integral is of finite variation since it can be written as the difference of monotone functions:

$$\int_0^t \phi(s) df(s) = \int_0^t \phi(s)^+ dg(s) + \int_0^t \phi(s)^- dh(s) - \int_0^t \phi(s) - dg(s) - \int_0^t \phi(s)^+ dh(s).$$

Also note if f is finite variation and $g(t) = \|f\|_{t,\mathrm{var}}$ then for $0 \le s \le t$ we have

$$|f(s) - f(t)| \le g(t) - g(s)$$

and hence for $h = g - f$,

$$|h(s) - h(t)| \le 2[g(t) - g(s)].$$

In particular, the measure μ_h associated with the non-decreasing process h is such that $\mu_h(A) \le 2\mu_g(A)$ for all A, and hence we have

$$\left\| \int \phi \, df \right\|_{t,\mathrm{var}} \le 3 \int_0^t |\phi(s)| dg(s).$$

We will use the notation $\int |\phi| \, |df|$ to denote the integral on the right-hand side above.

Definition. Let $X = X_0 + A + M$ be a continuous semimartingale. Let $L_{\mathrm{loc}}(X)$ denote the set of predictable processes such that

$$\int_0^t |Y_s| \, |dA_s| + \int_0^t Y_s^2 d\langle M \rangle_s < \infty \text{ a.s. for all } t \ge 0.$$

We single out a useful subset of $L_{\mathrm{loc}}(X)$:

Definition. A predictable process Y is *locally bounded* iff there exists a non-decreasing sequence of stopping times $T_N \uparrow \infty$ and constants $C_N > 0$ such that

$$|Y_t(\omega)| \mathbb{1}_{\{t \le T_N(\omega)\}} \le C_N \text{ for all } (t, \omega)$$

for each N.

Definition. Let $X = X_0 + A + M$ be a continuous semimartingale and $Y \in L_{\mathrm{loc}}(X)$. The integral $\int Y \, dX$ is defined as the continuous semimartingale with decomposition

$$\int Y \, dX = \int Y \, dA + \int Y \, dM.$$

The first integral on the right is a path-by-path Lebesgue–Stieltjes integral so is continuous and of finite variation, while the second integral is a continuous local martingale defined via Itô's isometry and localisation.

6. Summary of properties

Continuing the story, we need to consider the quadratic variation of a semimartingale:

Definition. Let X be a continuous semimartingale such that $X = X_0 + M + A$ where M is a continuous local martingale and A is a continuous adapted process of finite variation. Then

$$\langle X \rangle = \langle M \rangle.$$

This definition is consistent with our existing notion of quadratic variation by the following proposition:

Proposition. *Let X be a continuous semimartingale, and let*

$$\langle X\rangle_t^n = \sum_{k\geq 1}(X_{t\wedge t_k^n} - X_{t\wedge t_{k-1}^n})^2$$

where $t_k^n = k2^{-n}$. Then

$$\langle X\rangle^n \to \langle X\rangle \quad u.c.p.$$

as $n\to\infty$.

Proof. We already know that $\langle M\rangle^n \to \langle M\rangle = \langle X\rangle$ u.c.p. By expanding the squares, we have for any $t\leq T$,

$$\begin{aligned}
|\langle X\rangle_t^n - \langle M\rangle_t^n| &= \left|\sum_{k\geq 1}(A_{t\wedge t_k^n} - A_{t\wedge t_{k-1}^n})[2(M_{t\wedge t_k^n} - M_{t\wedge t_{k-1}^n}) + (A_{t\wedge t_k^n} - A_{t\wedge t_{k-1}^n})]\right| \\
&\leq \|A\|_{T,\mathrm{var}} \sup_{\substack{u,v\in[0,T]\\ |u-v|\leq 2^{-n}}} |2(M_u - M_v) + A_u - A_v| \\
&\to 0 \text{ a.s.}
\end{aligned}$$

by the uniform continuity of $2M+A$ on the compact $[0,T]$. □

A related notion we will also find useful:

Definition. Let X and Y be continuous semimartingales. Then the quadratic covariation is defined by

$$\langle X, Y\rangle = \frac{1}{4}\big(\langle X+Y\rangle - \langle X-Y\rangle\big).$$

(This is called the polarisation identity.)

Remark. The quadratic covariation notation resembles an inner product. Note however, that $\langle X,Y\rangle$ is not a number but a stochastic process, i.e a map $(t,\omega)\mapsto \langle X,Y\rangle_t(\omega)$.

We now collect together a list of useful facts about quadratic covariation.

Theorem. *Let X and Y be continuous semimartingales.*

(1) $\sum_{k\geq 1}(X_{t\wedge t_k^n} - X_{t\wedge t_{k-1}^n})(Y_{t\wedge t_k^n} - Y_{t\wedge t_{k-1}^n}) \to \langle X,Y\rangle_t$ *u.c.p.*

(2) *(Kunita–Watanabe inequality). If X and Y are continuous semimartingales and $H\in L_{\mathrm{loc}}(X)$ and $K\in L_{\mathrm{loc}}(Y)$, then*

$$\int_0^t H_s K_s d\langle X,Y\rangle_s \leq \left(\int_0^t H_s^2 d\langle X\rangle_s\right)^{1/2}\left(\int_0^t K_s^2 d\langle Y\rangle_s\right)^{1/2}$$

(3) *If A is of finite variation, then $\langle A, X\rangle = 0$.*

(4) *The map that sends the pair X,Y to $\langle X,Y\rangle$ is symmetric and bilinear.*

(5) *If M and N are in $\mathcal{M}^2$, then $MN - \langle M,N\rangle$ is a uniformly integrable martingale.*

(6) *(Characterisation) Suppose M and N are continuous local martingales. Then $\langle M,N\rangle$ is the unique continuous adapted process of finite variation A with $A_0 = 0$ such that $MN - A$ is a local martingale.*

(7) *If X and Y are independent, then $\langle X,Y\rangle = 0$.*

(8) *If $H\in L_{\mathrm{loc}}(X)$ then $\left\langle \int H dX\right\rangle = \int H^2 d\langle X\rangle$.*

(9) *(Kunita–Watanabe identity) If $H \in L_{\text{loc}}(X)$ then*
$$\left\langle Y, \int H dX \right\rangle = \int H d\langle X, Y \rangle.$$

Proof. (1) This is just a consequence of polarisation identity and the definition of quadratic variation.
(2) This proof was not done in the lectures. We start with a result from real analysis. Let f, ϕ and ψ be continuous, where ϕ and ψ are increasing and f is a finite variation function and such that

$$2ab|f(t) - f(s)| \le a^2[\phi(t) - \phi(s)] + b^2[\psi(t) - \psi(s)] \tag{*}$$

for all $0 \le s \le t$ and all real a, b. Then

$$\int_0^t \alpha(s)\beta(s) df(s) \le \left(\int_0^t \alpha(s)^2 d\phi(s)\right)^{1/2} \left(\int_0^t \beta(s)^2 d\psi(s)\right)^{1/2} \tag{**}$$

for all $t \ge 0$ and any $\alpha \in L^2(d\phi)$ and $\beta \in L^2(d\psi)$.

To see this, it enough to suppose f is increasing. Fix a, b and let
$$F(t) = a^2\phi(t) + b^2\psi(t) - 2abf(t)$$
Note that by equation (*) the function F is increasing, and hence corresponds to a Lebesgue–Stieltjes measure μ_F. In particular, since $\mu_F(A) \ge 0$ we have
$$2ab\mu_f(A) \le a^2\mu_\phi(A) + b^2\mu_\psi(A)$$
for any Borel A. If α and β are simple functions, then the above equation implies

$$2\int \alpha\beta df \le \int \alpha^2 d\phi + \int \beta^2 d\psi. \tag{***}$$

The validity of the above inequality for non-negative α, β follows from the monotone convergence theorem, and for general α, β by the inequality $\alpha\beta \le |\alpha|\ |\beta|$.

Finally, in equation (***), replace α with $c\alpha$ and replace β with β/c for a positive real c. Minimising the resulting expression over $c > 0$ yields inequality (**).

The Kunita–Watanabe inequality follows from the fact there is an almost sure set Ω_0 such that equation (*) holds where $f = \langle X, Y\rangle(\omega)$, $\phi = \langle X\rangle(\omega)$ and $\psi = \langle \psi\rangle(\omega)$. Indeed, by the u.c.p. convergence of $\langle X\rangle^{(n)}, \langle Y\rangle^{(n)}$ and $\langle X, Y\rangle^{(n)}$ there is an almost sure set Ω_0 and a subsequence $(n_k)_k$ such that
$$\langle X\rangle_t^{(n_k)}(\omega) \to \langle X\rangle_t(\omega)$$
simultaneously for all $t \ge 0$ with similar statements applying to $\langle Y\rangle$ and $\langle X, Y\rangle$. We are done since it is easy to see that
$$2ab\langle X, Y\rangle_t^{(n)}(\omega) \le a^2\langle X\rangle_t^{(n)} + b^2\langle X\rangle_t^{(n)}(\omega).$$
(3) If $X = X_0 + A' + M$ then $\langle X + A\rangle = \langle X - A\rangle = \langle M\rangle$, and so this follows from the polarisation identity.
(4) The symmetry and bilinearity are obvious from (1).
(5) Since $\mathcal{M}^2$ is a vector space, if $M, N \in \mathcal{M}^2$ then both $M + N$ and $M - N$ are in $\mathcal{M}^2$. By an earlier result on quadratic variation of square integrable martingales, both

$(M+N)^2 - \langle M+N\rangle$ and $(M-N)^2 - \langle M-N\rangle$ are uniformly integrable martingales. Hence, so is

$$MN - \langle M, N\rangle = \frac{1}{4}\big[(M+N)^2 - (M-N)^2\big] - \frac{1}{4}\big[\langle M+N\rangle - \langle M-N\rangle\big].$$

(6) This follows from localising (5) and the uniqueness of the semimartingale decomposition.
(7) We may assume X and Y are local martingales by (3). First suppose X and Y are square-integrable martingales. Let $\mathbb{F}^{X,Y}$ be the filtration generated by X and Y. Since $\mathcal{F}_t^{X,Y} \subseteq \mathcal{F}_t$ we have

$$\mathbb{E}(X_\infty|\mathcal{F}_t^{X,Y}) = \mathbb{E}[\mathbb{E}(X_\infty|\mathcal{F}_t)|\mathcal{F}_t^{X,Y}] = X_t$$

and hence X is a $\mathbb{F}^{X,Y}$ martingale. Similarly, so is Y.

By the assumed independence of the processes X and Y, the random variables X_∞ and Y_∞ are conditionally independent given $\mathcal{F}_t^{X,Y}$. Indeed, recall that $\mathcal{F}_t^{X,Y}$ is generated by events of the form $\{X_u \in A\} \cap \{Y_v \in B\}$ where A, B are Borel and $0 \le u, v \le t$. Now the product XY is also an $\mathbb{F}^{X,Y}$ martingale since

$$\mathbb{E}(X_\infty Y_\infty|\mathcal{F}_t^{X,Y}) = \mathbb{E}(X_\infty|\mathcal{F}_t^{X,Y})\mathbb{E}(Y_\infty|\mathcal{F}_t^{X,Y}) = X_t Y_t.$$

As we have proved in the first chapter, the martingale property is not affected by replacing a given filtration $\mathbb{F}$ by the smallest completed right-continuous filtration $\mathbb{F}^*$ containing $\mathbb{F}$, we now know that XY is a martingale with respect to a filtration satisfying the usual conditions, so the previously proven results of stochastic calculus are applicable. In particular, by (5), the quadratic variation is $\langle X, Y\rangle = 0$, at least when computed with respect to this filtration. But the limit in (1) makes no reference to the filtration, so the $\langle X, Y\rangle = 0$.

For general local martingales X and Y with with $X_0 = Y_0 = 0$, let

$$S_k = \inf\{t \ge 0 : |X_t| > k\} \text{ and } T_k = \inf\{t \ge 0 : |Y_t| > k\}$$

so that X^{S_k} and Y^{T_k} are independent bounded martingales. In particular, the product $X^{S_k}Y^{T_k}$ is a martingale. Therefore, the stopped process $(X^{S_k}Y^{T_k})^{S_k\wedge T_k} = (XY)^{S_k \wedge T_k}$ is a martingale, and hence XY is a local martingale.
(8) It is enough to consider the case where X is a local martingale. And by localisation, we can assume $X \in \mathcal{M}^2$ and $H \in L^2(X)$. We must show that $M = (\int H dX)^2 - \int H^2 d\langle X\rangle$ is a martingale. We already know that this is the case when H is simple. So let $(H_n)_n \in \mathcal{S}$ be such that $H_n \to H$ in $L^2(X)$. Since $\int H_n dX \to \int H dX$ in $\mathcal{M}^2$ we have

$$\left(\int_0^\infty H_n dX\right)^2 \to \left(\int_0^\infty H dX\right)^2 \text{ in } L^1(\Omega)$$

Similarly,

$$\int_0^\infty H_n^2 d\langle X\rangle \to \int_0^\infty H_n^2 d\langle X\rangle \text{ in } L^1(\Omega).$$

The remainder of the proof is on the example sheet.
(9) Again, we need only consider the case where X and Y are local martingales. By (5), we must show that $Z = Y\int H dX - \int H d\langle X, Y\rangle$ is a local martingale. By localisation, we can assume X and Y are square-integrable martingales and $H \in L^2(X)$, in which case we must show that Z is a true martingale. When $H = K\mathbb{1}_{(s_0,s_1]}$ for a bounded $\mathcal{F}_{s_0}$-measurable K we have

$$Z_t = KY_t(X_{s_1\wedge t} - X_{s_0 \wedge t}) - K(\langle X, Y\rangle_{s_1\wedge t} - \langle X, Y\rangle_{s_0\wedge t})$$

which is martingale by (5) and a routine calculation with conditional expectations. By linearity Z is a martingale whenever H is simple.

Finally, let $(H_n)_n \in \mathcal{S}$ be such that $H_n \to H$ in $L^2(X)$. Since $\int H_n dX$ in $\mathcal{M}^2$ we have

$$\mathbb{E}\left|Y_\infty \int_0^\infty (H - H_n) dX_s\right| \le \mathbb{E}(Y_\infty^2)^{1/2} \mathbb{E}\left(\int_0^\infty (H - H_n)^2 d\langle X\rangle_s\right)^{1/2} \to 0$$

and

$$\mathbb{E}\left|\int_0^\infty (H - H^n) d\langle X, Y\rangle_s\right| \le \mathbb{E}\left(\int_0^\infty (H - H_n)^2 d\langle X\rangle_s\right)^{1/2} \mathbb{E}(\langle Y\rangle_\infty)^{1/2} \to 0$$

by the Kunita–Watanabe inequality and the Cauchy–Schwarz inequality. The proof concludes as in (8). $\square$

Here are some more properties:

Proposition. *Let X be a continuous semimartingale and $Y \in L_{\mathrm{loc}}(X)$. Let K be a $\mathcal{F}_s$-measurable random variable. Then $KY\mathbb{1}_{(s,\infty)} \in L_{\mathrm{loc}}(X)$ and*

$$\int KY\mathbb{1}_{(s,\infty)} dX = K \int Y\mathbb{1}_{(s,\infty)} dX.$$

Theorem. *Let X be a continuous semimartingale and let $B \in L_{\mathrm{loc}}(X)$ and $A \in L_{\mathrm{loc}}(\int B\, dX)$. Then $AB \in L_{\mathrm{loc}}(X)$ and*

$$\int A\ d\left(\int B\ dX\right) = \int AB\ dX.$$

Proof. On the example sheet you are asked to prove this claim when the integrator is of finite variation. Assuming that the chain rule formula holds in this case, we have

$$\int A^2 d\left\langle \int BdX \right\rangle = \int A^2 d \int B^2 d\langle X\rangle = \int A^2 B^2 d\langle X\rangle$$

and hence $AB \in L_{\mathrm{loc}}(X)$.

We only consider the case where X is a local martingale. The claim is true if $A \in \mathcal{S}$ is simple and predictable just by comparing the sums defining both integrals and using the above proposition. By localisation we assume that $X \in \mathcal{M}^2$ and let $A^n \in \mathcal{S}$ such that $A^n \to A$ in $L^2(\int BdX)$. But

$$\begin{aligned}
\mathbb{E}\left[\left(\int_0^\infty (A_t - A_t^n) d\int_0^t B_s dX_s\right)^2\right] &= \mathbb{E}\left[\int_0^\infty (A_t - A_t^n)^2 d\left\langle \int tBdX \right\rangle_t\right] \\
&= \mathbb{E}\left[\int_0^\infty (A_t - A_t^n)^2 B_t^2 d\langle X\rangle_t\right] \\
&= \mathbb{E}\left[\left(\int_0^\infty (A_t - A_t^n) B_t dX_t\right)^2\right]
\end{aligned}$$

$\square$

Remark. We will adopt a more compact differential notation. For instance, if

$$Y_t = Y_0 + \int_0^t B_s dX_s$$

we will write

$$dY = B\ dX$$

which should only be considered short-hand for the corresponding integral equation. Indeed, recall that our chief example of a continuous martingale is Brownian motion, and the sample paths of Brownian motion are *nowhere* differential. Hence the differential notation can only be considered formally. Nevertheless, the notation is useful since the above chain rule can be expressed as

$$A\ dY = AB\ dX.$$

7. Itô's formula

We are now ready for Itô's formula, probably the single most useful result of stochastic calculus. It says that a smooth function of n semimartingales is again a semimartingale with an explicit decomposition.

THEOREM (Itô's formula). *Let $X = (X^1, \ldots, X^n)$ be an n-dimensional continuous semimartigale, let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable. Then*

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

REMARK. Notice that all of the integrands continuous and adapted and hence locally bounded and predictable, and in particular, all integrands are well-defined.

REMARK. Itô's formula is equivalently expressed as

$$\boxed{df(X) = \sum_{i=1}^n \frac{\partial f}{\partial x^i} dX^i + \frac{1}{2}\sum_{i,j=1}^n \frac{\partial^2 f}{\partial x^i \partial x^j} d\langle X^i, X^j\rangle}$$

If X is of finite variation, then $\langle X^i, X^j\rangle = 0$, and Itô's formula reduces to chain rule of ordinary calculus. The extra quadratic co-variation term that appears in the general case is sometimes called Itô's correction term.

EXAMPLE. Let X be a scalar continuous semimartingale, and let $Y = e^X$. Then

$$dY = Y dX + \frac{1}{2} Y d\langle X\rangle.$$

We will use this fact many times.

REMARK. Before giving a detailed proof, let us first consider the underlying reason why Itô's formula works. Simply put, it is Taylor's formula expanded to the second order.

Indeed, expand $f(X_t)$ about the point X_s, where $0 \le s \le t$ so that

$$f(X_t) = f(X_s) + \sum_{i=1}^n \frac{\partial f}{\partial x^i}(X_s)(X^i_t - X^i_s) + \frac{1}{2}\sum_{i,j=1}^n \frac{\partial^2 f}{\partial x^i \partial x^j}(X_s)(X^i_t - X^i_s)(X^j_t - X^j_s) + \ldots$$

by Taylor's theorem. In ordinary calculus, the second order terms are much smaller than the first order terms, so that in the limit $s \to t$, they can be ignored. However, in stochastic calculus, non-trivial local martingales have positive quadratic variation, and so the second order terms usually contribute a non-zero term to the limit. Hence, we are left with the Itô correction terms involving second derivatives and quadratic covariations. Nevertheless, the third and higher order terms are still relatively small and can be ignored.

To turn this argument into a proof, we would need to control the third and higher order terms that we have discarded. Instead, we present a simpler proof that uses the density of polynomials in the set of continuous functions.

We will prove Itô's formula via a series of lemmas.

Lemma. *Let X be a continuous semimartingale, and Y locally bounded, adapted and left-continuous. Then*

$$\sum_{k\geq 1} Y_{t^n_{k-1}}(X_{t\wedge t^n_k} - X_{t\wedge t^n_{k-1}}) \to \int_0^t Y_s dX_s \ u.c.p$$

where $t^n_k = k2^{-n}$.

Remark. This lemma differs from similar results proven so far because, rather than asserting by density the existence of an approximating sequence $(Y^n)_n$ of simple predictable process, we are given one explicitly. This is where we will use the left-continuity of Y.

Proof of the lemma. If $X = X_0 + A + M$ where A is of finite variation and M is a local martingale, it is enough to prove

$$\sum_{k\geq 1} Y_{t^n_k}(A_{t\wedge t^n_k} - A_{t\wedge t^n_{k-1}}) \to \int_0^t Y_s dA_s \text{ u.c.p}$$

and

$$\sum_{k\geq 1} Y_{t^n_k}(M_{t\wedge t^n_k} - M_{t\wedge t^n_{k-1}}) \to \int_0^t Y_s dM_s \text{ u.c.p}$$

First we consider the convergence of the 'dM' integral. As usual, we can assume Y is uniformly bounded and that M is square integrable. Now note that the predictable process

$$Y^n = \sum_{k\geq 1} Y_{t^n_k} \mathbb{1}_{(t^n_{k-1}, t^n_k]}$$

is bounded and converges to Y pointwise by left-continuity. Hence

$$\mathbb{E}\int_0^\infty (Y^n_s - Y_s)^2 d\langle M\rangle_s \to 0$$

by the dominated convergence theorem and hence $\int Y^n dM \to \int Y dM$ in $\mathcal{M}^2$. The u.c.p. convergence in the general case follows from a composition of Chebychev's inequalty, Doob's inequality and localisation.

Now the 'dA' integral. Assuming A is of bounded variation and that Y is bounded then

$$\sup_{t\geq 0}\left|\int_0^t (Y^n_s - Y_s) dA_s\right| \leq 3\int_0^\infty |Y^n_s - Y_s||dA_s| \to 0$$

by the dominated convergence theorem. The general case follows from localisation.

□

Now we come to the stochastic integration by parts formula. It says that the product of continuous semimartingales is again a semimartingale with an explicit decomposition.

Lemma (Stochastic integration by parts or product formula). *Let X and Y be continuous semimartingales. Then*

$$X_tY_t = X_0Y_0 + \int_0^t X_s dY_s + \int_0^t Y_s dX_s + \langle X, Y\rangle_t.$$

Proof of the integration by parts formula. First we consider the case $X = Y$. Then by the approximation lemma

$$\begin{aligned}
2\int_0^t X_s dX_s &= \lim_n \sum_{k\geq 1} 2X_{t^n_{k-1}}(X_{t\wedge t^n_k} - X_{t\wedge t^n_{k-1}}) \\
&= \lim_n \sum_{k\geq 1} X^2_{t\wedge t^n_k} - X^2_{t\wedge t^n_{k-1}} - (X_{t\wedge t^n_k} - X_{t\wedge t^n_{k-1}})^2 \\
&= X_t^2 - X_0^2 - \langle X\rangle_t
\end{aligned}$$

where $t^n_k = k2^{-n}$.

Now we apply the polarisation identity:

$$\begin{aligned}
2\int_0^t (X_s + Y_s) d(X_s + Y_s) = (X_t + Y_t)^2 - (X_0 + Y_0)^2 - \langle X + Y\rangle_t \\
2\int_0^t (X_s - Y_s) d(X_s - Y_s) = (X_t - Y_t)^2 - (X_0 - Y_0)^2 - \langle X - Y\rangle_t
\end{aligned}$$

The result follows from subtracting the above equations and dividing by 4. □

Proof of Itô's formula. We do the $n = 1$ case. The other cases are similar. First we prove Itô's formula for monomials. We proceed by induction. Itô's formula holds for $X^0 = 1$. So, suppose

$$d(X^m) = mX^{n-1}dX + \frac{m(m-1)}{2}X^{m-2}d\langle X\rangle$$

Now note that by bilinearity and the Kunita–Watanabe identity

$$\begin{aligned}
d\langle X^m, X\rangle &= mX^{m-1}d\langle X\rangle + \frac{m(m-1)}{2}X^{m-2}d\langle\langle X\rangle, X\rangle \\
&= mX^{m-1}d\langle X\rangle
\end{aligned}$$

where we have used to fact that $\langle X\rangle$ is of finite variation to assert $\langle\langle X\rangle, X\rangle = 0$. Now by the the integration by parts formula

$$\begin{aligned}
d(X^{m+1}) =& d(X^m X) \\
=& X^m dX + X d(X^m) + d\langle X^m, X\rangle \\
=& X^m dX + \left(mX^m dX + \frac{m(m-1)}{2}X^m d\langle X\rangle\right) + mX^{m-1}d\langle X\rangle \\
=& (m+1)X^n dX + \frac{(m+1)n}{2}X^{n-1}d\langle X\rangle,
\end{aligned}$$

completing the induction. Note that we have used the stochastic chain rule (lecture 11) to go from the second to third line. By linearity, we have also proven Itô's formula for polynomials.

Now, for general $f \in C^2$, let us suppose $X = X_0 + M + A$ where X_0 is a bounded $\mathcal{F}_0$-measurable random variable, M is a bounded martingale and A is of bounded variation. Therefore, there is a constant $N > 0$ such that $|X_t(\omega)| \leq N$ for all (t, ω). Now by the Weierstrass approximation theorem, given an $n > 0$, there exists a polynomial p_n such that the C^2 function $h_n = f - p_n$ satisfies the bound

$$|h_n(x)| + |h_n'(x)| + |h_n''(x)| \leq 1/n \text{ for all } x \in [-N, N]$$

where the h' denotes the derivative of h, etc. Since Itô's formula holds for polynomials, we have

$$\begin{aligned} f(X_t) - f(X_0) - & \int_0^t f'(X_s) dX_s - \frac{1}{2} \int_0^t f''(X_s) d\langle X \rangle_s \\ & = h_n(X_t) - h_n(X_0) - \int_0^t h_n'(X_s) dX_s - \frac{1}{2} \int_0^t h_n''(X_s) d\langle X \rangle_s \end{aligned}$$

But by a now familiar argument, the terms on the right-side converge u.c.p. to 0 as $n \to \infty$.

Now, for general $X = X_0 + M + A$, let

$$T_N = \inf\{t \geq 0 : |X_t| > N, \|A\|_{t,\mathrm{var}} > N\}.$$

We have shown

$$\begin{aligned} f(X_t^{T_N}) &= f(X_0) + \int_0^t f'(X_s^{T_N}) dX_s^{T_N} - \frac{1}{2} \int_0^t f''(X_s^{T_N}) d\langle X^{T_N} \rangle_s \\ &= f(X_0) + \int_0^{t \wedge T_N} f'(X_s) dX_s + \frac{1}{2} \int_0^{t \wedge T_N} f''(X_s) d\langle X \rangle_s \end{aligned}$$

(Note that the above formula holds even on the event $\{|X_0| > N\}$, since in this case we have $T_N = 0$ and both integrals vanish.) Sending $N \to \infty$ completes the proof. $\square$

CHAPTER 4

Applications to Brownian motion

1. Lévy's characterisation of Brownian motion

We now see a striking application of Itô's formula.

THEOREM (Lévy's characterisation of Brownian motion). *Let X be a continuous d-dimensional local martingale such that $X_0 = 0$ and*

$$\langle X^j, X^j \rangle_t = \begin{cases} t & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

Then X is a standard d-dimensional Brownian motion.

PROOF. Fix a constant vector $\theta \in \mathbb{R}^d$ and let

$$M_t = e^{\mathrm{i}\theta \cdot X_t + \|\theta\|^2 t/2}.$$

By Itô's formula,

$$\begin{aligned} dM_t &= M_t \left(\mathrm{i}\ \theta \cdot dX_t + \frac{\|\theta\|^2}{2} dt \right) - \frac{1}{2} M_t \sum_{i,j=1}^{d} \theta^i \theta^j d\langle X^i, X^j \rangle_t \\ &= \mathrm{i}\ M_t \theta \cdot dX_t \end{aligned}$$

and so M is a continuous local martingale, as it is the stochastic integral with respect to a continuous local martingale. On the other hand, since $|M_t| = e^{\|\theta\|^2 t/2}$ and hence $\mathbb{E}(\sup_{s \in [0,t]} |M_s|) < \infty$ the local martingale M is in fact a true martingale. Thus for all $0 \leq s \leq t$ we have

$$\mathbb{E}(M_t | \mathcal{F}_s) = M_s$$

which implies

$$\mathbb{E}(e^{\mathrm{i}\ \theta \cdot (X_t - X_s)} | \mathcal{F}_s) = e^{-\|\theta\|^2 (t-s)/2}.$$

That X is a Brownian motion in the filtration is a consequence of a result proven in a previous chapter. $\square$

The next result follows directly from Lévy's characterisation theorem.

THEOREM (Dambis, Dubins–Schwarz 1965). *Let X be a scalar continuous local martingale for a filtration $\mathbb{F}$, such that $X_0 = 0$ and $\langle X \rangle_\infty = \infty$ a.s. Define a family of stopping times by*

$$T(s) = \inf\{t \geq 0 : \langle X \rangle_t > s\},$$

and a family of random variables by

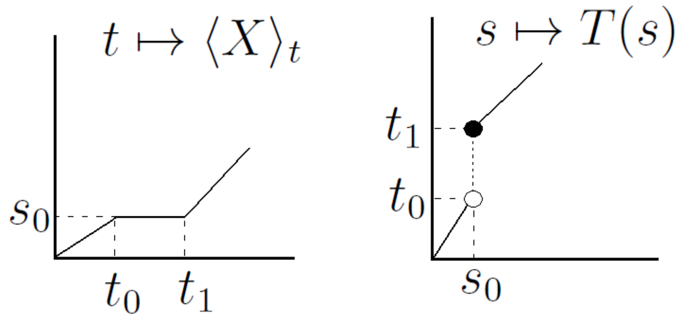
$$W_s = X_{T(s)}$$

and a family of sigma-fields by

$$\mathcal{G}_s = \mathcal{F}_{T(s)}.$$

Then $\mathbb{G} = (\mathcal{G}_s)_{s\geq 0}$ is a filtration and the process W is a Brownian motion in $\mathbb{G}$.

Proof. For a fixed outcome $\omega \in \Omega$, the map $t \to \langle X\rangle_t(\omega)$ is continuous and non-decreasing. Hence, the map $s \to T(s,\omega)$ is increasing and right-continuous. The relationship is illustrated below. Since $T(s) \leq T(s')$ a.s. whenever $0 \leq s \leq s'$, we have the inclusion



$\mathcal{F}_{T(s)} \subseteq \mathcal{F}_{T(s')}$ and so $\mathbb{G}$ is a filtration.

Now we will show that W is continuous. To see what we need to prove, consider the figure. For the given realisation, we have

$$\langle X\rangle_{t_1} = \langle X\rangle_{t_0}$$

so that $W_{s_0} = X_{t_1}$ while $\lim_{s\uparrow s_0} W_s = X_{t_0}$. To show continuity, we must show $X_{t_1} = X_{t_0}$. That is to say, that X and $\langle X\rangle$ have the same intervals of constancy.

So fix $t_0 \geq 0$ and let

$$T = \inf\{u \geq t_0 : \langle X\rangle_u > \langle X\rangle_{t_0}\}$$

so that $T(\omega) = t_1$ in the figure. Define a process $Y = (Y_u)_{u\geq 0}$ by

$$\begin{aligned} Y_u &= X_{u\wedge T\vee t_0} - X_{u\vee t_0} \\ &= \int_0^u \mathbb{1}_{(t_0,T]}(r)dX_r \end{aligned}$$

Note that Y is a continuous local martingale and that

$$\begin{aligned} \langle Y\rangle_\infty &= \int_0^\infty \mathbb{1}_{(t_0,T]}(r)d\langle X\rangle_r \\ &= \langle X\rangle_T - \langle X\rangle_{t_0} = 0 \end{aligned}$$

by the definition of T and the continuity of $\langle X\rangle$. Hence, Y is almost surely constant and $X_T = X_{t_0}$. This does the job for a fixed t_0.

Now let

$$S_r = \inf\{t \geq r : \langle X\rangle_t > \langle X\rangle_r\}$$

and

$$T_r = \inf\{t \geq r : X_t \neq X_r\}.$$

We have show that $T_r = S_r$ a.s. for each r. Hence $T_r = S_r$ for all rational r a.s. But since S and T are right-continuous, we have $S_r = T_r$ for all r a.s. So X and $\langle X \rangle$ have the same intervals of constancy afterall. In particular, we can conclude that W is continuous.

Now we show that W is a local martingale in $\mathbb{G}$. Let

$$\tau_N = \inf\{t \geq 0 : |X_t| > N\}$$

so that X^{τ_N} is a bounded $\mathbb{F}$-martingale for each N. Let $\sigma_N = \langle X \rangle_{\tau_N}$ so that

$$\begin{aligned}
\mathbb{E}(W_{\sigma_N \wedge s_1} | \mathcal{G}_{s_0}) &= \mathbb{E}(X_{\tau_N \wedge T(s_1)} | \mathcal{F}_{T(s_0)}) \\
&= X_{\tau_N \wedge T(s_0)} \\
&= W_{\sigma_N \wedge s_0}
\end{aligned}$$

for any $0 \leq s_0 \leq s_1$, by the optional sampling theorem. Hence W^{σ_N} is a $\mathbb{G}$-martingale. Note that $\sigma_N \to \infty$ since $\langle X \rangle_\infty = \infty$, and hence we have shown W is a local martingale.

Finally, note that $\langle W \rangle_s = \langle X \rangle_{T(s)} = s$ for all $s \geq 0$, and hence W is a Brownian motion by Lévy's characterisation theorem. □

We now rewrite the conclusion of the above theorem, and remove the assumption that $\langle X \rangle_\infty = \infty$ a.s.

Theorem. *Let X be a continuous local martingale with $X_0 = 0$. Then X is a time-changed Brownian motion in the sense that there exists a Brownian motion W, possibly defined on an extended probability space, such that $X_t = W_{\langle X \rangle_t}$.*

Proof. It is an exercise to show that $\lim_{t \to \infty} X_t(\omega)$ exists on the set $\{\langle X \rangle_\infty < \infty\}$. Therefore, the process

$$W_s = X_{T(s)} + \int_0^s \mathbb{1}_{\{u > \langle X \rangle_\infty\}} dB_u$$

is well-defined, where B is a Brownian motion independent of X. Note that W is a local martingale and

$$\begin{aligned}
\langle W \rangle_s &= \langle X \rangle_{T(s)} + \int_0^s \mathbb{1}_{\{u > \langle X \rangle_\infty\}} du \\
&= s \wedge \langle X \rangle_\infty + (s - \langle X \rangle_\infty)^+ \\
&= s
\end{aligned}$$

so W is a Brownian motion by Lévy's characterisation. □

1.1. Remark on the conformal invariance of complex Brownian motion. This section is an attempt to illustrate the application of a variant of the Dambis–Dubins–Schwarz theorem to complex Brownian motion.

Let X and Y be independent real Brownian motions and let $W = X + \mathrm{i}Y$ be a complex Brownian motion. Let $f : \mathbb{C} \to \mathbb{C}$ be holomorphic. Our aim is to show that there exists another complex Brownian motion $\hat{W}$ and a non-decreasing process A such that

$$f(W_t) = f(0) + \hat{W}_{A_t}$$

Let u and v be real functions such that

$$f(x + \mathrm{i}y) = u(x, y) + \mathrm{i}v(x, y).$$

Recall that u and v are C^∞, satisfy the Cauchy–Riemann equations

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \text{ and } \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}$$

and are harmonic

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 = \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2}.$$

Hence, by Itô's formula we have

$$\begin{aligned} df(W) &= \left(\frac{\partial u}{\partial x} dX + \frac{\partial u}{\partial y} dY + \frac{1}{2}\frac{\partial^2 u}{\partial x^2} d\langle X \rangle + \frac{\partial^2 u}{\partial x \partial y} d\langle X, Y \rangle + \frac{1}{2}\frac{\partial^2 u}{\partial x^2} d\langle Y \rangle \right) \\ &\quad + \mathrm{i} \left(\frac{\partial v}{\partial x} dX + \frac{\partial v}{\partial y} dY + \frac{1}{2}\frac{\partial^2 v}{\partial x^2} d\langle X \rangle + \frac{\partial^2 v}{\partial x \partial y} d\langle X, Y \rangle + \frac{1}{2}\frac{\partial^2 v}{\partial x^2} d\langle Y \rangle \right) \\ &= \left(\frac{\partial u}{\partial x} dX + \frac{\partial u}{\partial y} dY \right) + \mathrm{i} \left(-\frac{\partial u}{\partial y} dX + \frac{\partial u}{\partial x} dY \right). \end{aligned}$$

If we define real processes U and V by $f(W) = f(0) + U + \mathrm{i}V$, we see that

$$\langle U, V \rangle = 0$$

and

$$\langle U \rangle = \langle V \rangle = A$$

where A is the non-decreasing process given by

$$dA_t = \left[\left(\frac{\partial u}{\partial x} \right)^2 + \left(\frac{\partial u}{\partial y} \right)^2 \right] dt = |f'(W_t)|^2 dt.$$

As before, we have

$$U_t = \hat{X}_{A_t} \text{ and } V_t = \hat{Y}_{A_t}$$

where $\hat{X}$ and $\hat{Y}$ are independent Brownian motions.

An application of this representation is this claim. Let W be a two-dimensional Brownian motion. Then for any $a \in \mathbb{R}^2$ such that $a \neq 0$, we have

$$\mathbb{P}(W_t = a \text{ for any } t \geq 0) = 0.$$

The idea of the proof is to identify $W = (X, Y)$ with the complex Brownian motion $X + \mathrm{i}Y$ and the point $a = (\alpha, \beta)$ with the complex number $\alpha + \mathrm{i}\beta$. Consider the holomorphic function $f(w) = a(1 - e^w)$ so by the above argument, we have $f(W) = \hat{W}_A$ for another complex Brownian motion $\hat{W}$ and a non-decreasing process A. Assuming[1] $A_\infty = \infty$ a.s., we have the following equalities

$$\begin{aligned} \mathbb{P}(W_t = a \text{ for any } t \geq 0) &= \mathbb{P}(\hat{W}_t = a \text{ for any } t \geq 0) \\ &= \mathbb{P}(\hat{W}_{A_t} = a \text{ for any } t \geq 0) \\ &= \mathbb{P}(e^{W_t} = 0 \text{ for any } t \geq 0) \end{aligned}$$

from which the claim follows.

[1] In this case, we see that $dA_t = |f'(W_t)|^2 dt = |a|^2 e^{2X_t} dt$. By the recurrence of scalar Brownian motion we can conclude $A_\infty = \infty$.

2. Changes of measure and the Cameron–Martin–Girsanov theorem

The notion of a continuous semimartingale is rather robust. We have seen that if X is a semimartingale, then so is the stochastic integral $\int Y\, dX$. Also, Itô's formula shows that if f is smooth enough, then $f(X)$ is again a semimartingale. In this section, if we will see that often when we change the underlying probability measure $\mathbb{P}$, the process X remains a semimartingale under the new probability measure $\mathbb{Q}$. To make this all precise we first introduce a definition:

Definition. Let $(\Omega, \mathcal{F})$ be a measurable space and let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures on $(\Omega, \mathcal{F})$. The measures $\mathbb{P}$ and $\mathbb{Q}$ are *equivalent*, written $\mathbb{P} \sim \mathbb{Q}$, iff

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

(or equivalently, iff $\mathbb{P}(A) = 1 \Leftrightarrow \mathbb{Q}(A) = 1.$)

It turns out that equivalent measures can be characterized by the following theorem. When there are more than one probability measure floating around, we use the notation $\mathbb{E}^{\mathbb{P}}$ to denote expected value with respect to $\mathbb{P}$, etc.

Theorem (Radon–Nikodym theorem). *The probability measure $\mathbb{Q}$ is equivalent to the probability measure $\mathbb{P}$ if and only if there exists a random variable Z such that $\mathbb{P}(Z > 0) = 1$ and*

$$\mathbb{Q}(A) = \mathbb{E}^{\mathbb{P}}(Z \mathbb{1}_A)$$

for each $A \in \mathcal{F}$.

Note that $\mathbb{E}^{\mathbb{P}}(Z) = 1$ by putting $A = \Omega$ in the conclusion of theorem. Also, by the usual rules of integration theory, if ξ is a non-negative random variable then

$$\mathbb{E}^{\mathbb{Q}}(\xi) = \mathbb{E}^{\mathbb{P}}(Z\xi).$$

If $\mathbb{Q} \sim \mathbb{P}$, then the random variable Z is called the *density*, or the *Radon–Nikodym derivative*, of $\mathbb{Q}$ with respect to $\mathbb{P}$, and is often denoted

$$Z = \frac{d\mathbb{Q}}{d\mathbb{P}}.$$

In fact, by the definition of equivalence of measure, we also have $\mathbb{Q}(Z > 0) = 1$ and hence $\mathbb{P}$ also has a density with respect to $\mathbb{Q}$ given by

$$\frac{d\mathbb{P}}{d\mathbb{Q}} = \frac{1}{Z}.$$

In particular,

$$\mathbb{P}(A) = \mathbb{E}^{\mathbb{Q}}\left(\frac{1}{Z}\mathbb{1}_A\right)$$

for all $A \in \mathcal{F}$.

Example. To anticipate the Cameron–Martin–Girsanov theorem, we first consider a very simple and explicit example. It turns out that the main features of the theorem are already present in this example, though the method of proof of the general theorem will use the efficient tools of stochastic calculus developed in the previous lectures.

Let X be an n-dimensional random vector with the $N(0, I)$ distribution under a measure $\mathbb{P}$. Fix a vector $a \in \mathbb{R}^n$ and let

$$Z = e^{a \cdot X - \|a\|^2/2}.$$

Note that Z is positive and $\mathbb{E}^{\mathbb{P}}(Z) = 1$. So we can define an equivalent measure $\mathbb{Q}$ by

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = e^{a \cdot X - \|a\|^2/2}.$$

How does X look under this new measure? Well, pick a bounded measurable f and consider the integral

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}}[f(X)] &= \mathbb{E}^{\mathbb{P}}[e^{a \cdot X - \|a\|^2/2} f(X)] \\ &= \int_{\mathbb{R}^n} \frac{1}{(2\pi)^{n/2}} e^{a \cdot x - \|a\|^2/2} f(x) e^{-\|x\|^2/2} dx \\ &= \int_{\mathbb{R}^n} \frac{1}{(2\pi)^{n/2}} f(x) e^{-\|x-a\|^2/2} dx. \end{aligned}$$

We recognise the density of the $N(a, I)$ distribution in the final integral. Hence, we have the conclusion that $X - a \sim N(0, I)$ under $\mathbb{Q}$.

Let X be a continuous martingale under a measure $\mathbb{P}$. We now explore what X looks like under an equivalent measure $\mathbb{Q}$. The measures $\mathbb{Q}$ that we consider will be such that

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

where Z is a positive, continuous, uniformly integrable martingale with $Z_0 = 1$.

Theorem. *For a given probability measure $\mathbb{P}$, suppose Z is a strictly positive $\mathbb{P}$-uniformly integrable martingale. Let X be a continuous $\mathbb{P}$-local martingale. Define an equivalent measure $\mathbb{Q}$ by*

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

Then the process $X - \langle X, \log Z\rangle$ is a $\mathbb{Q}$-local martingale.

Proof. Note by Itô's formula

$$d \log Z = \frac{dZ}{Z} - \frac{d\langle Z\rangle}{2Z^2}$$

so by the Kunita–Watanabe identity we have

$$d\langle X, \log Z\rangle = \frac{d\langle X, Z\rangle}{Z}.$$

Let $\hat{X} = X - \langle X, \log Z\rangle$. The key observation is that $\hat{X}Z$ is a $\mathbb{P}$-local martingale because of the calculation

$$\begin{aligned} d(\hat{X}Z) &= Z(dX - d\langle X, M\rangle) + \hat{X}\ dZ + d\langle X, Z\rangle \\ &= Z\ dX + \hat{X}\ dZ \end{aligned}$$

By localisation we can assume that $\hat{X}$ is bounded. Since Z is a uniformly integrable $\mathbb{P}$-martingale then Z is in class D, i.e.

$$\{Z_T : T \text{ bounded stopping time }\}$$

is uniformly $\mathbb{P}$-integrable for all $t \geq 0$. By boundedness, the collection

$$\{\hat{X}_T Z_T : T \text{ bounded stopping time }\}$$

is also $\mathbb{P}$-uniformly integrable, and hence the local martingale $\hat{X}Z$ is actually a uniformly integrable $\mathbb{P}$-martingale. In particular, $\hat{X}_t Z_t \to \hat{X}_\infty Z_\infty$ in $L^1(\mathbb{P})$ and $\mathbb{P}$-a.s. (and $\mathbb{Q}$-a.s. by equivalence) By Bayes's formula we have

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}}(\hat{X}_\infty|\mathcal{F}_t) &= \frac{\mathbb{E}^{\mathbb{P}}(Z_\infty \hat{X}_\infty|\mathcal{F}_t)}{\mathbb{E}^{\mathbb{P}}(Z_\infty|\mathcal{F}_t)} \\ &= \hat{X}_t\end{aligned}$$

This shows that $\hat{X} = X - \langle X, \log Z\rangle$ is a $\mathbb{Q}$-martingale. □

A useful result which we have already used implicitly is the following:

Proposition (Representation of positive local martingales). *A positive, continuous process Z is a local martingale if and only if $Z = e^{M - \langle M\rangle/2}$ for some continuous local martingale M.*

Proof. Suppose Z is a positive local martinglae. Let

$$M_t = \int_0^t \frac{dZ_s}{Z_s},$$

so that

$$\langle M\rangle_t = \int_0^t \frac{d\langle Z\rangle_s}{Z_s^2}.$$

Now by Itô's formula we have

$$d \log Z = dM - \frac{1}{2} d\langle M\rangle.$$

Conversely, if $Z = e^{M - \langle M\rangle/2}$ for a local martingale M, then Itô's formula says

$$dZ = Z dM.$$

and hence Z is a local martingale. □

Definition. For a continuous semimartingale X, we will use the notation $\mathcal{E}(X) = e^{X - \langle X\rangle/2}$ to denote the stochastic (or Doléans-Dade) exponential.

An important application of the previous theorem is this natural generalisation of the above Gaussian example:

Theorem (Cameron–Martin–Girsanov). *Fix a probability measure $\mathbb{P}$ and let W be an n-dimensional a $\mathbb{P}$-Brownian motion and let α be an n-dimensional predictable process such that*

$$\int_0^\infty \|\alpha_s\|^2 ds < \infty \quad \mathbb{P} - a.s.$$

Let

$$\begin{aligned}Z_t &= \mathcal{E}\left(\int \alpha \cdot dW\right)_t \\ &= e^{\int_0^t \alpha_s \cdot dW_s - \frac{1}{2}\int_0^t \|\alpha_s\|^2 ds}\end{aligned}$$

and suppose that Z is a $\mathbb{P}$-uniformly integrable martingale. Let

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

The process

$$\hat{W}_t = W_t - \int_0^t \alpha_s ds$$

defines a $\mathbb{Q}$-Brownian motion.

PROOF. By the previous theorem, the process $\hat{W}$ is a $\mathbb{Q}$-local martingale. Also, note by example sheet 3, u.c.p convergence and hence the calculation of quadratic variation is unaffected by equivalent changes of measure. Now note that

$$\langle \hat{W}^i, \hat{W}^j \rangle_t = \langle W^i, W^j \rangle_t = \begin{cases} t & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

so $\hat{W}$ is a $\mathbb{Q}$-Brownian motion by Lévy's characterisation. □

We now discuss when Z is a true martingale, but not necessarily uniformly integrable.

DEFINITION. Probability measures $\mathbb{P}$ and $\mathbb{Q}$ on a measurable space $(\Omega, \mathcal{F})$ with filtration $\mathbb{F}$ are *locally equivalent* iff for every $t \geq 0$, the restrictions $\mathbb{P}|_{\mathcal{F}_t}$ and $\mathbb{Q}|_{\mathcal{F}_t}$ are equivalent.

THEOREM (Radon–Nikodym, local version). *If the measures $\mathbb{P}$ and $\mathbb{Q}$ are locally equivalent then there exists a positive $\mathbb{P}$-martingale Z with $\mathbb{E}(Z_0) = 1$ such that*

$$\frac{d\mathbb{Q}|_{\mathcal{F}_t}}{d\mathbb{P}|_{\mathcal{F}_t}} = Z_t.$$

PROOF. To lighten the notation, let $\mathbb{P}_t = \mathbb{P}|_{\mathcal{F}_t}$ and $\mathbb{Q}_t = \mathbb{Q}|_{\mathcal{F}_t}$. Note that by the 'standard machine' of measure theory,

$$\mathbb{E}^{\mathbb{P}_t}(X) = \mathbb{E}^{\mathbb{P}}(X)$$

if X is non-negative and $\mathcal{F}_t$-measurable.

Now, by the Radon–Nikodym theorem, for every $t \geq 0$ there exists an $\mathcal{F}_t$ measurable positive random variable Z_t such that

$$\frac{d\mathbb{Q}_t}{d\mathbb{P}_t} = Z_t.$$

We need only show that Z is a $\mathbb{P}$-martingale. Let $0 \leq s \leq t$ and let $A \in \mathcal{F}_s$. By the definition of the Radon–Nikodym density, we have

$$\mathbb{Q}_s(A) = \mathbb{E}^{\mathbb{P}_s}(Z_s \mathbb{1}_A) = \mathbb{E}^{\mathbb{P}}(Z_s \mathbb{1}_A)$$

Also, since $A \in \mathcal{F}_t$ as well, we have

$$\mathbb{Q}_t(A) = \mathbb{E}^{\mathbb{P}}(Z_t \mathbb{1}_A)$$

Of course $\mathbb{Q}_s(A) = \mathbb{Q}_t(A)$ since $A \in \mathcal{F}_s \subseteq \mathcal{F}_t$. Since A was arbitrary, we have shown

$$\mathbb{E}^{\mathbb{P}}(Z_t | \mathcal{F}_s) = Z_s.$$

□

Remark. Locally equivalent measures $\mathbb{P}$ and $\mathbb{Q}$ are actually equivalent if $\mathcal{F} = \sigma\big(\bigcup_{t\geq 0} \mathcal{F}_t\big)$ and the Radon–Nikodym density martingale Z is uniformly integrable. In this case there exists a positive random variable Z_∞ such that $Z_t \to Z_\infty$ a.s. and in $L^1(\Omega, \mathcal{F}, \mathbb{P})$ and such that

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

Example. Let $\mathbb{P}$ a probability measure such that the process W is a real Brownian motion, and let $\mathbb{Q}$ be a probability measure such that the process $\hat{W}$ is a Brownian motion, where $\hat{W}_t = W_t + at$ for some constant $a \neq 0$. The measures $\mathbb{P}$ and $\mathbb{Q}$ are locally equivalent with density process

$$\frac{d\mathbb{Q}|_{\mathcal{F}_t}}{d\mathbb{P}|_{\mathcal{F}_t}} = e^{-aW_t - a^2 t/2}$$

by the Cameron–Martin theorem. Note that the density process is a $\mathbb{P}$ martingale, but it is not uniformly integrable, so the measures $\mathbb{Q}$ and $\mathbb{P}$ are not equivalent. Indeed, note that

$$\mathbb{P}\left(\frac{W_t}{t} \to 0 \text{ as } t \to \infty\right) = 1$$

but

$$\mathbb{Q}\left(\frac{W_t}{t} \to 0 \text{ as } t \to \infty\right) = \mathbb{Q}\left(\frac{\hat{W}_t}{t} \to a \text{ as } t \to \infty\right) = 0.$$

For these theorems to be useful, we need some criterion to check whether a positive local martingale is actually uniformly integrable.

Proposition. *Let M be a continuous local martingale with $M_0 = 0$ and let*

$$Z = e^{M - \langle M \rangle/2}.$$

(1) *Z is a local martingale and supermartingale.*
(2) *Z is a true martingale if and only if $\mathbb{E}(Z_t) = 1$ for all $t \geq 0$.*
(3) *$Z_t \to Z_\infty$ a.s., and Z is a uniformly integrable martingale if and only if $\mathbb{E}(Z_\infty) = 1$.*
(4) *(Novikov's criterion (1973)) If $\mathbb{E}(e^{\frac{1}{2}\langle M \rangle_\infty}) < \infty$, then Z is a uniformly integrable martingale.*

Proof. Items (1) and (2) are on example sheet 2. For (3), non-negative martingales are bounded in L^1 and hence converge by the martingale convergence theorem. And if $\mathbb{E}(Z_\infty) = 1 = \lim_{t\uparrow\infty} \mathbb{E}(Z_t)$ then $Z_t \to Z_\infty$ by in L^1 Scheffé's lemma. Convergence in L^1 implies uniform integrability by Vitali's theorem. $\square$

To prove Novikov's theorem, we first establish two lemmas. This proof is inspired by a paper of Krylov.

Lemma. *Let X be a continuous local martingale with $X_0 = 0$. Then for any $p, q > 1$ and stopping time T, we have*

$$\mathbb{E}\left(\mathcal{E}(X)_T^p\right) \leq \mathbb{E}\big(e^{\frac{pq(pq-1)}{2(q-1)}\langle X \rangle_\infty}\big)^{\frac{q-1}{q}}.$$

PROOF. First note the identity

$$\mathcal{E}(X)^p = \mathcal{E}(pqX)^{1/q} \exp\left(\tfrac{pq(pq-1)}{2(q-1)}\langle X\rangle\right)^{\frac{q-1}{q}}.$$

By Hölder's inequality we have

$$\mathbb{E}\left(\mathcal{E}(X)_T^p\right) \le \mathbb{E}\left(\mathcal{E}(pqX)_T\right)^{\frac{1}{q}} \mathbb{E}(e^{\frac{pq(pq-1)}{2(q-1)}\langle X\rangle_T})^{\frac{q-1}{q}}.$$

Now $\mathcal{E}(pqX)$ is a supermartingale, so $\mathbb{E}(\mathcal{E}(pqX)_T) \le 1$ by the optional sampling theorem. The result follows from the inequality $\langle X\rangle_T \le \langle X\rangle_\infty$. □

LEMMA. *Let X be a continuous local martingale with $X_0 = 0$. Then for any $p > 1$, we have*

$$\mathbb{E}\left(\mathcal{E}(pX)_\infty\right) \ge \mathbb{E}\left(\mathcal{E}(X)_\infty\right)^p \mathbb{E}(e^{\frac{p^2}{2}\langle X\rangle_\infty})^{-\frac{p-1}{p}}.$$

PROOF. As above, note the identity

$$\mathcal{E}(X) = \mathcal{E}(pX)^{\frac{1}{p}} \exp\left(\tfrac{p}{2}\langle X\rangle\right)^{\frac{p-1}{p}}.$$

Then Hölder's inequality implies

$$\begin{aligned}\mathbb{E}[\mathcal{E}(X)_\infty] &\le \mathbb{E}[\mathcal{E}(pX)_\infty]^{\frac{1}{p}} \mathbb{E}[e^{\frac{p}{2}\langle X\rangle_\infty}]^{\frac{p-1}{p}} \\ &\le \mathbb{E}[\mathcal{E}(pX)_\infty]^{\frac{1}{p}} \mathbb{E}[e^{\frac{p^2}{2}\langle X\rangle_\infty}]^{\frac{p-1}{p^2}}\end{aligned}$$

where the second line follows from Jensen's inequality. □

PROOF OF NOVIKOV'S CRITERION. Suppose $\mathbb{E}[e^{\frac{1}{2}\langle M\rangle_\infty}] < \infty$.
Note that

$$\frac{pq(pq-1)}{q-1} > 1$$

for all $p, q > 1$, and

$$\lim_{q\downarrow 1}\lim_{p\downarrow 1} \frac{pq(pq-1)}{q-1} = 1$$

Therefore, for any $0 < a < 1$ we can find $p, q > 1$ such that $a^2 \frac{pq(pq-1)}{q-1} = 1$ and hence

$$\mathbb{E}\left(\mathcal{E}(aM)_T^p\right) \le \mathbb{E}(e^{\frac{1}{2}\langle M\rangle_\infty})^{\frac{q-1}{q}} < \infty$$

by the first lemma. In particular, the collection

$$\{\mathcal{E}(aM)_T : T \text{ stopping time }\}$$

is bounded in L^p, and hence uniformly integrable. That means the local martingale $\mathcal{E}(aM)$ is in class D, implying it is a uniformly integrable martingale and hence

$$\mathbb{E}\left[\mathcal{E}(aM)_\infty\right] = 1.$$

By the second lemma we have

$$\begin{aligned}\mathbb{E}\left(\mathcal{E}(M)_\infty\right) &\ge \mathbb{E}\left(\mathcal{E}(aM)_\infty\right)^{1/a} \mathbb{E}(e^{\frac{1}{2a^2}\langle aM\rangle_\infty})^{a-1} \\ &= \mathbb{E}(e^{\frac{1}{2}\langle M\rangle_\infty})^{a-1}\end{aligned}$$

The conclusion follows upon sending $a \uparrow 1$. □

3. The martingale representation theorem

Now we ask explore the characterisation of local martingales, not in a general probability space, but in the special but important case where all measurable events are generated by a Brownian motion.

THEOREM (Itô's martingale representation theorem). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on which a d-dimensional Brownian motion $W = (W_t)_{t\geq 0}$ is defined, and let the filtration $(\mathcal{F}_t)_{t\geq 0}$ be the (completed, right-continuous) filtration generated by W. Assume $\mathcal{F} = \sigma\big(\bigcup_{t\geq 0} \mathcal{F}_t\big)$.*

Let $M = (M_t)_{t\geq 0}$ be a càdlàg locally square-integrable local martingale. Then there exists an Leb $\times\, \mathbb{P}$*-unique predictable d-dimensional process $\alpha = (\alpha_t)_{t>0}$ such that $\int_0^t \|\alpha_s\|^2 ds < \infty$ almost surely for all $t \geq 0$ and*

$$M_t = M_0 + \int_0^t \alpha_s \cdot dW_s.$$

In particular, the local martingale M is continuous.

PROOF. (Uniqueness) Suppose

$$M_t = M_0 + \int_0^t \alpha_s' \cdot dW_s$$

for another predictable process α'. Then by subtracting these two representations we have

$$\int_0^t (\alpha_s - \alpha_s') \cdot dW_s = 0 \text{ for all } t \geq 0.$$

Since right-hand side is a square-integrable martingale, we can apply Itô's isometry

$$\mathbb{E} \int_0^\infty \|\alpha_s - \alpha_s'\|^2 ds = 0$$

from which the uniqueness follows. $\square$

For existence we proceed via a series of lemmas:

LEMMA (from local martingales to martingales). *Suppose M is a locally square integrable local martingale, so that there exists an increasing sequence $(T_n)_n$ of stopping times such that the stopped processes M^{T_n} are square-integrable martingales. If we can find an integral representation $M^{T_n} = M_0 + \int \alpha^n dW$ for each n, then there exists a predictable $\alpha \in L_{\mathrm{loc}}(W)$ such that*

$$M = M_0 + \int \alpha dW.$$

PROOF. Note that we have

$$(M^{T_{n+1}})^{T_n} = M_0 + \int \alpha^{n+1} \mathbb{1}_{(0,T_n]} dW$$

and hence $\alpha^n = \alpha^{n+1}\mathbb{1}_{(0,T_n]}$ by uniqueness. In particular, we can find the unique integral representation of M by setting

$$\alpha_t(\omega) = \alpha_t^n(\omega) \text{ on } \{(t,\omega) : t \leq T_n(\omega)\}.$$

$\square$

Let M be a locally square-integrable local martingale. By lemma the above, in order to show that M has the stochastic representation property, it is enough to show that for every stopping time T such that M^T is a square integrable martingale there exists an $\alpha \in L^2(W)$ such that $M^T = M_0 + \int \alpha \cdot dW$. Hence we may assume M is a square-integrable martingale.

LEMMA (from martingales to random variables). *Suppose that $X \in L^2(\mathbb{P})$ has the property that*

$$X = x + \int_0^\infty \alpha_s \cdot dW_s$$

for a constant x and a predictable process α such that

$$\mathbb{E} \int_0^\infty \|\alpha_s\|^2 ds < \infty.$$

Then

$$\mathbb{E}(X|\mathcal{F}_t) = x + \int_0^t \alpha_s \cdot dW_s$$

PROOF. Let

$$M_t = x + \int_0^t \alpha_s \cdot dW_s.$$

Note that M is a continuous local martingale, as it is the stochastic integral with respect to the martingale W with quadratic variation

$$\langle M \rangle_t = \int_0^t \|\alpha_s\|^2 ds.$$

Since $\langle M \rangle_\infty$ is integrable by assumption, the local martingale X is in fact square integrable martingale. Hence

$$\mathbb{E}(X|\mathcal{F}_t) = M_t.$$

□

Let M be a square-integrable martingale. By the martingale convergence theorem, there is a square-integrable random variable X such that $M_t = \mathbb{E}(X|\mathcal{F}_t)$. By the last lemma, in order to show the integral representation of M, it is enough to show that for every $X \in L^2(\mathbb{P})$ exists a real x and $\alpha \in L^2(W)$ such that $X = x + \int_0^\infty \alpha_s \cdot dW_s$.

LEMMA (from random variables to approximations). *Suppose $X^n \to X$ in $L^2(\mathbb{P})$ and each X^n has the form*

$$X^n = x^n + \int_0^\infty \alpha_s^n dW_s.$$

where $\alpha^n \in L^2(W)$. Then X has the form

$$X = x + \int_0^\infty \alpha_s dW_s.$$

where $\alpha \in L^2(W)$.

Proof. Note that by Itô's isometry

$$\mathbb{E}[(X^n - X^m)^2] = (x^n - x^m)^2 + \mathbb{E}\int_0^\infty \|\alpha_s^n - \alpha_s^m\| ds \to 0.$$

Hence the sequences (x^n) and $(\alpha^n)_n$ are Cauchy in $\mathbb{R}$ and $L^2(W)$ respectively. The conclusion follows from completeness. $\square$

From the above lemma, to show that a random variable $X \in L^2(\mathbb{P})$ has the integral representation property, it is enough to exhibit a sequence $X^n \to X$ in $L^2(\mathbb{P})$ such that each X^n has the integral representation property. In particular, to show that every element of $L^2(\mathbb{P})$ has the integral representation property, it is enough to show that a dense subset of $L^2(\mathbb{P})$ has the integral representation property.

Lemma (density of cylindrical functions). *Random variables of the form*

$$f(W_{t_1} - W_{t_0}, \ldots, W_{t_N} - W_{t_{N-1}})$$

are dense in $L^2(\mathbb{P})$.

Proof. Introduce the notation $t_k^n = k2^{-n}$ and let

$$\mathcal{G}_n = \sigma(W_{t_k^n} - W_{t_{k-1}^n}, 1 \le k \le 2^{2n}).$$

Note that $\mathcal{G}_n \subseteq \mathcal{G}_{n+1}$ and $\mathcal{F} = \left(\bigcup_n \mathcal{G}_n\right)$. Given a $X \in L^2(\mathbb{P})$, let $X_n = \mathbb{E}(X|\mathcal{G}_n)$. By the martingale convergence theorem we have $X_n \to X$ in $L^2(\mathbb{P})$. But by measurability, there exists a function $f : (\mathbb{R}^d)^N \to \mathbb{R}$ such that

$$X_n = f(W_{t_k^n} - W_{t_{k-1}^n}, 1 \le k \le N)$$

where $N = 2^{2n}$. $\square$

To establish the integral representation property for every element of $L^2(\mathbb{P})$, the above lemma says we only need to show that every $X \in L^2(\mathbb{P})$ of the form

$$X = f(W_{t_k} - W_{t_{k-1}}, 1 \le k \le N)$$

has the integral representation property.

Lemma (density of exponentials). *Fix a dimension d and let μ be a probability measure on $\mathbb{R}^d$. Consider the complex vector space of functions on $\mathbb{R}^d$ defined by*

$$E = \operatorname{span}\{e^{\mathrm{i}\theta\cdot x} : \theta \in \mathbb{R}^d\}$$

Then E is dense in the $L^2(\mu)$.

Proof. Let $f \in L^2(\mu)$ be orthogonal to E, that is to say

$$\int f(x)e^{\mathrm{i}\theta\cdot x}\mu(dx) = 0 \text{ for all } \theta \in \mathbb{R}^d.$$

This means that $f(x) = 0$ for μ-a.e. x. First of all, by considering the real and imaginary parts of f separately, we may assume f is real valued. Now let $d\nu_\pm = f^\pm d\mu$. Note that $\nu_\pm$ are finite measures since $f \in L^2 \subseteq L^1$. Now, the above equation and the uniqueness of characteristic functions of finite measures implies $\nu_+ = \nu_-$, or equivalently $f^+ = f^-$ μ-a.e. But $\mu\{f^+ > 0, f^- > 0) = 0$ and hence $f = 0$ μ-a.e.

Hence the closure of E is $\bar{E} = E^{\perp\perp} = \{0\}^\perp = L^2(\mu)$ as desired. $\square$

By the above lemma, to establish the integral representation of cylindrical random variables, it is enough to establish the integral representation of exponentials.

LEMMA. *Let*

$$X = e^{\mathrm{i} \sum_{k=1}^N \theta_k \cdot (W_{t_k} - W_{t_{k-1}})}$$

for fixed $\theta_k \in \mathbb{R}^n$ *and* $0 \leq t_0 < \ldots < t_N$*. Then there exists a constant* x *and a predictable process* α *valued in* $\mathbb{C}^n$ *such that*

$$\mathbb{E} \int_0^\infty \|\alpha_s\|^2 ds < \infty.$$

and

$$X = x + \int_0^\infty \alpha_s \cdot dW_s.$$

PROOF. Let

$$\beta = \mathrm{i} \sum_k \theta_k \mathbb{1}_{(t_{k-1}, t_k]}$$

and

$$M = \int \beta dW.$$

Note that

$$X = C\mathcal{E}(M)_\infty$$

where $C = e^{-\sum_k \|\theta_k\|^2 (t_k - t_{k-1})/2} = \mathbb{E}(X)$.

By Itô's formula

$$\mathcal{E}(M)_\infty = 1 + \int_0^\infty \mathcal{E}(M)_s \beta_s \cdot dW_s$$

we have the desired integral representation with

$$\alpha = C\mathcal{E}(M)\beta$$

since

$$\int_0^\infty \|\alpha_s\|^2 ds \leq C^4 \sum_k \|\theta_k\|^2 (t_k - t_{k-1}) < \infty.$$

$\square$

This concludes the proof of the martingale representation theorem.

4. Brownian local time

Let W be a one-dimensional Brownian motion, and fix $x_0 \in \mathbb{R}$. For the sake of motivation, let

$$f(x) = |x - x_0|$$

with 'derivative'

$$f'(x) = \mathrm{sign}(x - x_0) = \begin{cases} +1 & \text{if } x > x_0 \\ 0 & \text{if } x = x_0 \\ -1 & \text{if } x < x_0 \end{cases}$$

and 'second derivative'

$$f''(x) = 2\delta(x - x_0)$$

where δ is the Dirac delta function. If Itô's formula was applicable, we would have

$$(*) \qquad |W_t - x_0| = |x_0| + \int_0^t \mathrm{sign}(W_s - x_0)dW_s + \int_0^t \delta(W_s - x_0)ds.$$

Now remember, the Dirac delta function is operationally defined by

$$\int_{-\infty}^{\infty} g(x)\delta(x - x_0)dx = g(x_0)$$

for smooth functions g. Of course, the correct way of viewing the above equation is to interpret the notation $\delta(x - x_0)dx$ as the measure with a single atom, assigning unit mass to $x = x_0$. Is there a similar interpretation of equation (*)?

Definition. The local time of the Brownian motion W at the point x up to time t is defined by

$$L_t^x = |W_t - x| - |x| - \int_0^t \mathrm{sign}(W_s - x)dW_s.$$

For fixed x, it turns out that $(L_t^x)_{t\geq 0}$ is a non-decreasing process. Furthermore, it increases precisely on the x level set of the Brownian motion. Since this set is of Lebesgue-measure 0 almost surely, the local time is another time scale, singular with respect to the natural dt time scale. But it can be computed in terms of the Lebesgue measure of the amount of time the Brownian motion spends in a small neighbourhood of x:

Theorem.

$$\frac{1}{2\varepsilon}\mathrm{Leb}(s \in [0,t] : |W_s - x| \leq \varepsilon) \to L_t^x \;\; u.c.p.$$

Proof. Without loss assume $x = 0$. Let

$$f_\varepsilon(x) = \begin{cases} |x| & \text{if } |x| > \varepsilon \\ \frac{1}{2\varepsilon}x^2 + \frac{\varepsilon}{2} & \text{if } |x| \leq \varepsilon \end{cases}$$

Itô's formula yields

$$\begin{aligned} f_\varepsilon(W_t) &= f_\varepsilon(0) + \int_0^t f_\varepsilon'(W_s)dW_t + \frac{1}{2}\int_0^t f_\varepsilon''(W_s)ds \\ &= \frac{\varepsilon}{2} + \int_0^t f_\varepsilon'(W_s)dW_t + \frac{1}{2\varepsilon}\mathrm{Leb}(s \in [0,t] : |W_s| \leq \varepsilon). \end{aligned}$$

Since $f_\varepsilon(x) \to |x|$ uniformly, we have

$$f_\varepsilon(W_t) - \frac{\varepsilon}{2} \to |W_t| \text{ uniformly almost surely.}$$

Also note that

$$f_\varepsilon'(x) \to \mathrm{sign}(x) \text{ for all } x \in \mathbb{R}.$$

and that

$$|f_\varepsilon'(x) - \mathrm{sign}(x)| \leq 1 \text{ for all } (x, \varepsilon)$$

so that

$$\mathbb{E}\left[\sup_{0\leq s\leq t}\left(\int_0^s [f_\varepsilon'(W_s) - \mathrm{sign}(W_s)]dW_s\right)^2\right] \leq 4\mathbb{E}\left[\int_0^t [f_\varepsilon'(W_s) - \mathrm{sign}(W_s)]^2 ds\right] \to 0$$

by the dominated convergence theorem. In particular, the stochastic integral converges uniformly on compacts in probability. $\square$

REMARK. The proof above is not quite rigorous, since we have proved Itô's formula only for twice-*continuously* differentiable functions, but the function f_ε has a discontinuous second derivative. So actually, we really should deal with the discontinuities of f_ε'' by approximating it with a smooth function, then take the limit as above...

There are lots of interesting facts about Brownian local time, but we will not need them in this course. Here is one stated without proof.

THEOREM. *Let W be a Brownian motion, L^0 its local time at* 0 *and $M_t = \sup_{0 \le s \le t} W_t$ its running maximum. Then the two dimensional processes $(|W|, L^0)$ and $(M - W, M)$ have the same law.*

Finally, we can also define local time for continuous semimartingales:

DEFINITION. If X is a continuous semimartingale, the local time of X at the point x up to time t is

$$L_t^x = |X_t - x| - |X_0 - x| - \int_0^t g(X_s - x) dX_s$$

where $g = \mathbb{1}_{[0,\infty)} - \mathbb{1}_{(-\infty,0)} = \text{sign} + \mathbb{1}_{\{0\}}$.

The following theorem shows that the above definition is useful:

THEOREM (Tanaka's formula). *Let f be the difference of two convex functions (so that f is continuous, has a right-derivative f_+' at each point, and the function f_+' is of finite variation) Let X be a continuous semimartingale and (L_t^x) its local time. Then*

$$f(X_t) = f(X_0) + \int_0^t f_+'(X_s) dX_s + \int_{-\infty}^{\infty} L_t^x df_+'(x).$$

CHAPTER 5

Stochastic differential equations

1. Definitions of solution

The remaing part of the course is to study so-called stochastic differential equations, i.e. equations of the form

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

where the functions $b : \mathbb{R}^n \to \mathbb{R}^n$ and $\sigma : \mathbb{R}^n \to \mathbb{R}^{n\times d}$ are given and W is an d-dimensional Brownian motion.

Recall our motivation. If X satisfies equation (*) then it should be the case that

$$X_{t+\Delta t} = X_t + \int_t^{t+\Delta t} b(X_s)ds + \int_t^{t+\Delta t} \sigma(X_s)dW_s.$$

Assuming that b and σ are sufficiently well-behaved, in particular, so that the dW integral is a square-integrable martingale, we should be able to conclude (by Itô's isometry) that

$$\mathbb{E}(X_{t+\Delta t}|\mathcal{F}_t) \approx X_t + b(X_t)\Delta t \text{ and } \mathrm{Cov}(X_{t+\Delta t}|\mathcal{F}_t) \approx \sigma(X_t)\sigma(X_t)^\top \Delta t$$

when $\Delta t > 0$ is small. Indeed, it is the formal calculation above which often leads applied mathematicians, physicists, economists, etc to consider equation (*) in the first place.

A solution of the SDE (*) consists of the following ingredients:

(1) a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a (complete right-continuous) filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$,
(2) a d-dimensional Brownian motion W compatible with $\mathbb{F}$,
(3) an adapted X process such that $b(X)$ and $\sigma(X)$ are predictable and

$$\int_0^t \|b(X_s)\|ds < \infty \text{ and } \int_0^t \|\sigma(X_s)\|^2 ds < \infty \text{ a.s. for all } t \geq 0$$

where $\|\sigma\|^2 = \mathrm{trace}(\sigma\sigma^{\mathrm{T}})$ is the Frobenius (a.k.a. Hilbert–Schmidt) matrix norm, and

$$X_t = X_0 + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dW_s$$

for all $t \geq 0$.

Item (3) is obviously the integral version of the formal differential equation (*). The integrability conditions are to ensure that the dt integral can be interpreted as a pathwise Lebesgue integral and the dW integral is well-defined according to our stochastic integration theory.

It turns out there are two useful notions solution to an SDE. The first one seems very natural to me:

Definition. A *strong solution* of the SDE (*) takes as the data the functions b and σ along with the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and the Brownian motion W. The filtration $\mathbb{F}$ is assumed to be generated by W. The output is the process X.

Note that assuming that X is adapted to the filtration generated by W says that X is a functional of the Brownian sample path W. That is, given only the infinitesimal characteristics of the dynamics (the functions b and σ) and the realisation of the 'noise', the process X can be reconstructed. Strong solutions are consistent with a notion of causality, in that the driving noise 'causes' the random fluctuations of X. It is this assumption which distinguishes strong solutions from weak solutions:

Definition. A *weak solution* of the SDE (*) takes as the data the functions b and σ. The output is the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the filtration $\mathbb{F}$, the Brownian motion W and the process X.

From the point of stochastic modelling, weak solutions are in some sense more natural. Indeed, from an applied perspective, the natural input is the infinitesimal characteristics b and σ. The Brownian motion W could be considered an auxiliary process, since the process of interest is really X.

Example (Tanaka's example). The purpose of the following example is illustrate the difference between the notions of weak and strong solutions.

Consider the SDE with $n = d = 1$ and

$$dX_t = g(X_t)dW_t, \quad X_0 = 0 \tag{**}$$

where

$$\begin{aligned} g(x) &= \mathrm{sign(x)} + \mathbb{1}_{\{x=0\}} \\ &= \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0. \end{cases} \end{aligned}$$

First we show that this equation has a weak solution. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on which a Brownian motion X is defined. Let $\mathbb{F}$ be any filtration compatible with X. Define a local martingale W by

$$W_t = \int_0^t g(X_s)dX_s.$$

Note that the quadratic variation of W is

$$\langle W \rangle_t = \int_0^t g(X_s)^2 d\langle X \rangle_s = t$$

and hence W is a Brownian motion by Lévy's characterisation. Now note that

$$\begin{aligned} X_t &= \int_0^t g(X_s)^2 dX_s \\ &= \int_0^t g(X_s)dW_s. \end{aligned}$$

Therefore, the setup $(\Omega, \mathcal{F}, \mathbb{P})$, $\mathbb{F}$, W and X form a weak solution of the SDE (**).

Now we show that Tanaka's SDE does not have a strong solution. First we need a little fact. Let X be a Brownian motion. Then

$$\int_0^t g(X_s)dX_s - \int_0^t \mathrm{sign}(X_s)dX_s = 0$$

since the expected quadratic variation of the left-hand side is

$$\mathbb{E}\int_0^t \mathbb{1}_{\{X_s\}}dx = \int_0^t \mathbb{P}(X_s = 0)ds = 0.$$

Now let X is be any solution of (*). Note that X is a Brownian motion since

$$\langle X\rangle_t = \int_0^t g(X_s)^2 ds = t.$$

The crucial observation is that W can be recovered from X by

$$\begin{aligned} W_t &= \int_0^t g(X_s)dX_s \\ &= \int_0^t \mathrm{sign}(X_s)dX_s \\ &= |X_t| - \lim_{\epsilon\downarrow 0}\frac{1}{2\epsilon}\mathrm{Leb}(s\in[0,t] : |X_s|\le\epsilon). \end{aligned}$$

In particular, the random variable W_t is measurable with respect to $\sigma(|X_s| : 0 \le s \le t)$.

Suppose, for the sake of finding a contradiction, that X is a strong solution. Then X is adapted to the filtration generated by W. But W is determined by $|X|$. In particular, this says that we can determine the sign of the random variable X_t just by observing $(|X_s|)_{s\ge 0}$, an absurdity!

The existence of weak, but not strong, solutions might come as a surprise. Indeed, consider the SDE

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

It is natural to approximate this equation setting $\hat{X}_0 = X_0$ and discretising

$$\hat{X}_{t_k} = \hat{X}_{t_{k-1}} + b(\hat{X}_{t_{k-1}})(t_k - t_{k-1}) + \sigma(\hat{X}_{t_{k-1}})(W_{t_k} - W_{t_{k-1}})$$

for a family of times $0 \le t_0 < t_1 < \ldots$. By construction, the random variable $\hat{X}_{t_k}$ is measurable with respect to the sigma-field generated by the random variables $X_0, W_{t_0}, \ldots, W_{t_k}$. One could say that there is a 'causality principle', in the sense that the Brownian motion is 'driving' the dynamics of $\hat{X}$.

The bizarre phenomenon is that somehow this measurable dependence on the driving noise may not hold for the SDE. In particular, in some cases such as Tanaka's example, it is impossible to reconstruct X_t only from X_0 and $(W_s)_{0\le s\le t}$ – some other external randomisation is necessary.

2. Notions of uniqueness

Just as there are at least two notions of solution of an SDE, there are at least two notions of uniqueness of those solutions.

Again, we're studying the SDE

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

Definition. The SDE (*) has the pathwise uniqueness property iff for any two solutions X and X' defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$, filtration $\mathbb{F}$ and Brownian motion W such that $X_0 = X'_0$ a.s., we must have

$$\mathbb{P}(X_t = X'_t \text{ for all } t \geq 0) = 1.$$

We see that the notion of pathwise uniqueness can be too strict for some equations. Here is another notion of uniqueness, which might be more suitable for stochastic modelling:

Definition. The SDE (*) has the uniqueness in law property iff for any two weak solutions $(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F}, X, W)$ and $(\Omega', \mathcal{F}', \mathbb{P}', \mathbb{F}', X', W')$ such that $X_0 \sim X'_0$, we must have

$$(X_t)_{t\geq 0} \sim (X'_t)_{t\geq 0}.$$

Example (Tanaka's example, continued.). Again we're considering the SDE

$$dX_t = g(X_t)dW_t, \quad X_0 = 0 \tag{**}$$

where $g(x) = \text{sign(x)} + \mathbb{1}_{\{x=0\}}$. Suppose that X is a weak solution of Tanaka's SDE. Note that

$$\begin{aligned} d(-X_t) &= -g(X_t)dW_t \\ &= [g(-X_t) - 2\mathbb{1}_{\{X_s=0\}}]dW_t \end{aligned}$$

As before, we conclude that

$$\int_0^t \mathbb{1}_{\{X_s=0\}}dW_t = 0$$

by computing the expected quadratic variation by Fubini's theorem. Hence

$$d(-X_t) = g(-X_t)dW_t.$$

In particular $-X$ is also weak solution of the SDE (**), and the pathwise uniqueness property does not hold.

On the other hand, note that every weak solution of the Tanaka's SDE are Brownian motions. Hence, the SDE does have the uniqueness in law property.

Example. Consider the ODE

$$dX_t = 2\sqrt{X_t}dt, \quad X_0 = 0$$

There is no uniqueness, since there is a whole family of solutions

$$X_t = \begin{cases} 0 & \text{if } t \leq T \\ (t-T)^2 & \text{if } t > T. \end{cases}$$

The problem is the function $b(x) = 2\sqrt{x}$ is not smooth at $x = 0$, since b' is unbounded. Therefore, it seems that a natural condition to impose to ensure path-wise uniqueness is smoothness.

THEOREM. *The SDE (*) has path-wise uniqueness if the functions b and σ are locally Lipschitz, in that for every $N > 0$ there exists a constant $K_N > 0$ such that*

$$\|b(x) - b(y)\| \leq K_N \|x - y\| \text{ and } \|\sigma(x) - \sigma(y)\| \leq K_N \|x - y\|$$

for all x, y such that $\|x\| \leq N$ and $\|y\| \leq N$.

Before we launch into the proof, we will need a lemma that is of great importance in classical ODE theory.

THEOREM (Gronwall's lemma). *Suppose there are constants $a \in \mathbb{R}$ and $b > 0$ such that the locally integrable function f satisfies*

$$f(t) \leq a + b \int_0^t f(s) ds \text{ for all } t \geq 0.$$

Then

$$f(t) \leq a e^{bt} \text{ for all } t \geq 0.$$

PROOF. By the assumption of local integrability we can apply Fubini's theorem to conclude

$$\begin{aligned}\int_{s=0}^t \int_{u=0}^s b e^{b(t-s)} f(u) du \, ds &= \int_{u=0}^t \int_{s=u}^t b e^{b(t-s)} f(u) ds \, du \\ &= \int_{u=0}^t (e^{b(t-u)} - 1) f(u) du.\end{aligned}$$

Hence, we have

$$\begin{aligned}\int_0^t f(s) ds &= \int_0^t e^{(t-s)} \left(f(s) - b \int_0^s f(u) du \right) ds \\ &\leq \int_0^t e^{b(t-s)} a \, ds \\ &= \frac{a}{b}(e^{bt} - 1)\end{aligned}$$

and so the result follows from substituting this bound into the inequality

$$\begin{aligned}f(t) &\leq a + b \int_0^t f(s) ds \\ &\leq a + a(e^{bt} - 1) \\ &= a e^{bt}.\end{aligned}$$

□

PROOF OF PATH-WISE UNIQUENESS. Let X and X' be two solutions of (*) defined on the same probability space with $X_0 = X'_0$. Fix an $N > 0$ and let $T_N = \inf\{t \geq 0 : |X_t| > N \text{ or } |X'_t| > N\}$. Finally,

$$f(t) = \mathbb{E}(\|X_{t \wedge T_N} - X'_{t \wedge T_N}\|^2).$$

We will show that $f(t) = 0$ for all $t \geq 0$. By the continuity of the sample paths of X and X' this will imply

$$\mathbb{P}(\sup_{0 \leq t \leq T_N} \|X_s - X'_s\| = 0) = 1.$$

By sending $N \to \infty$, we will conclude that X and X' are indistinguishable.

Now, by Itô's formula, we have

$$\begin{aligned}\|X_{t\wedge T_N} - X'_{t\wedge T_N}\|^2 = &\int_0^{t\wedge T_N} 2(X_s - X'_s)\cdot\{[b(X_s) - b(X'_s)]ds + [\sigma(X_s) - \sigma(X'_s)]dW_s\} \\ &+ \int_0^{t\wedge T_N} \|\sigma(X_s) - \sigma(X'_s)\|^2 ds\end{aligned}$$

Note that the integrand of the dW integral is bounded on $\{t \leq T_N\}$, so the dW integral is a mean-zero martingale. Hence, computing the expectation of both sides and using the Lipschitz bound yields

$$\begin{aligned}f(t) =&\mathbb{E}\int_0^{t\wedge T_N} 2(X_s - X'_s)\cdot[b(X_s) - b(X'_s)]ds + \mathbb{E}\int_0^{t\wedge T_N} \|\sigma(X_s) - \sigma(X'_s)\|^2 ds \\ \leq&(2K_N + K_N^2)\int_0^t f(s)ds\end{aligned}$$

The conclusion now follows from Gronwall's lemma. □

The following theorem shows that uniqueness in law really is a weaker notion than pathwise uniqueness.

Theorem (Yamada–Watanabe). *If an SDE has the pathwise uniqueness property, then it has the uniqueness in law property.*

Sketch of the proof. The essential idea is to take two weak solutions, and then couple them onto the same probability space in order to exploit the pathwise uniqueness.

Let $(\Omega, \mathbb{P}, \mathcal{F}, \mathbb{F}, W, X)$ and $(\Omega', \mathbb{P}', \mathcal{F}', \mathbb{F}', W', X')$ be two weak solutions of the SDE. Suppose $X_0 \sim X'_0 \sim \lambda$ for a probability measure λ on $\mathbb{R}^n$.

Let C^k denotes the space of continuous functions from $[0,\infty)$ into $\mathbb{R}^k$. We can define a probability measures on $\mathbb{R}^n \times C^d \times C^n$ by the rule

$$\mu(A \times B \times C) = \mathbb{P}(X_0 \in A, W \in B, X \in C).$$

Since X_0 and W are independent under $\mathbb{P}$, we factorise this measure as

$$\mu(dx, dw, dy) = \lambda(dx)\mathbb{W}(dw)\nu(x, w; dy)$$

where $\mathbb{W}$ is the Wiener measure on $\mathbb{R}^d$ and where ν is the regular conditional probability measures on C^n such that

$$\nu(X_0, W; C) = \mathbb{P}(X \in C|X_0, W).$$

Define μ' similarly and note that

$$\mu'(dx, dw, dy) = \lambda(dx)\mathbb{W}(dw)\nu'(x, w; dy)$$

where ν' is defined analogously.

Now let $\hat{\Omega} = \mathbb{R}^n \times C^d \times C^n \times C^n$ be the sample space and $\hat{\mathbb{P}}$ be the probability measure defined by

$$\hat{\mathbb{P}}(dx, dw, dy, dy') = \lambda(dx)\mathbb{W}(dw)\nu(x, w; dy)\nu'(x, w; dy').$$

Finally define $\hat{X}_0(x,w,y,y') = x$, $\hat{W}_t(x,w,y,y') = w(t)$, and $\hat{X}_t(x,w,y,y') = y(t)$ and $\hat{X}'_t(x,w,y,y') = y'(t)$. Note that $\hat{X}$ and $\hat{X}'$ are two solutions of the SDE such that $\hat{X}_0 = \hat{X}'_0$ $\hat{\mathbb{P}}$-a.s. By pathwise uniqueness, we have

$$\hat{\mathbb{P}}(\hat{X}_t = \hat{X}'_t \text{ for all } t \geq 1) = 1.$$

Hence

$$\begin{aligned}\mathbb{P}(X \in C) &= \hat{\mathbb{P}}(\hat{X} \in C) \\ &= \hat{\mathbb{P}}(\hat{X}' \in C) \\ &= \mathbb{P}'(X' \in C)\end{aligned}$$

proving that X and X' have the same law. □

It remains to find easy-to-check sufficient conditions that an SDE has the uniqueness in law property. It turns out that this condition is intimately related to the Markov property, as well as the existence of solutions to certain PDEs. We will return this later.

3. Strong existence

Just as smoothness of the functions b and σ was sufficient for uniqueness, it is not very surprising that smoothness is also sufficient for existence. But now we will assume not only local smoothness, but a global smoothness.

To see what kind of bad behaviour the global Lipschitz assumption rules out, we consider an example with locally, but not globally, Lipschitz coefficients:

$$dX_t = X_t^2 dt.$$

The unique solution is given by

$$X_t = \frac{X_0}{1 - X_0 t}.$$

If $X_0 > 0$, then the solution only exists on the bounded interval $[0, 1/X_0)$.

Theorem (Itô). *Suppose b and σ are globally Lipschitz, in the sense that there exists a constant $K > 0$ such that*

$$\|b(x) - b(y)\| \leq K\|x - y\| \text{ and } \|\sigma(x) - \sigma(y)\| \leq K\|x - y\| \text{ for all } x, y \in \mathbb{R}^d.$$

Then there exists a unique strong solution to the SDE

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

Furthermore, if $\mathbb{E}(\|X_0\|^p) < \infty$ for some $p \geq 2$ then

$$\mathbb{E}(\sup_{0\leq s\leq t} \|X_s\|^p) < \infty$$

for all $t \geq 0$.

Here is one way to proceed. First note the following: suppose that we can prove the existence of a strong solution X^1 to the SDE for any initial condition X_0 on an interval $[0,T]$, where $T > 0$ is not random and does not depend on X_0. That is to say, suppose we can find a measurable map sending X_0 and the Brownian motion $(W_t)_{0\leq t\leq T}$ to the process X^1 where

$$X^1_t = X_0 + \int_0^t b(X^1_s)ds + \int_0^t \sigma(X^1_s)dW_s \text{ for } 0 \leq t \leq T$$

Then we can now use this same map with the intial condition X^1_T and the Brownian motion $(W_{u+T} - W_T)_{u\in[0,T]}$ to construct a process X^2 so that

$$X^2_u = X^1_T + \int_0^u b(X^2_s)ds + \int_0^u \sigma(X^2_s)d(W_{s+T} - W_T) \text{ for } 0 \leq u \leq T$$

Now we set

$$X_t = X_0 + X^1_t \mathbb{1}_{\{0<t\leq T\}} + X^2_{t-T}\mathbb{1}_{\{T<t\leq 2T\}}.$$

The process X is a solution of the SDE on $[0, 2T]$. Indeed, since $X = X^1$ on the interval $[0, T]$, and X^1 solves the SDE, we need only check that X solves the SDE on $[T, 2T]$:

$$\begin{aligned}
X_t &= X^1_T + \int_0^{t-T} b(X^2_s)ds + \int_0^{t-T} \sigma(X^2_s)d(W_{s+T} - W_T)\\
&= X_0 + \int_0^T b(X^1_s)ds + \int_0^T \sigma(X^1_s)dW_s + \int_T^t b(X^2_{s-T})ds + \int_T^t \sigma(X^2_{s-T})dW_s\\
&= X_0 + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dW_s.
\end{aligned}$$

By continuing this procedure, we can patch together solutions on the intervals $[(N-1)T, NT]$ for $N = 1, 2, \ldots$ to get a solution on $[0, \infty)$.

With the above goal in mind, we recall Banach's fixed point theorem, also called the contraction mapping theorem. We state it in the form relevant for us.

THEOREM. *Suppose B is a Banach space with norm $\|\cdot\|$, and $F : B \to B$ is such that*

$$\|F(x) - F(y)\| \leq c\|x - y\| \text{ for all } x, y \in B$$

for some constant $0 < c < 1$. Then there exists a unique fixed point $x^* \in B$ such that

$$F(x^*) = x^*.$$

PROOF OF STRONG EXISTENCE. Fix an intial condition X_0 and let F be defined by

$$F(Y)_t = X_0 + \int_0^t b(Y_s)ds + \int_0^t \sigma(Y_s)dW_s$$

for adapted continuous processes $Y = (Y_t)_{t\geq 0}$. Indeed, if Y^* is a fixed point of F then Y^* is a solution of the SDE.

The art of the proof of this theorem is to find a reasonable Banach space B of adapted processes on which to apply the Banach fixed point theorem. The specific choice of B is somewhat arbitrary, since no matter how we arrive at a solution, it must be unique by the strong uniqueness theorem from last lecture.

Now, we find a convenience Banach space. For an adapted continuous Y, let

$$\|Y\| = \mathbb{E}(\sup_{t\in[0,T]} \|Y_t\|^2)^{1/2}$$

for a non-random $T > 0$ to be determined later. We consider the vector space B defined by

$$B = \{Y : \|Y\| < \infty\}.$$

Then B is a Banach space, i.e. complete, for the norm. (The proof of this fact is similar to the proof that the space $\mathcal{M}^2$ of continuous square-integrable martingales is complete.) We now assume that X_0 is square-integrable without loss[1].

Now we show that F is a contraction:

$$\|F(X) - F(Y)\|^2 \leq 2\mathbb{E}\left(\sup_{t\in[0,T]}\left\|\int_0^t [b(X_s) - b(Y_s)]ds\right\|^2\right) + 2\mathbb{E}\left(\sup_{t\in[0,T]}\left\|\int_0^t [\sigma(X_s) - \sigma(Y_s)]dW_s\right\|^2\right)$$

Here we use the Lipschitz assumption combined with Jensen's inequality:

$$\begin{aligned}\mathbb{E}\left(\sup_{0\leq t\leq T}\left\|\int_0^t b(X_s) - b(Y_s)ds\right\|^2\right) &\leq T\mathbb{E}\left(\int_0^T \|b(X_s) - b(Y_s)\|^2 ds\right) \\ &\leq K^2T^2\ \mathbb{E}(\sup_{t\in[0,T]}\|X_t - Y_t\|^2)]\end{aligned}$$

and here we use Lipschitz assumption combined with Burkholder's inequality:

$$\begin{aligned}\mathbb{E}\left(\sup_{0\leq t\leq T}\left\|\int_0^t [\sigma(X_s) - \sigma(Y_s)]dW_s\right\|^2\right) &\leq 4\mathbb{E}\left(\int_0^T \|\sigma(X_s) - \sigma(Y_s)\|^2 ds\right) \\ &\leq 4K^2T\ \mathbb{E}(\sup_{t\in[0,T]}\|X_t - Y_t\|^2)]\end{aligned}$$

and hence

$$\|F(X) - F(Y)\|^2 \leq (2T + 8)K^2T\ \|X - Y\|^2.$$

So, if we choose T small enough that $c^2 = (2T + 8)K^2T < 1$, we can apply the Banach fixed point theorem as promised – once we check that the map F sends B to B. Clearly for any $X \in B$, the process $F(X)$ is adapted and continuous. Also, we have the triangle inequality,

$$\begin{aligned}\|F(X)\| &\leq \|F(X) - F(0)\| + \|F(0)\| \\ &\leq c\|X\| + \|F(0)\|\end{aligned}$$

so it is enough to show $F(0) \in B$. But

$$F(0)_t = X_0 + tb(0) + \sigma(0)W_t$$

which is easily seen to belong to B. □

[1]... since otherwise we could use the norm

$$\|Y\| = \mathbb{E}(e^{-\|X_0\|}\sup_{t\in[0,T]}\|Y_t\|^2)^{1/2}$$

instead...

4. Connection to partial differential equations

One of the most useful aspects of the theory of stochastic differential equations is their link to partial differential equations.

The main idea for this section is contained in this result:

THEOREM (Feynman–Kac formula, Kolmogorov equation). *Let X be a weak solution of the SDE*

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

and suppose $v : [0,T] \times \mathbb{R}^n \to \mathbb{R}$ is C^2, bounded and satisfies the PDE

$$\frac{\partial v}{\partial t} + \sum_i b^i \frac{\partial v}{\partial x^i} + \frac{1}{2} \sum_{i,j} a^{ij} \frac{\partial^2 v}{\partial x^i \partial x^j} = 0$$

where

$$a^{ij} = \sum_k \sigma^{ik} \sigma^{jk}$$

with terminal condition

$$v(T,x) = \phi(x) \text{ for all } x \in \mathbb{R}^n.$$

Then

$$v(t, X_t) = \mathbb{E}\left[\phi(X_T)|\mathcal{F}_t\right]$$

PROOF. Let $M_t = v(t, X_t)$ and note that by Itô's formula and the fact that v satisfies a certain PDE, we have

$$dM_t = \sum_{ij} \sigma^{ij} \frac{\partial v}{\partial x^i} dW^j.$$

Hence M is a local martingale. But since M is bounded by assumption, M is a true martingale. □

The main reason for studying stochastic differential equations is that solutions to SDEs provide examples of continuous time Markov processes.

DEFINITION. An adapted process X is called a Markov process in the filtration $\mathbb{F}$ iff

$$\mathbb{E}[\phi(X_t)|\mathcal{F}_s] = \mathbb{E}[\phi(X_t)|X_s]$$

for all bounded measurable functions $\phi : \mathbb{R}^n \to \mathbb{R}$ and $0 \leq s \leq t$.

Recall that if U and V are random variables such that V is measurable with respect to the sigma-field $\sigma(U)$ generated by U, then there exists a measurable function f such that $V = f(U)$. In particular, that for integrable ξ, the conditional expectation $\mathbb{E}(\xi|U)$ is defined to be a $\sigma(U)$-measurable random variable, and hence we have $\mathbb{E}(\xi|U) = f(U)$ for some f.

We the above comment is mind, if X is Markov, then for any bounded measurable ϕ and $0 \leq s \leq t$, there exists a function $P_{s,t}\phi$ such that

$$\mathbb{E}[\phi(X_t)|\mathcal{F}_s] = P_{s,t}\phi(X_s).$$

DEFINITION. The Markov process is called time-homogeneous if $P_{s,t}\phi = P_{0,t-s}\phi$ for all bounded measurable ϕ and $0 \leq s \leq t$.

PROPOSITION. *If X is a continuous Markov process, then the conditional law of a Markov process $(X_t)_{t\in[s,\infty)}$ given the whole history $\mathcal{F}_s$ only depends on X_s.*

PROOF. This argument is in essence the same as that on page 21.

Note that since X is continuous we need only show that its conditional finite-dimensional distributions only depend on X_s. Therefore, fixing $s = t_0 < t_1 \dots < t_k$ we consider the conditional joint characteristic function of $X_{t_1}, \dots, X_{t_k}$. Let $\theta_1, \dots, \theta_k \in \mathbb{R}^n$, and note by iterating expectations that we have

$$\begin{aligned}
\mathbb{E}[e^{i\theta_1 \cdot X_{t_1} + \dots + i\theta_k \cdot X_{t_k}} | \mathcal{F}_{t_0}] &= \mathbb{E}[e^{i\theta_1 \cdot X_{t_1} + \dots + i\theta_k \cdot X_{t_{k-1}}} \mathbb{E}(e^{i\theta_k \cdot X_{t_k}} | \mathcal{F}_{t_{k-1}}) | \mathcal{F}_{t_0}] \\
&= \mathbb{E}[e^{i\theta_1 \cdot X_{t_1} + \dots + i\theta_k \cdot X_{t_{k-1}}} g_{k-1}(X_{t_{k-1}}) | \mathcal{F}_{t_0}] \\
&= \dots \\
&= g_0(X_{t_0}) \\
&= \mathbb{E}[e^{i\theta_1 \cdot X_{t_1} + \dots + i\theta_k \cdot X_{t_k}} | X_{t_0}]
\end{aligned}$$

where g_j is the bounded measurable function defined recursively by $g_k = 1$ and

$$\mathbb{E}(e^{i\theta_j \cdot X_{t_j}} g_j(X_{t_j}) | X_{t_{j-1}}) = g_{j-1}(X_{t_{j-1}}).$$

□

For the rest of the section we fix some notation: We are concerned with the SDE

$$dX = b(X)dt + \sigma(X)dW, \quad X_0 = x \tag{SDE}$$

It turns out that there is a differential operator $\mathcal{L}$ associated to (*) defined by

$$\mathcal{L} = \sum_i b^i \frac{\partial}{\partial x^i} + \frac{1}{2} \sum_{i,j} a^{ij} \frac{\partial^2}{\partial x^i \partial x^j}.$$

where

$$a^{ij} = \sum_k \sigma^{ik} \sigma^{jk}$$

This operator $\mathcal{L}$ is called the *generator* of X, and is the continuous space analogue of the so-called Q-matrix from the theory of Markov processes on countable state spaces.

Finally, we are concerned with the PDE

$$\frac{\partial u}{\partial t} = \mathcal{L}u, \quad u(0, x) = \phi(x) \tag{PDE}$$

This PDE is usually called Kolmogorov's equation .

THEOREM. *Suppose the equation (SDE) has a weak solution X for every choice of initial condition $X_0 = x \in \mathbb{R}^n$. Furthermore, suppose the equation (PDE) has a bounded C^2 solution u for each bounded ϕ. Then equation (SDE) has the uniqueness in law property, the process X is Markov, and the equation (PDE) has a unique solution given by*

$$u(t, x) = \mathbb{E}\left[\phi(X_t) | X_0 = x\right]$$

PROOF. Fix a non-random $T > 0$ and a bounded ϕ. Let u be a bounded solution to (PDE) and set $v(s, x) = u(T - s, x)$. Note that

$$\frac{\partial v}{\partial s} + \mathcal{L}v = 0, \quad v(T, x) = \phi(x)$$

so by the theorem at the beginning of this section we have

$$v(s, X_s) = \mathbb{E}[\phi(X_T) | \mathcal{F}_s].$$

Since the left side is X_s measurable, we have

$$\mathbb{E}[\phi(X_T)|\mathcal{F}_s] = \mathbb{E}[\phi(X_T)|X_s].$$

Since this holds for all $0 \le s \le T$ and all bounded ϕ, we see that X is Markov.

Now setting $s = 0$, we have

$$u(T, X_0) = v(0, X_0) = \mathbb{E}[\phi(X_T)|X_0]$$

showing that the bounded solutions to equation (PDE) is uniquely determined by the initial condition ϕ.

Now, to see that uniqueness in law property of equation (SDE), we note as before, that by continuity the law of the solution is determined by its finite-dimensional distributions, and that these are determined by joint characteristic functions, which can be computed recursively as the *unique* solution of equation (PDE) with the appropriate boundary condition. □

Remark. Notice the interplay of existence and uniqueness above. The existence of the solution of the SDE implies the uniqueness of the solution of the PDE, since if there was a solution to the PDE, it would have to be given by the specific formula. On the other hand, the existence of the solution of the PDE implies the uniqueness of law of the SDE since the law of the solution of the SDE is characterised completely by the solutions of the PDE.

The above theorem has a converse:

Theorem. *Suppose X is a weak solution of (SDE) and is a time-homogeneous Markov process such that $\mathbb{P}(X_t \in A|X_0 = x) > 0$ for $t > 0$, any open set A and any initial condition $x \in \mathbb{R}^n$. Let*

$$u(t, x) = \mathbb{E}\left[\phi(X_t)|X_0 = x\right].$$

for bounded ϕ. If u is C^2, then u satisfies equation (PDE).

Proof. Fix a non-random $T > 0$, and let $v(s, x) = u(T - s, x)$ as before. Now

$$\begin{aligned} v(s, x) &= \mathbb{E}\left[\phi(X_{T-s})|X_0 = x\right] \\ &= \mathbb{E}\left[\phi(X_T)|X_s = x\right] \end{aligned}$$

by time-homogeneity. Hence

$$v(s, X_s) = \mathbb{E}\left[\phi(X_T)|\mathcal{F}_s\right]$$

by the Markov property. Hence $M = v(s, X_s)$ defines a martingale. Finally, since v is C^2 by assumption we can apply Itô's formula:

$$dM_s = \left(\frac{\partial v}{\partial s} + \mathcal{L}v\right) ds + \sum_{ij} \sigma^{ij} \frac{\partial}{\partial x^i} dW^j.$$

Since M is a martingale, the drift term must vanish for Leb $\times$ $\mathbb{P}$-almost every (t, ω) by the uniqueness of the semimartingale decomposition. The conclusion follows from the assumption that the derivatives of the v are continuous and that X visits every point with positive probability. □

There are lots of ways to generalise the above discussion. For instance, THEOREM. *Suppose X is a weak solution of*

$$dX = b(X)dt + \sigma(X)dW$$

and that u is C^2 solution of

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

where ϕ, g, f are bounded. Suppose that $\sup_{0\le t\le T}|u(t,x)|$ is bounded for each $T > 0$.

Then

$$u(t,x) = \mathbb{E}\left[e^{\int_0^t g(X_s)ds}\phi(X_t) + \int_0^t e^{\int_0^t g(X_u)du} f(X_s)ds | X_0 = x\right]$$

4.1. Formulation in terms of the transition density. We will now consider the density formulation of the Kolmogorov equations. That is, we study the equations that arise when the Markov process X has a density $p(t,x,y)$ with respect to Lebesgue measure, i.e.

$$\mathbb{E}[\phi(X_t)|X_0 = x] = \int \phi(y)p(t,x,y)dy$$

In this section, we will use the word *claim* for mathematical statements which are not true in general, but for which there exists a non-empty but unspecified set of hypotheses such that the statement can be proven. And rather than proofs, we include *ideas*, which involve some formal manipulations without proper justification. The claims can be turned into theorems by identifying conditions sufficient to justify these manipulations.

CLAIM (Kolmogorov's backward equation). *The transition density, if it exists, should satisfy the PDE*

$$\frac{\partial p}{\partial t} = \mathcal{L}_x p$$

with initial condition

$$p(0,x,y) = \delta(x-y).$$

IDEA. Fix a ϕ and let

$$\mathbb{E}[\phi(X_t)|X_0 = x] = u(t,x).$$

In the previous section, we showed that if u is smooth enough then

$$\frac{\partial u}{\partial t} = \mathcal{L}u.$$

Writing u in terms of the transition density, and supposing that we can interchange integration and differentiation, we have

$$\int \frac{\partial p}{\partial t}(t,x,y)\phi(y)dy = \int \mathcal{L}_x p(t,x,y)\phi(y)dy.$$

Since ϕ is arbitrary, we must have that p satisfies the PDE. $\square$

Kolmogorov's backward equation is simply

$$\frac{dP_t}{dt} = \mathcal{L}P_t$$

where $(P_t)_{t\geq 0}$ is the transition semigroup of X, i.e. where

$$(P_t f)(x) = \mathbb{E}[f(X_t)|X_0 = x].$$

Formally speaking, the solution of this (operator valued) ODE is given by

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

Hence, we should expect the semigroup to satisfy Kolmogorov's forward equation

$$\frac{dP_t}{dt} = P_t\mathcal{L}$$

as well. To formulate this equation in terms of the transition density, we first need to define some notation.

DEFINITION. The formal adjoint of $\mathcal{L}$ is the second order partial differential operator $\mathcal{L}^*$ defined by

$$\mathcal{L}^*\phi = -\sum_i \frac{\partial}{\partial x^i}(b^i\phi) + \frac{1}{2}\sum_{i,j}\frac{\partial^2}{\partial x^i \partial x^j}(a^{ij}\phi).$$

REMARK. Let $\langle \phi, \psi\rangle_{L^2} = \int \phi(x)\psi(x)dx$ be the usual inner product on the space L^2 of square-integrable functions on $\mathbb{R}^n$. Note that if ϕ and ψ are smooth and compactly supported, then

$$\begin{aligned}
\langle \phi, \mathcal{L}^*\psi\rangle_{L^2} &= \int\left[-\phi(x)\sum_i \frac{\partial}{\partial x^i}(b^i\psi)(x) + \frac{1}{2}\phi(x)\sum_{i,j}\frac{\partial^2}{\partial x^i\partial x^j}(a^{ij}\psi)(x)\right]dx \\
&= \int\left[\psi(x)\sum_i b^i(x)\frac{\partial\phi}{\partial x^i}(x) + \frac{1}{2}\psi(x)\sum_{i,j} a^{ij}(x)\frac{\partial^2\phi}{\partial x^i\partial x^j}(x)\right]dx \\
&= \langle \mathcal{L}\phi, \psi\rangle_{L^2}
\end{aligned}$$

by integrating by parts.

CLAIM (Kolmogorov's forward equation/Fokker–Planck equation). *The transition density p should satisfy the PDE*

$$\frac{\partial p}{\partial t} = \mathcal{L}^*_y p.$$

with initial condition

$$p(0, x, y) = \delta(x - y).$$

IDEA. Suppose ϕ is smooth and compactly supported. By Itô's formula

$$\phi(X_t) = \phi(X_0) + \int_0^t (\mathcal{L}\phi)(X_s)ds + M_t$$

where M is a local martingale. Since M is bounded on bounded intervals, M is a true martingale and therefore

$$\begin{aligned}\int p(t,x,y)\phi(y)dy &= \phi(x) + \int_0^t \int p(s,x,y)\mathcal{L}_y\phi(y)dy \\ &= \phi(x) + \int_0^t \int \mathcal{L}_y^* p(s,x,y)f(y)dy\end{aligned}$$

by formal integration by parts. Now formally differentiate with respect to t. □

EXAMPLE. Consider the SDE

$$dX_t = dW_t, \quad X_0 = x$$

where W is a n-dimensional Brownian motion, so that $X_t = x + W_t$. Since the increments of X are normally distributed, it is easy to see that in this case the transition density is given by

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

Also, in this case, the generator is given by

$$\mathcal{L} = \frac{1}{2}\sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2} = \frac{1}{2}\Delta$$

where Δ is the Laplacian operator. Notice that $\mathcal{L} = \mathcal{L}^*$ in this case. You should check that the given density solves both of the corresponding Kolmogorov equations.

We now turn our attention to invariant measures. A probability measure μ on $\mathbb{R}^n$ is invariant for the Markov process defined by the SDE

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

if $X_0 \sim \mu$ implies $X_t \sim \mu$ for all $t \geq 0$.

CLAIM. *Suppose $\hat{p}$ is a probability density and satisfies*

$$\mathcal{L}^*\hat{p} = 0.$$

Then $\hat{p}$ is an invariant density.

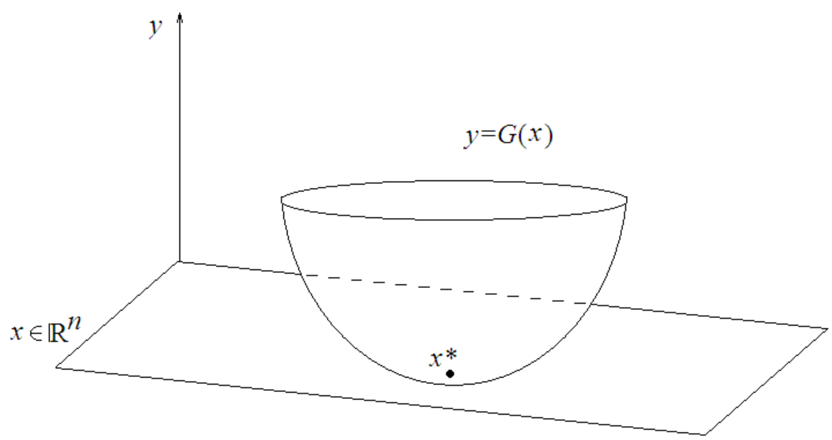
IDEA. Let

$$u(t,x) = \int \mathbb{E}[\phi(X_t)|X_0 = x]$$

and suppose u satisfies the Kolmogorov equation $\frac{\partial u}{\partial t} = \mathcal{L}u$. Then

$$\begin{aligned}\int u(t,x)\hat{p}(x)dx - \int \phi(x)\hat{p}(x)dx &= \int \int_0^t \frac{\partial u}{\partial t}(s,x)\hat{p}(x)ds\ ds \\ &= \int_0^t \int \mathcal{L}u(t,x)\hat{p}(x)dx \\ &= \int_0^t \int u(t,x)\mathcal{L}^*\hat{p}(x)dx \\ &= 0.\end{aligned}$$

by Fubini's theorem and integration by parts. □



Example. Consider the gradient system

$$dX_t = -\nabla G(X_t)dt + \varepsilon dW_t$$

where $n = d$, $b = -\nabla G$, $\sigma = \varepsilon I$ and $G : \mathbb{R}^n \mapsto \mathbb{R}$ is a smooth, strictly convex function taking its minimum at the unique point x^*.

In the case where $\varepsilon = 0$, one can show that for all initial condition $X_0 \in \mathbb{R}^n$, we have

$$X_t \to x^* \text{ as } t \to \infty.$$

What happens when $\varepsilon > 0$?

The generator in this case is

$$\mathcal{L} = -\nabla G \cdot \nabla + \frac{\varepsilon^2}{2}\Delta$$

One can then confirm that that an invariant density is

$$\hat{p}(x) = Ce^{-2G(x)/\varepsilon^2}$$

where $C > 0$ is the normalising constant.

In example sheet 3, you were asked to find a Gaussian invariant density for the Ornstein–Uhlenbeck process, corresponding to the case $G(x) = \|x\|^2/2$.

4.2. An application of stochastic analysis to PDEs. This section is not examinable. Our goal is to see a probablistic approach to finding sufficient conditions on the coefficients b and σ such that the PDE

$$\frac{\partial u}{\partial t} = \mathcal{L}u, \quad u(0,\cdot) = \phi$$

has a solution. We have seen that if

$$u(t,x) = \mathbb{E}[\phi(X_t)|X_0 = x]$$

and *if u is smooth*, then u is a solution.

As a warm-up, suppose that $n = 1$, that $b = 0$, that $\sigma = 1$ and hence $X_t = X_0 + W_t$. In this case

$$u(t,x) = \int \phi(x + \sqrt{t}z)\frac{e^{-z^2/2}}{\sqrt{2\pi}}dz$$

Assuming ϕ is smooth with all derivatives bound, we have by integration by parts

$$\begin{aligned}\frac{\partial u}{\partial t}(t,x) &= \int \phi'(x + \sqrt{t}z)\frac{e^{-z^2/2}}{\sqrt{2\pi}}dz \\ &= \int \phi(x + \sqrt{t}z)\frac{z}{\sqrt{t}}\frac{e^{-z^2/2}}{\sqrt{2\pi}}dz \\ &= \mathbb{E}\left[\phi(x + W_t)\frac{W_t}{t}\right].\end{aligned}$$

The important point is that even if ϕ is not smooth, the last equation makes sense whenever $t > 0$.

We will suppose that b and σ are smooth with bounded derivatives. This implies that the SDE has a strong solution. We will also suppose that σ is bounded from below. What we will see is that if

$$u(t,x) = \mathbb{E}[\phi(X^x_t)]$$

where X^x is the SDE started at $X_0 = x$, then for each (t,x) there exists an integrable random variable π^x_t such that

$$\frac{\partial u}{\partial t}(t,x) = \mathbb{E}[\phi(X^x_t)\pi^x_t]$$

with similar statements for $n > 1$. In particular, the function $u(t,\cdot)$ is differentiable even is the bounded function ϕ is not differentiable.

To begin to see how to prove such a statement, we very briefly discuss Malliavin calculus. Our setting is a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ supporting a Brownian motion W which generates the filtration.

Definition. A random variable ξ is called *smooth* if there is smooth function $F : \mathbb{R}^k \to \mathbb{R}$ with all bounded derivative and $0 \le t_0 < \ldots < t_k$ such that

$$\xi = F(W_{t_1} - W_{t_0}, \ldots, W_{t_k} - W_{t_{k-1}}).$$

Definition. The Malliavin derivative of a smooth random variable is a process $D\xi$ defined by

$$D_t\xi = \sum_{i=1}^{k} \frac{\partial F}{\partial x_i}(W_{t_1} - W_{t_0}, \ldots, W_{t_k} - W_{t_{k-1}})\mathbb{1}_{(t_{i-1},t_i]}(t)$$

The key property of the Malliavin derivative is the following integration by parts formula:

THEOREM. *Suppose ξ is smooth and α is a simple predictable process. Then*

$$\mathbb{E}\left[\int_0^\infty (D_t\xi)\alpha_t\ dt\right] = \mathbb{E}\left[\xi\int_0^\infty \alpha_t\ dW_t\right].$$

PROOF. Suppose

$$\alpha = \sum_{i=1}^k K_i \mathbb{1}_{(t_{i-1},t_i]}$$

where K_i is bounded and $\mathcal{F}_{t_{i-1}}$-measurable. Hence

$$\begin{aligned}\mathbb{E}\left[\int_0^\infty (D_t\xi)\alpha_t\ dt\right] &= \mathbb{E}\sum_{i=1}^k \frac{\partial F}{\partial x_i}(W_{t_1}-W_{t_0},\ldots,W_{t_k}-W_{t_{k-1}})K_{t_i}(t_i-t_{i-1}) \\ &= \mathbb{E}\sum_{i=1}^k F(W_{t_1}-W_{t_0},\ldots,W_{t_k}-W_{t_{k-1}})K_{t_i}(W_{t_i}-W_{t_{i-1}}) \\ &= \mathbb{E}\left[\xi\int_0^\infty \alpha_t\ dW_t\right].\end{aligned}$$

where we have used the Gaussian integration by parts formula. □

It is the above integration by parts formula is what allows us to extend the definition of the Malliavin derivative to a larger class of random variables. We skip the details, but the idea is that this extended definition has the convenient property that

$$\mathbb{E}\left[\int_0^\infty (D_t\xi)\alpha_t\ dt\right] = \mathbb{E}\left[\xi\int_0^\infty \alpha_t\ dW_t\right].$$

holds for all ξ in the domain $\mathbb{D}$ of the derivative operator D and all $\alpha \in L^2(W)$.

Now we turn our attention to the solutions of the SDE:

$$X_t^x = x + \int_0^t b(X_s^x)ds + \int_0^t \sigma(X_s^x)dW_s.$$

Let

$$Y_t^x = \frac{\partial X_t^x}{\partial x}.$$

Differentiating the integral equation (assuming integration and differentiation can be interchanged) yields

$$Y_t^x = 1 + \int_0^t b'(X_s^x)Y_s^x\ ds + \int_0^t \sigma'(X_s^x)Y_s^x\ dW_s.$$

On the other hand, computing the Malliavin derivative of both sides of the integral equation yields

$$D_tX_T^x = \int_t^T b'(X_s^x)D_tX_s^x\ ds + \int_t^t \sigma(X_s^x)D_tX_s^x\ dW_s + \sigma(X_t^x)$$

for $0 \le t \le T$, from which we have

$$D_tX_T^x = \frac{Y_T^x}{Y_t^x}\sigma(X_t^x).$$

In particular, we have the identity

$$\begin{aligned} Y_T^x &= \frac{Y_t^x}{\sigma(X_t^x)} D_t X_T^x \text{ for all } 0 \le t \le T \\ &= \frac{1}{T} \int_0^T \frac{Y_t^x}{\sigma(X_t^x)} D_t X_T^x dt. \end{aligned}$$

$$\begin{aligned} \frac{\partial u}{\partial t}(t,x) &= \mathbb{E}\left[\phi'(X_t^x) Y_t^x\right] \\ &= \mathbb{E}\left[\frac{1}{t}\int_0^t \phi'(X_t^x) D_s X_t^x \frac{Y_s^x}{\sigma(X_s^x)} ds\right] \\ &= \mathbb{E}\left[\frac{1}{t}\int_0^t D_s \phi(X_t^x) \frac{Y_s^x}{\sigma(X_s^x)} ds\right] \\ &= \mathbb{E}\left[\phi(X_t^x)\pi_t^x\right] \end{aligned}$$

where

$$\pi_t^x = \frac{1}{t}\int_0^t \frac{Y_s^x}{\sigma(X_s^x)} dW_s.$$

This is called the Bismut–Elworthy–Li formula. This little calculation is just the beginning of an interesting story.

Index