

Part III — Stochastic Calculus and Applications

Based on lectures by R. Bauerschmidt

Notes taken by Dexter Chua

Lent 2018

These notes are not endorsed by the lecturers, and I have modified them (often significantly) after lectures. They are nowhere near accurate representations of what was actually lectured, and in particular, all errors are almost surely mine.

- *Brownian motion*. Existence and sample path properties.
- *Stochastic calculus for continuous processes*. Martingales, local martingales, semi-martingales, quadratic variation and cross-variation, Itô's isometry, definition of the stochastic integral, Kunita–Watanabe theorem, and Itô's formula.
- *Applications to Brownian motion and martingales*. Lévy characterization of Brownian motion, Dubins–Schwartz theorem, martingale representation, Girsanov theorem, conformal invariance of planar Brownian motion, and Dirichlet problems.
- *Stochastic differential equations*. Strong and weak solutions, notions of existence and uniqueness, Yamada–Watanabe theorem, strong Markov property, and relation to second order partial differential equations.

Pre-requisites

Knowledge of measure theoretic probability as taught in Part III Advanced Probability will be assumed, in particular familiarity with discrete-time martingales and Brownian motion.

Contents

0	Introduction	3
1	The Lebesgue–Stieltjes integral	6
2	Semi-martingales	9
2.1	Finite variation processes	9
2.2	Local martingale	11
2.3	Square integrable martingales	15
2.4	Quadratic variation	17
2.5	Covariation	22
2.6	Semi-martingale	24
3	The stochastic integral	25
3.1	Simple processes	25
3.2	Itô isometry	26
3.3	Extension to local martingales	28
3.4	Extension to semi-martingales	30
3.5	Itô formula	32
3.6	The Lévy characterization	35
3.7	Girsanov’s theorem	37
4	Stochastic differential equations	41
4.1	Existence and uniqueness of solutions	41
4.2	Examples of stochastic differential equations	45
4.3	Representations of solutions to PDEs	47
	Index	51

0 Introduction

Ordinary differential equations are central in analysis. The simplest class of equations tend to look like

$$\dot{x}(t) = F(x(t)).$$

Stochastic differential equations are differential equations where we make the function F “random”. There are many ways of doing so, and the simplest way is to write it as

$$\dot{x}(t) = F(x(t)) + \eta(t),$$

where η is a random function. For example, when modeling noisy physical systems, our physical bodies will be subject to random noise. What should we expect the function η to be like? We might expect that for $|t - s| \gg 0$, the variables $\eta(t)$ and $\eta(s)$ are “essentially” independent. If we are interested in physical systems, then this is a rather reasonable assumption, since random noise is random!

In practice, we work with the idealization, where we claim that $\eta(t)$ and $\eta(s)$ are independent for $t \neq s$. Such an η exists, and is known as *white noise*. However, it is not a function, but just a Schwartz distribution.

To understand the simplest case, we set $F = 0$. We then have the equation

$$\dot{x} = \eta.$$

We can write this in integral form as

$$x(t) = x(0) + \int_0^t \eta(s) ds.$$

To make sense of this integral, the function η should at least be a signed measure. Unfortunately, white noise isn’t. This is bad news.

We ignore this issue for a little bit, and proceed as if it made sense. If the equation held, then for any $0 = t_0 < t_1 < \dots$, the increments

$$x(t_i) - x(t_{i-1}) = \int_{t_{i-1}}^{t_i} \eta(s) ds$$

should be independent, and moreover their variance should scale linearly with $|t_i - t_{i-1}|$. So maybe this x should be a Brownian motion!

Formalizing these ideas will take up a large portion of the course, and the work isn’t always pleasant. Then why should we be interested in this continuous problem, as opposed to what we obtain when we discretize time? It turns out in some sense the continuous problem is easier. When we learn measure theory, there is a lot of work put into constructing the Lebesgue measure, as opposed to the sum, which we can just define. However, what we end up is much easier — it’s easier to integrate $\frac{1}{x^3}$ than to sum $\sum_{n=1}^{\infty} \frac{1}{n^3}$. Similarly, once we have set up the machinery of stochastic calculus, we have a powerful tool to do explicit computations, which is usually harder in the discrete world.

Another reason to study stochastic calculus is that a lot of continuous time processes can be described as solutions to stochastic differential equations. Compare this with the fact that functions such as trigonometric and Bessel functions are described as solutions to ordinary differential equations!

There are two ways to approach stochastic calculus, namely via the Itô integral and the Stratonovich integral. We will mostly focus on the Itô integral, which is more useful for our purposes. In particular, the Itô integral tends to give us martingales, which is useful.

To give a flavour of the construction of the Itô integral, we consider a simpler scenario of the Wiener integral.

Definition (Gaussian space). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Then a subspace $S \subseteq L^2(\Omega, \mathcal{F}, \mathbb{P})$ is called a *Gaussian space* if it is a closed linear subspace and every $X \in S$ is a centered Gaussian random variable.

An important construction is

Proposition. Let H be any separable Hilbert space. Then there is a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a Gaussian subspace $S \subseteq L^2(\Omega, \mathcal{F}, \mathbb{P})$ and an isometry $I : H \rightarrow S$. In other words, for any $f \in H$, there is a corresponding random variable $I(f) \sim N(0, (f, f)_H)$. Moreover, $I(\alpha f + \beta g) = \alpha I(f) + \beta I(g)$ and $(f, g)_H = \mathbb{E}[I(f)I(g)]$.

Proof. By separability, we can pick a Hilbert space basis $(e_i)_{i=1}^\infty$ of H . Let $(\Omega, \mathcal{F}, \mathbb{P})$ be any probability space that carries an infinite independent sequence of standard Gaussian random variables $X_i \sim N(0, 1)$. Then send e_i to X_i , extend by linearity and continuity, and take S to be the image. \square

In particular, we can take $H = L^2(\mathbb{R}_+)$.

Definition (Gaussian white noise). A *Gaussian white noise* on \mathbb{R}_+ is an isometry WN from $L^2(\mathbb{R}_+)$ into some Gaussian space. For $A \subseteq \mathbb{R}_+$, we write $WN(A) = WN(\mathbf{1}_A)$.

Proposition.

- For $A \subseteq \mathbb{R}_+$ with $|A| < \infty$, $WN(A) \sim N(0, |A|)$.
- For disjoint $A, B \subseteq \mathbb{R}_+$, the variables $WN(A)$ and $WN(B)$ are independent.
- If $A = \bigcup_{i=1}^\infty A_i$ for disjoint sets $A_i \subseteq \mathbb{R}_+$, with $|A| < \infty, |A_i| < \infty$, then

$$WN(A) = \sum_{i=1}^\infty WN(A_i) \text{ in } L^2 \text{ and a.s.}$$

Proof. Only the last point requires proof. Observe that the partial sum

$$M_n = \sum_{i=1}^n WN(A)$$

is a martingale, and is bounded in L^2 as well, since

$$\mathbb{E}M_n^2 = \sum_{i=1}^n \mathbb{E}WN(A_i)^2 = \sum_{i=1}^n |A_i| \leq |A|.$$

So we are done by the martingale convergence theorem. The limit is indeed $WN(A)$ because $\mathbf{1}_A = \sum_{n=1}^\infty \mathbf{1}_{A_n}$. \square

The point of the proposition is that WN really looks like a random measure on \mathbb{R}_+ , except it is *not*. We only have convergence almost surely above, which means we have convergence on a set of measure 1. However, the set depends on which A and A_i we pick. For things to actually work out well, we must have a fixed set of measure 1 for which convergence holds for all A and A_i .

But perhaps we can ignore this problem, and try to proceed. We define

$$B_t = WN([0, t])$$

for $t \geq 0$.

Exercise. This B_t is a standard Brownian motion, except for the continuity requirement. In other words, for any t_1, t_2, \dots, t_n , the vector $(B_{t_i})_{i=1}^n$ is jointly Gaussian with

$$\mathbb{E}[B_s B_t] = s \wedge t \text{ for } s, t \geq 0.$$

Moreover, $B_0 = 0$ a.s. and $B_t - B_s$ is independent of $\sigma(B_r : r \leq s)$. Moreover, $B_t - B_s \sim N(0, t - s)$ for $t \geq s$.

In fact, by picking a good basis of $L^2(\mathbb{R}_+)$, we can make B_t continuous.

We can now try to define some stochastic integral. If $f \in L^2(\mathbb{R}_+)$ is a step function,

$$f = \sum_{i=1}^n f_i \mathbf{1}_{[s_i, t_i]}$$

with $s_i < t_i$, then

$$WN(f) = \sum_{i=1}^n f_i (B_{t_i} - B_{s_i})$$

This motivates the notation

$$WN(f) = \int f(s) dB_s.$$

However, extending this to a function that is not a step function would be problematic.

1 The Lebesgue–Stieltjes integral

In calculus, we are able to perform integrals more exciting than simply $\int_0^1 h(x) dx$. In particular, if $h, a : [0, 1] \rightarrow \mathbb{R}$ are C^1 functions, we can perform integrals of the form

$$\int_0^1 h(x) da(x).$$

For them, it is easy to make sense of what this means — it’s simply

$$\int_0^1 h(x) da = \int_0^1 h(x)a'(x) dx.$$

In our world, we wouldn’t expect our functions to be differentiable, so this is not a great definition. One reasonable strategy to make sense of this is to come up with a measure that should equal “da”.

An immediate difficulty we encounter is that $a'(x)$ need not be positive all the time. So for example, $\int_0^1 1 da$ could be a negative number, which one wouldn’t expect for a usual measure! Thus, we are naturally lead to think about *signed measures*.

From now on, we always use the Borel σ -algebra on $[0, T]$ unless otherwise specified.

Definition (Signed measure). A *signed measure* on $[0, T]$ is a difference $\mu = \mu_+ - \mu_-$ of two positive measures on $[0, T]$ of disjoint support. The decomposition $\mu = \mu_+ - \mu_-$ is called the *Hahn decomposition*.

In general, given two measures μ_1 and μ_2 with not necessarily disjoint supports, we may still want to talk about $\mu_1 - \mu_2$.

Theorem. For any two finite measures μ_1, μ_2 , there is a signed measure μ with $\mu(A) = \mu_1(A) - \mu_2(A)$.

If μ_1 and μ_2 are given by densities f_1, f_2 , then we can simply decompose μ as $(f_1 - f_2)^+ dt + (f_1 - f_2)^- dt$, where $+$ and $-$ denote the positive and negative parts respectively. In general, they need not be given by densities with respect to dx , but they are always given by densities with respect to some other measure.

Proof. Let $\nu = \mu_1 + \mu_2$. By Radon–Nikodym, there are positive functions f_1, f_2 such that $\mu_i(dt) = f_i(t)\nu(dt)$. Then

$$(\mu_1 - \mu_2)(dt) = (f_1 - f_2)^+(t) \cdot \nu(dt) + (f_1 - f_2)^-(t) \cdot \nu(dt). \quad \square$$

Definition (Total variation). The total variation of a signed measure $\mu = \mu_+ - \mu_-$ is $|\mu| = \mu_+ + \mu_-$.

We now want to figure out how we can go from a function to a signed measure. Let’s think about how one would attempt to define $\int_0^1 f(x) dg$ as a Riemann sum. A natural option would be to write something like

$$\int_0^t h(s) da(s) = \lim_{m \rightarrow \infty} \sum_{i=1}^{n_m} h(t_{i-1}^{(m)}) (a(t_i^{(m)}) - a(t_{i-1}^{(m)}))$$

for any sequence of subdivisions $0 = t_0^{(m)} < \dots < t_{n_m}^{(m)} = t$ of $[0, t]$ with $\max_i |t_i^{(m)} - t_{i-1}^{(m)}| \rightarrow 0$.

In particular, since we want the integral of $h = 1$ to be well-behaved, the sum $\sum (a(t_i^{(m)}) - a(t_{i-1}^{(m)}))$ must be well-behaved. This leads to the notion of

Definition (Total variation). The *total variation* of a function $a : [0, T] \rightarrow \mathbb{R}$ is

$$V_a(t) = |a(0)| + \sup \left\{ \sum_{i=1}^n |a(t_i) - a(t_{i-1})| : 0 = t_0 < t_1 < \dots < t_n = T \right\}.$$

We say a has *bounded variation* if $V_a(T) < \infty$. In this case, we write $a \in BV$.

We include the $|a(0)|$ term because we want to pretend a is defined on all of \mathbb{R} with $a(t) = 0$ for $t < 0$.

We also define

Definition (Càdlàg). A function $a : [0, T] \rightarrow \mathbb{R}$ is *càdlàg* if it is right-continuous and has left-limits.

The following theorem is then clear:

Theorem. There is a bijection

$$\left\{ \text{signed measures on } [0, T] \right\} \longleftrightarrow \left\{ \begin{array}{l} \text{càdlàg functions of bounded} \\ \text{variation } a : [0, T] \rightarrow \mathbb{R} \end{array} \right\}$$

that sends a signed measure μ to $a(t) = \mu([0, t])$. To construct the inverse, given a , we define

$$a_{\pm} = \frac{1}{2}(V_a \pm a).$$

Then a_{\pm} are both positive, and $a = a_+ - a_-$. We can then define μ_{\pm} by

$$\begin{aligned} \mu_{\pm}[0, t] &= a_{\pm}(t) - a_{\pm}(0) \\ \mu &= \mu_+ - \mu_- \end{aligned}$$

Moreover, $V_{\mu[0, t]} = |\mu|[0, t]$.

Example. Let $a : [0, 1] \rightarrow \mathbb{R}$ be given by

$$a(t) = \begin{cases} 1 & t < \frac{1}{2} \\ 0 & t \geq \frac{1}{2} \end{cases}.$$

This is càdlàg, and its total variation is $v_0(1) = 2$. The associated signed measure is

$$\mu = \delta_0 - \delta_{1/2},$$

and the total variation measure is

$$|\mu| = \delta_0 + \delta_{1/2}.$$

We are now ready to define the Lebesgue–Stieltjes integral.

Definition (Lebesgue–Stieltjes integral). Let $a : [0, T] \rightarrow \mathbb{R}$ be càdlàg of bounded variation and let μ be the associated signed measure. Then for $h \in L^1([0, T], |\mu|)$, the *Lebesgue–Stieltjes integral* is defined by

$$\int_s^t h(r) da(r) = \int_{(s,t]} h(r)\mu(dr),$$

where $0 \leq s \leq t \leq T$, and

$$\int_s^t h(r) |da(r)| = \int_{(s,t]} h(r)|\mu|(dr).$$

We also write

$$h \cdot a(t) = \int_0^t h(r) da(r).$$

To let $T = \infty$, we need the following notation:

Definition (Finite variation). A càdlàg function $a : [0, \infty) \rightarrow \mathbb{R}$ is of finite variation if $a|_{[0,T]} \in BV[0, 1]$ for all $T > 0$.

Fact. Let $a : [0, T] \rightarrow \mathbb{R}$ be càdlàg and BV, and $h \in L^1([0, T], |da|)$, then

$$\left| \int_0^T h(s) da(s) \right| \leq \int_a^b |h(s)| |da(s)|,$$

and the function $h \cdot a : [0, T] \rightarrow \mathbb{R}$ is càdlàg and BV with associated signed measure $h(s) da(s)$. Moreover, $|h(s) da(s)| = |h(s)| |da(s)|$.

We can, unsurprisingly, characterize the Lebesgue–Stieltjes integral by a Riemann sum:

Proposition. Let a be càdlàg and BV on $[0, t]$, and h bounded and left-continuous. Then

$$\begin{aligned} \int_0^t h(s) da(s) &= \lim_{m \rightarrow \infty} \sum_{i=1}^{n_m} h(t_{i-1}^{(m)}) (a(t_i^{(m)}) - a(t_{i-1}^{(m)})) \\ \int_0^t h(s) |da(s)| &= \lim_{m \rightarrow \infty} \sum_{i=1}^{n_m} h(t_{i-1}^{(m)}) |a(t_i^{(m)}) - a(t_{i-1}^{(m)})| \end{aligned}$$

for any sequence of subdivisions $0 = t_0^{(m)} < \dots < t_{n_m}^{(m)} = t$ of $[0, t]$ with $\max_i |t_i^{(m)} - t_{i-1}^{(m)}| \rightarrow 0$.

Proof. We approximate h by h_m defined by

$$h_m(0) = 0, \quad h_m(s) = h(t_{i-1}^{(m)}) \text{ for } s \in (t_{i-1}^{(m)}, t_i^{(m)}].$$

Then by left continuity, we have

$$h(s) = \lim_{n \rightarrow \infty} h_m(s)$$

by left continuity, and moreover

$$\lim_{m \rightarrow \infty} \sum_{i=1}^{n_m} h(t_{i-1}^{(m)}) (a(t_i^{(m)}) - a(t_{i-1}^{(m)})) = \lim_{m \rightarrow \infty} \int_{(0,t]} h_m(s)\mu(ds) = \int_{(0,t]} h(s)\mu(ds)$$

by dominated convergence theorem. The statement about $|da(s)|$ is left as an exercise. \square

2 Semi-martingales

The title of the chapter is “semi-martingales”, but we are not going even meet the definition of a semi-martingale till the end of the chapter. The reason is that a semi-martingale is essentially defined to be the sum of a (local) martingale and a finite variation process, and understanding semi-martingales mostly involves understanding the two parts separately. Thus, for most of the chapter, we will be studying local martingales (finite variation processes are rather more boring), and at the end we will put them together to say a word or two about semi-martingales.

From now on, $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ will be a filtered probability space. Recall the following definition:

Definition (Càdlàg adapted process). A *càdlàg adapted process* is a map $X : \Omega \times [0, \infty) \rightarrow \mathbb{R}$ such that

- (i) X is càdlàg, i.e. $X(\omega, \cdot) : [0, \infty) \rightarrow \mathbb{R}$ is càdlàg for all $\omega \in \Omega$.
- (ii) X is adapted, i.e. $X_t = X(\cdot, t)$ is \mathcal{F}_t -measurable for every $t \geq 0$.

Notation. We will write $X \in \mathcal{G}$ to denote that a random variable X is measurable with respect to a σ -algebra \mathcal{G} .

2.1 Finite variation processes

The definition of a finite variation function extends immediately to a finite variation process.

Definition (Finite variation process). A *finite variation process* is a càdlàg adapted process A such that $A(\omega, \cdot) : [0, \infty) \rightarrow \mathbb{R}$ has finite variation for all $\omega \in \Omega$. The *total variation process* V of a finite variation process A is

$$V_t = \int_0^t |dA_s|.$$

Proposition. The total variation process V of a càdlàg adapted process A is also càdlàg, finite variation and adapted, and it is also increasing.

Proof. We only have to check that it is adapted. But that follows directly from our previous expression of the integral as the limit of a sum. Indeed, let $0 = t_0^{(m)} < t_1^{(m)} < \dots < t_{n_m}^{(m)} = t$ be a (nested) sequence of subdivisions of $[0, t]$ with $\max_i |t_i^{(m)} - t_{i-1}^{(m)}| \rightarrow 0$. We have seen

$$V_t = \lim_{m \rightarrow \infty} \sum_{i=1}^{n_m} |A_{t_i^{(m)}} - A_{t_{i-1}^{(m)}}| + |A(0)| \in \mathcal{F}_t. \quad \square$$

Definition $((H \cdot A)_t)$. Let A be a finite variation process and H a process such that for all $\omega \in \Omega$ and $t \geq 0$,

$$\int_0^t H_s(\omega) |dA_s(\omega)| < \infty.$$

Then define a process $((H \cdot A)_t)_{t \geq 0}$ by

$$(H \cdot A)_t = \int_0^t H_s dA_s.$$

For the process $H \cdot A$ to be adapted, we need a condition.

Definition (Previsible process). A process $H : \Omega \times [0, \infty) \rightarrow \mathbb{R}$ is *previsible* if it is measurable with respect to the *previsible σ -algebra* \mathcal{P} generated by the sets $E \times (s, t]$, where $E \in \mathcal{F}_s$ and $s < t$. We call the generating set Π .

Very roughly, the idea is that a previsible event is one where whenever it happens, you know it a finite (though possibly arbitrarily small) before.

Definition (Simple process). A process $H : \Omega \times [0, \infty) \rightarrow \mathbb{R}$ is *simple*, written $H \in \mathcal{E}$, if

$$H(\omega, t) = \sum_{i=1}^n H_{i-1}(\omega) \mathbf{1}_{(t_{i-1}, t_i]}(t)$$

for random variables $H_{i-1} \in \mathcal{F}_{i-1}$ and $0 = t_0 < \dots < t_n$.

Fact. Simple processes and their limits are previsible.

Fact. Let X be a càdlàg adapted process. Then $H_t = X_{t-}$ defines a left-continuous process and is previsible.

In particular, continuous processes are previsible.

Proof. Since X is càdlàg adapted, it is clear that H is left-continuous and adapted. Since H is left-continuous, it is approximated by simple processes. Indeed, let

$$H_t^n = \sum_{i=1}^{2^n} H_{(i-1)2^{-n}} \mathbf{1}_{((i-1)2^{-n}, i2^{-n}]}(t) \wedge n \in \mathcal{E}.$$

Then $H_t^n \rightarrow H$ for all t by left continuity, and previsibility follows. \square

Exercise. Let H be previsible. Then

$$H_t \in \mathcal{F}_{t-} = \sigma(\mathcal{F}_s : s < t).$$

Example. Brownian motion is previsible (since it is continuous).

Example. A Poisson process (N_t) is not previsible since $N_t \notin \mathcal{F}_{t-}$.

Proposition. Let A be a finite variation process, and H previsible such that

$$\int_0^t |H(\omega, s)| |dA(\omega, s)| < \infty \text{ for all } (\omega, t) \in \Omega \times [0, \infty).$$

Then $H \cdot A$ is a finite variation process.

Proof. The finite variation and càdlàg parts follow directly from the deterministic versions. We only have to check that $H \cdot A$ is adapted, i.e. $(H \cdot A)(\cdot, t) \in \mathcal{F}_t$ for all $t \geq 0$.

First, $H \cdot A$ is adapted if $H(\omega, s) = \mathbf{1}_{(u, v]}(s) \mathbf{1}_E(\omega)$ for some $u < v$ and $E \in \mathcal{F}_u$, since

$$(H \cdot A)(\omega, t) = \mathbf{1}_E(\omega) (A(\omega, t \wedge v) - A(\omega, t \wedge u)) \in \mathcal{F}_t.$$

Thus, $H \cdot A$ is adapted for $H = \mathbf{1}_F$ when $F \in \Pi$. Clearly, Π is a π system, i.e. it is closed under intersections and non-empty, and by definition it generates the

previsible σ -algebra \mathcal{P} . So to extend the adaptedness of $H \cdot A$ to all previsible H , we use the monotone class theorem.

We let

$$\mathcal{V} = \{H : \Omega \times [0, \infty) \rightarrow \mathbb{R} : H \cdot A \text{ is adapted}\}.$$

Then

- (i) $1 \in \mathcal{V}$
- (ii) $1_F \in \mathcal{V}$ for all $F \in \Pi$.
- (iii) \mathcal{V} is closed under monotone limits.

So \mathcal{V} contains all bounded \mathcal{P} -measurable functions. \square

So the conclusion is that if A is a finite variation process, then as long as reasonable finiteness conditions are satisfied, we can integrate functions against dA . Moreover, this integral was easy to define, and it obeys all expected properties such as dominated convergence, since ultimately, it is just an integral in the usual measure-theoretic sense. This crucially depends on the fact that A is a finite variation process.

However, in our motivating example, we wanted to take A to be Brownian motion, which is *not* of finite variation. The work we will do in this chapter and the next is to come up with a stochastic integral where we let A be a martingale instead. The heuristic idea is that while martingales can vary wildly, the martingale property implies there will be some large cancellation between the up and down movements, which leads to the possibility of a well-defined stochastic integral.

2.2 Local martingale

From now on, we assume that $(\Omega, \mathcal{F}, (\mathcal{F}_t)_t, \mathbb{P})$ satisfies the *usual conditions*, namely that

- (i) \mathcal{F}_0 contains all \mathbb{P} -null sets
- (ii) $(\mathcal{F}_t)_t$ is right-continuous, i.e. $\mathcal{F}_t = (\mathcal{F}_{t+} = \bigcap_{s>t} \mathcal{F}_s)$ for all $t \geq 0$.

We recall some of the properties of continuous martingales.

Theorem (Optional stopping theorem). Let X be a càdlàg adapted integrable process. Then the following are equivalent:

- (i) X is a martingale, i.e. $X_t \in L^1$ for every t , and

$$\mathbb{E}(X_t \mid \mathcal{F}_s) = X_s \text{ for all } t > s.$$
- (ii) The *stopped process* $X^T = (X_t^T) = (X_{T \wedge t})$ is a martingale for all stopping times T .
- (iii) For all stopping times T, S with T bounded, $X_T \in L^1$ and $\mathbb{E}(X_T \mid \mathcal{F}_S) = X_{T \wedge S}$ almost surely.
- (iv) For all bounded stopping times T , $X_T \in L^1$ and $\mathbb{E}(X_T) = \mathbb{E}(X_0)$.

For X uniformly integrable, (iii) and (iv) hold for all stopping times.

In practice, most of our results will be first proven for bounded martingales, or perhaps square integrable ones. The point is that the square-integrable martingales form a Hilbert space, and Hilbert space techniques can help us say something useful about these martingales. To get something about a general martingale M , we can apply a cutoff $T_n = \inf\{t > 0 : M_t \geq n\}$, and then M^{T_n} will be a martingale for all n . We can then take the limit $n \rightarrow \infty$ to recover something about the martingale itself.

But if we are doing this, we might as well weaken the martingale condition a bit — we only need the M^{T_n} to be martingales. Of course, we aren't doing this just for fun. In general, martingales will not always be closed under the operations we are interested in, but local (or maybe semi-) martingales will be. In general, we define

Definition (Local martingale). A càdlàg adapted process X is a *local martingale* if there exists a sequence of stopping times T_n such that $T_n \rightarrow \infty$ almost surely, and X^{T_n} is a martingale for every n . We say the sequence T_n *reduces* X .

Example.

- (i) Every martingale is a local martingale, since by the optional stopping theorem, we can take $T_n = n$.
- (ii) Let (B_t) to be a standard 3d Brownian motion on \mathbb{R}^3 . Then

$$(X_t)_{t \geq 1} = \left(\frac{1}{|B_t|} \right)_{t \geq 1}$$

is a local martingale but not a martingale.

To see this, first note that

$$\sup_{t \geq 1} \mathbb{E} X_t^2 < \infty, \quad \mathbb{E} X_t \rightarrow 0.$$

Since $\mathbb{E} X_t \rightarrow 0$ and $X_t \geq 0$, we know X cannot be a martingale. However, we can check that it is a local martingale. Recall that for any $f \in C_b^2$,

$$M^f = f(B_t) - f(B_1) - \frac{1}{2} \int_0^t \Delta f(B_s) ds$$

is a martingale. Moreover, $\Delta \frac{1}{|x|} = 0$ for all $x \neq 0$. Thus, if $\frac{1}{|x|}$ didn't have a singularity at 0, this would have told us X_t is a martingale. Thus, we are safe if we try to bound $|B_s|$ away from zero.

Let

$$T_n = \inf \left\{ t \geq 1 : |B_t| < \frac{1}{n} \right\},$$

and pick $f_n \in C_b^2$ such that $f_n(x) = \frac{1}{|x|}$ for $|x| \geq \frac{1}{n}$. Then $X_t^T - X_1^{T_n} = M_{t \wedge T_n}^{f_n}$. So X^{T_n} is a martingale.

It remains to show that $T_n \rightarrow \infty$, and this follows from the fact that $\mathbb{E} X_t \rightarrow 0$.

Proposition. Let X be a local martingale and $X_t \geq 0$ for all t . Then X is a supermartingale.

Proof. Let (T_n) be a reducing sequence for X . Then

$$\begin{aligned} \mathbb{E}(X_t | \mathcal{F}_s) &= \mathbb{E}\left(\liminf_{n \rightarrow \infty} X_{t \wedge T_n} | \mathcal{F}_s\right) \\ &\leq \lim_{n \rightarrow \infty} \mathbb{E}(X_{t \wedge T_n} | \mathcal{F}_s) \\ &= \liminf_{T_n \rightarrow \infty} X_{s \wedge T_n} \\ &= X_s. \end{aligned} \quad \square$$

Recall the following result from Advanced Probability:

Proposition. Let $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. Then the set

$$\chi = \{\mathbb{E}(X | \mathcal{G}) : \mathcal{G} \subseteq \mathcal{F} \text{ a sub-}\sigma\text{-algebra}\}$$

is uniformly integrable, i.e.

$$\sup_{Y \in \chi} \mathbb{E}(|Y| \mathbf{1}_{|Y| > \lambda}) \rightarrow 0 \text{ as } \lambda \rightarrow \infty.$$

Recall also the following important result about uniformly integrable random variables:

Theorem (Vitali theorem). $X_n \rightarrow X$ in L^1 iff (X_n) is uniformly integrable and $X_n \rightarrow X$ in probability.

With these, we can state the following characterization of martingales in terms of local martingales:

Proposition. The following are equivalent:

- (i) X is a martingale.
- (ii) X is a local martingale, and for all $t \geq 0$, the set

$$\chi_t = \{X_T : T \text{ is a stopping time with } T \leq t\}$$

is uniformly integrable.

Proof.

- (a) \Rightarrow (b): Let X be a martingale. Then by the optional stopping theorem, $X_T = \mathbb{E}(X_t | \mathcal{F}_T)$ for any bounded stopping time $T \leq t$. So χ_t is uniformly integrable.
- (b) \Rightarrow (a): Let X be a local martingale with reducing sequence (T_n) , and assume that the sets χ_t are uniformly integrable for all $t \geq 0$. By the optional stopping theorem, it suffices to show that $\mathbb{E}(X_T) = \mathbb{E}(X_0)$ for any bounded stopping time T .

So let T be a bounded stopping time, say $T \leq t$. Then

$$\mathbb{E}(X_0) = \mathbb{E}(X_0^{T_n}) = \mathbb{E}(X_T^{T_n}) = \mathbb{E}(X_{T \wedge T_n})$$

for all n . Now $T \wedge T_n$ is a stopping time $\leq t$, so $\{X_{T \wedge T_n}\}$ is uniformly integrable by assumption. Moreover, $T_n \wedge T \rightarrow T$ almost surely as $n \rightarrow \infty$, hence $X_{T \wedge T_n} \rightarrow X_T$ in probability. Hence by Vitali, this converges in L^1 . So

$$\mathbb{E}(X_T) = \mathbb{E}(X_0). \quad \square$$

Corollary. If $Z \in L^1$ is such that $|X_t| \leq Z$ for all t , then X is a martingale. In particular, every bounded local martingale is a martingale.

The definition of a local martingale does not give us control over what the reducing sequence $\{T_n\}$ is. In particular, it is not necessarily true that X^{T_n} will be bounded, which is a helpful property to have. Fortunately, we have the following proposition:

Proposition. Let X be a *continuous* local martingale with $X_0 = 0$. Define

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

Then S_n is a stopping time, $S_n \rightarrow \infty$ and X^{S_n} is a bounded martingale. In particular, (S_n) reduces X .

Proof. It is clear that S_n is a stopping time, since (if it is not clear)

$$\{S_n \leq t\} = \bigcap_{k \in \mathbb{N}} \left\{ \sup_{s \leq t} |X_s| > n - \frac{1}{k} \right\} = \bigcap_{k \in \mathbb{N}} \bigcup_{s < t, s \in \mathbb{Q}} \left\{ |X_s| > n - \frac{1}{k} \right\} \in \mathcal{F}_t.$$

It is also clear that $S_n \rightarrow \infty$, since

$$\sup_{s \leq t} |X_s| \leq n \leftrightarrow S_n \geq t,$$

and by continuity and compactness, $\sup_{s \leq t} |X_s|$ is finite for every (ω, t) .

Finally, we show that X^{S_n} is a martingale. By the optional stopping theorem, $X^{T_n \wedge S_n}$ is a martingale, so X^{S_n} is a local martingale. But it is also bounded by n . So it is a martingale. \square

An important and useful theorem is the following:

Theorem. Let X be a continuous local martingale with $X_0 = 0$. If X is also a finite variation process, then $X_t = 0$ for all t .

This would rule out interpreting $\int H_s dX_s$ as a Lebesgue–Stieltjes integral for X a non-zero continuous local martingale. In particular, we cannot take X to be Brownian motion. Instead, we have to develop a new theory of integration for continuous local martingales, namely the Itô integral.

On the other hand, this theorem is very useful. We will later want to define the stochastic integral with respect to the sum of a continuous local martingale and a finite variation process, which is the appropriate generality for our theorems to make good sense. This theorem tells us there is a unique way to decompose a process as a sum of a finite variation process and a continuous local martingale (if it can be done). So we can simply define this stochastic integral by using the Lebesgue–Stieltjes integral on the finite variation part and the Itô integral on the continuous local martingale part.

Proof. Let X be a finite-variation continuous local martingale with $X_0 = 0$. Since X is finite variation, we can define the total variation process (V_t) corresponding to X , and let

$$S_n = \inf\{t \geq 0 : V_t \geq n\} = \inf\left\{t \geq 0 : \int_0^1 |dX_s| \geq n\right\}.$$

Then S_n is a stopping time, and $S_n \rightarrow \infty$ since X is assumed to be finite variation. Moreover, by optional stopping, X^{S_n} is a local martingale, and is also bounded, since

$$X_t^{S_n} \leq \int_0^{t \wedge S_n} |dX_s| \leq n.$$

So X^{S_n} is in fact a martingale.

We claim its L^2 -norm vanishes. Let $0 = t_0 < t_1 < \dots < t_n = t$ be a subdivision of $[0, t]$. Using the fact that X^{S_n} is a martingale and has orthogonal increments, we can write

$$\mathbb{E}((X_t^{S_n})^2) = \sum_{i=1}^k \mathbb{E}((X_{t_i}^{S_n} - X_{t_{i-1}}^{S_n})^2).$$

Observe that X^{S_n} is finite variation, but the right-hand side is summing the *square* of the variation, which ought to vanish when we take the limit $\max |t_i - t_{i-1}| \rightarrow 0$. Indeed, we can compute

$$\begin{aligned} \mathbb{E}((X_t^{S_n})^2) &= \sum_{i=1}^k \mathbb{E}((X_{t_i}^{S_n} - X_{t_{i-1}}^{S_n})^2) \\ &\leq \mathbb{E}\left(\max_{1 \leq i \leq k} |X_{t_i}^{S_n} - X_{t_{i-1}}^{S_n}| \sum_{i=1}^k |X_{t_i}^{S_n} - X_{t_{i-1}}^{S_n}|\right) \\ &\leq \mathbb{E}\left(\max_{1 \leq i \leq k} |X_{t_i}^{S_n} - X_{t_{i-1}}^{S_n}| V_{t \wedge S_n}\right) \\ &\leq \mathbb{E}\left(\max_{1 \leq i \leq k} |X_{t_i}^{S_n} - X_{t_{i-1}}^{S_n}| n\right). \end{aligned}$$

Of course, the first term is also bounded by the total variation. Moreover, we can make further subdivisions so that the mesh size tends to zero, and then the first term vanishes in the limit by continuity. So by dominated convergence, we must have $\mathbb{E}((X_t^{S_n})^2) = 0$. So $X_t^{S_n} = 0$ almost surely for all n . So $X_t = 0$ for all t almost surely. \square

2.3 Square integrable martingales

As previously discussed, we will want to use Hilbert space machinery to construct the Itô integral. The rough idea is to define the Itô integral with respect to a fixed martingale on simple processes via a (finite) Riemann sum, and then by calculating appropriate bounds on how this affects the norm, we can extend this to all processes by continuity, and this requires our space to be Hilbert. The interesting spaces are defined as follows:

Definition (\mathcal{M}^2). Let

$$\mathcal{M}^2 = \left\{ X : \Omega \times [0, \infty) \rightarrow \mathbb{R} : X \text{ is càdlàg martingale with } \sup_{t \geq 0} \mathbb{E}(X_t^2) < \infty \right\}.$$

$$\mathcal{M}_c^2 = \{ X \in \mathcal{M}^2 : X(\omega, \cdot) \text{ is continuous for every } \omega \in \Omega \}$$

We define an inner product on \mathcal{M}^2 by

$$(X, Y)_{\mathcal{M}^2} = \mathbb{E}(X_\infty Y_\infty),$$

which in particular induces a norm

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

We will prove this is indeed an inner product soon. Here recall that for $X \in \mathcal{M}^2$, the martingale convergence theorem implies $X_t \rightarrow X_\infty$ almost surely and in L^2 .

Our goal will be to prove that these spaces are indeed Hilbert spaces. First observe that if $X \in \mathcal{M}^2$, then $(X_t^2)_{t \geq 0}$ is a submartingale by Jensen, so $t \mapsto \mathbb{E}X_t^2$ is increasing, and

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

All the magic that lets us prove they are Hilbert spaces is Doob's inequality.

Theorem (Doob's inequality). Let $X \in \mathcal{M}^2$. Then

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

So once we control the limit X_∞ , we control the whole path. This is why the definition of the norm makes sense, and in particular we know $\|X\|_{\mathcal{M}^2} = 0$ implies that $X = 0$.

Theorem. \mathcal{M}^2 is a Hilbert space and \mathcal{M}_c^2 is a closed subspace.

Proof. We need to check that \mathcal{M}^2 is complete. Thus let $(X^n) \subseteq \mathcal{M}^2$ be a Cauchy sequence, i.e.

$$\mathbb{E}((X_\infty^n - X_\infty^m)^2) \rightarrow 0 \text{ as } n, m \rightarrow \infty.$$

By passing to a subsequence, we may assume that

$$\mathbb{E}((X_\infty^n - X_\infty^{n-1})^2) \leq 2^{-n}.$$

First note that

$$\begin{aligned} \mathbb{E} \left(\sum_n \sup_{t \geq 0} |X_t^n - X_t^{n-1}| \right) &\leq \sum_n \mathbb{E} \left(\sup_{t \geq 0} |X_t^n - X_t^{n-1}|^2 \right)^{1/2} && \text{(CS)} \\ &\leq \sum_n 2\mathbb{E} (|X_\infty^n - X_\infty^{n-1}|^2)^{1/2} && \text{(Doob's)} \\ &\leq 2 \sum_n 2^{-n/2} < \infty. \end{aligned}$$

So

$$\sum_{n=1}^{\infty} \sup_{t \geq 0} |X_t^n - X_t^{n-1}| < \infty \text{ a.s.} \quad (*)$$

So on this event, (X^n) is a Cauchy sequence in the space $(D[0, \infty), \|\cdot\|_{\infty})$ of càdlàg sequences. So there is some $X(\omega, \cdot) \in D[0, \infty)$ such that

$$\|X^n(\omega, \cdot) - X(\omega, \cdot)\|_{\infty} \rightarrow 0 \text{ for almost all } \omega.$$

and we set $X = 0$ outside this almost sure event $(*)$. We now claim that

$$\mathbb{E} \left(\sup_{t \geq 0} |X^n - X|^2 \right) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

We can just compute

$$\begin{aligned} \mathbb{E} \left(\sup_t |X^n - X|^2 \right) &= \mathbb{E} \left(\lim_{m \rightarrow \infty} \sup_t |X^n - X^m|^2 \right) \\ &\leq \liminf_{m \rightarrow \infty} \mathbb{E} \left(\sup_t |X^n - X^m|^2 \right) \quad (\text{Fatou}) \\ &\leq \liminf_{m \rightarrow \infty} 4\mathbb{E}(X_{\infty}^n - X_m^{\infty})^2 \quad (\text{Doob's}) \end{aligned}$$

and this goes to 0 in the limit $n \rightarrow \infty$ as well.

We finally have to check that X is indeed a martingale. We use the triangle inequality to write

$$\begin{aligned} \|E(X_t | \mathcal{F}_s) - X_s\|_{L^2} &\leq \|E(X_t - X_t^n | \mathcal{F}_s)\|_{L^2} + \|X_s^n - X_s\|_{L^2} \\ &\leq \mathbb{E}(\mathbb{E}((X_t - X_t^n)^2 | \mathcal{F}_s))^{1/2} + \|X_s^n - X_s\|_{L^2} \\ &= \|X_t - X_t^n\|_{L^2} + \|X_s^n - X_s\|_{L^2} \\ &\leq 2\mathbb{E} \left(\sup_t |X_t - X_t^n|^2 \right)^{1/2} \rightarrow 0 \end{aligned}$$

as $n \rightarrow \infty$. But the left-hand side does not depend on n . So it must vanish. So $X \in \mathcal{M}^2$.

We could have done exactly the same with continuous martingales, so the second part follows. \square

2.4 Quadratic variation

Physicists are used to dropping all terms above first-order. It turns out that Brownian motion, and continuous local martingales in general oscillate so wildly that second order terms become important. We first make the following definition:

Definition (Uniformly on compact sets in probability). For a sequence of processes (X^n) and a process X , we say that $X^n \rightarrow X$ u.c.p. iff

$$\mathbb{P} \left(\sup_{s \in [0, t]} |X_s^n - X_s| > \varepsilon \right) \rightarrow 0 \text{ as } n \rightarrow \infty \text{ for all } t > 0, \varepsilon > 0.$$

Theorem. Let M be a continuous local martingale with $M_0 = 0$. Then there exists a unique (up to indistinguishability) continuous adapted increasing process $(\langle M \rangle_t)_{t \geq 0}$ such that $\langle M \rangle_0 = 0$ and $M_t^2 - \langle M \rangle_t$ is a continuous local martingale. Moreover,

$$\langle M \rangle_t = \lim_{n \rightarrow \infty} \langle M \rangle_t^{(n)}, \quad \langle M \rangle_t^{(n)} = \sum_{i=1}^{\lceil 2^n t \rceil} (M_{t2^{-n}} - M_{(i-1)2^{-n}})^2,$$

where the limit u.c.p.

Definition (Quadratic variation). $\langle M \rangle$ is called the *quadratic variation* of M .

It is probably more useful to understand $\langle M \rangle_t$ in terms of the explicit formula, and the fact that $M_t^2 - \langle M \rangle_t$ is a continuous local martingale is a convenient property.

Example. Let B be a standard Brownian motion. Then $B_t^2 - t$ is a martingale. Thus, $\langle B \rangle_t = t$.

The proof is long and mechanical, but not hard. All the magic happened when we used the magical Doob's inequality to show that \mathcal{M}_c^2 and \mathcal{M}^2 are Hilbert spaces.

Proof. To show uniqueness, we use that finite variation and local martingale are incompatible. Suppose (A_t) and (\tilde{A}_t) obey the conditions for $\langle M \rangle$. Then $A_t - \tilde{A}_t = (M_t^2 - \tilde{A}_t) - (M_t^2 - A_t)$ is a continuous adapted local martingale starting at 0. Moreover, both A_t and \tilde{A}_t are increasing, hence have finite variation. So $A - \tilde{A} = 0$ almost surely.

To show existence, we need to show that the limit exists and has the right property. We do this in steps.

Claim. The result holds if M is in fact bounded.

Suppose $|M(\omega, t)| \leq C$ for all (ω, t) . Then $M \in \mathcal{M}_c^2$. Fix $T > 0$ deterministic. Let

$$X_t^n = \sum_{i=1}^{\lceil 2^n T \rceil} M_{(i-1)2^{-n}} (M_{i2^{-n} \wedge t} - M_{(i-1)2^{-n} \wedge t}).$$

This is defined so that

$$\langle M \rangle_{k2^{-n}}^{(n)} = M_{k2^{-n}}^2 - 2X_{k2^{-n}}^n.$$

This reduces the study of $\langle M \rangle^{(n)}$ to that of $X_{k2^{-n}}^n$.

We check that (X_t^n) is a Cauchy sequence in \mathcal{M}_c^2 . The fact that it is a martingale is an immediate computation. To show it is Cauchy, for $n \geq m$, we calculate

$$X_\infty^n - X_\infty^m = \sum_{i=1}^{\lceil 2^n T \rceil} (M_{(i-1)2^{-n}} - M_{\lfloor (i-1)2^{m-n} \rfloor 2^{-m}}) (M_{i2^{-n}} - M_{(i-1)2^{-n}}).$$

We now take the expectation of the square to get

$$\begin{aligned}
\mathbb{E}(X_\infty^n - X_\infty^m)^2 &= \mathbb{E} \left(\sum_{i=1}^{\lceil 2^n T \rceil} (M_{(i-1)2^{-n}} - M_{\lfloor (i-1)2^{m-n} \rfloor 2^{-m}})^2 (M_{i2^{-n}} - M_{(i-1)2^{-n}})^2 \right) \\
&\leq \mathbb{E} \left(\sup_{|s-t| \leq 2^{-m}} |M_t - M_s|^2 \sum_{i=1}^{\lceil 2^n T \rceil} (M_{i2^{-n}} - M_{(i-1)2^{-n}})^2 \right) \\
&= \mathbb{E} \left(\sup_{|s-t| \leq 2^{-m}} |M_t - M_s|^2 \langle M \rangle_T^{(n)} \right) \\
&\leq \mathbb{E} \left(\sup_{|s-t| \leq 2^{-m}} |M_t - M_s|^4 \right)^{1/2} \mathbb{E} \left((\langle M \rangle_T^{(n)})^2 \right)^{1/2} \\
&\hspace{15em} \text{(Cauchy-Schwarz)}
\end{aligned}$$

We shall show that the second factor is bounded, while the first factor tends to zero as $m \rightarrow \infty$. These are both not surprising — the first term vanishing in the limit corresponds to M being continuous, and the second term is bounded since M itself is bounded.

To show that the first term tends to zero, we note that we have

$$|M_t - M_s|^4 \leq 16C^4,$$

and moreover

$$\sup_{|s-t| \leq 2^{-m}} |M_t - M_s| \rightarrow 0 \text{ as } m \rightarrow \infty \text{ by uniform continuity.}$$

So we are done by the dominated convergence theorem.

To show the second term is bounded, we do (writing $N = \lceil 2^n T \rceil$)

$$\begin{aligned}
\mathbb{E} \left((\langle M \rangle_T^{(n)})^2 \right) &= \mathbb{E} \left(\left(\sum_{i=1}^N (M_{i2^{-n}} - M_{(i-1)2^{-n}})^2 \right)^2 \right) \\
&= \sum_{i=1}^N \mathbb{E} \left((M_{i2^{-n}} - M_{(i-1)2^{-n}})^4 \right) \\
&\quad + 2 \sum_{i=1}^N \mathbb{E} \left((M_{i2^{-n}} - M_{(i-1)2^{-n}})^2 \sum_{k=i+1}^N (M_{k2^{-n}} - M_{(k-1)2^{-n}})^2 \right)
\end{aligned}$$

We use the martingale property and orthogonal increments to rearrange the off-diagonal term as

$$\mathbb{E} \left((M_{i2^{-n}} - M_{(i-1)2^{-n}})(M_{N2^{-n}} - M_{i2^{-n}})^2 \right).$$

Taking some sups, we get

$$\begin{aligned}
\mathbb{E} \left((\langle M \rangle_T^{(n)})^2 \right) &\leq 12C^2 \mathbb{E} \left(\sum_{i=1}^N (M_{i2^{-n}} - M_{(i-1)2^{-n}})^2 \right) \\
&= 12C^2 \mathbb{E} \left((M_{N2^{-n}} - M_0)^2 \right) \\
&\leq 12C^2 \cdot 4C^2.
\end{aligned}$$

So done.

So we now have $X^n \rightarrow X$ in M_c^2 for some $X \in M_c^2$. In particular, we have

$$\left\| \sup_t |X_t^n - X_t| \right\|_{L^2} \rightarrow 0$$

So we know that

$$\sup_t |X_t^n - X_t| \rightarrow 0$$

almost surely along a subsequence Λ .

Let $N \subseteq \Omega$ be the events on which this convergence fails. We define

$$A_t^{(T)} = \begin{cases} M_t^2 - 2X_t & \omega \in \Omega \setminus N \\ 0 & \omega \in N \end{cases}.$$

Then $A^{(T)}$ is continuous, adapted since M and X are, and $(M_{t \wedge T}^2 - A_{t \wedge T}^{(T)})_t$ is a martingale since X is. Finally, $A^{(T)}$ is increasing since $M_t^2 - X_t^n$ is increasing on $2^{-n}\mathbb{Z} \cap [0, T]$ and the limit is uniform. So this $A^{(T)}$ basically satisfies all the properties we want $\langle M \rangle_t$ to satisfy, except we have the stopping time T .

We next observe that for any $T \geq 1$, $A_{t \wedge T}^{(T)} = A_{t \wedge T}^{(T+1)}$ for all t almost surely. This essentially follows from the same uniqueness argument as we had at the beginning of the proof. Thus, there is a process $(\langle M \rangle_t)_{t \geq 0}$ such that

$$\langle M \rangle_t = A_t^{(T)}$$

for all $t \in [0, T]$ and $T \in \mathbb{N}$, almost surely. Then this is the desired process. So we have constructed $\langle M \rangle$ in the case where M is bounded.

Claim. $\langle M \rangle^{(n)} \rightarrow \langle M \rangle$ u.c.p.

Recall that

$$\langle M \rangle_t^{(n)} = M_{2^{-n}\lfloor 2^n t \rfloor}^2 - 2X_{2^{-n}\lfloor 2^n t \rfloor}.$$

We also know that

$$\sup_{t \leq T} |X_t^n - X_t| \rightarrow 0$$

in L^2 , hence also in probability. So we have

$$\begin{aligned} |\langle M \rangle_t - \langle M \rangle_t^{(n)}| &\leq \sup_{t \leq T} |M_{2^{-n}\lfloor 2^n t \rfloor}^2 - M_t^2| \\ &\quad + \sup_{t \leq T} |X_{2^{-n}\lfloor 2^n t \rfloor}^n - X_{2^{-n}\lfloor 2^n t \rfloor}| + \sup_{t \leq T} |X_{2^{-n}\lfloor 2^n t \rfloor} - X_t|. \end{aligned}$$

The first and last terms $\rightarrow 0$ in probability since M and X are uniformly continuous on $[0, T]$. The second term converges to zero by our previous assertion. So we are done.

Claim. The theorem holds for M any continuous local martingale.

We let $T_n = \inf\{t \geq 0 : |M_t| \geq n\}$. Then (T_n) reduces M and M^{T_n} is a bounded martingale. So in particular M^{T_n} is a bounded continuous martingale. We set

$$A^n = \langle M^{T_n} \rangle.$$

Then (A_t^n) and $(A_{t \wedge T_n}^{n+1})$ are indistinguishable for $t < T_n$ by the uniqueness argument. Thus there is a process $\langle M \rangle$ such that $\langle M \rangle_{t \wedge T_n} = A_t^n$ are indistinguishable for all n . Clearly, $\langle M \rangle$ is increasing since the A^n are, and $M_{t \wedge T_n}^2 - \langle M \rangle_{t \wedge T_n}$ is a martingale for every n , so $M_t^2 - \langle M \rangle_t$ is a continuous local martingale.

Claim. $\langle M \rangle^{(n)} \rightarrow \langle M \rangle$ u.c.p.

We have seen

$$\langle M^{T_k} \rangle^{(n)} \rightarrow \langle M^{T_k} \rangle \text{ u.c.p.}$$

for every k . So

$$\mathbb{P} \left(\sup_{t \leq T} |\langle M \rangle_t^{(n)} - \langle M \rangle_t| > \varepsilon \right) \leq \mathbb{P}(T_k < T) + \mathbb{P} \left(\sup_{t \leq T} |\langle M^{T_k} \rangle_t^{(n)} - \langle M^{T_k} \rangle_t| > \varepsilon \right).$$

So we can first pick k large enough such that the first term is small, then pick n large enough so that the second is small. \square

There are a few easy consequences of this theorem.

Fact. Let M be a continuous local martingale, and let T be a stopping time. Then almost surely for all $t \geq 0$,

$$\langle M^T \rangle_t = \langle M \rangle_{t \wedge T}$$

Proof. Since $M_t^2 - \langle M \rangle_t$ is a continuous local martingale, so is $M_{t \wedge T}^2 - \langle M \rangle_{t \wedge T} = (M^T)_t^2 - \langle M \rangle_{t \wedge T}$. So we are done by uniqueness. \square

Fact. Let M be a continuous local martingale with $M_0 = 0$. Then $M = 0$ iff $\langle M \rangle = 0$.

Proof. If $M = 0$, then $\langle M \rangle = 0$. Conversely, if $\langle M \rangle = 0$, then M^2 is a continuous local martingale and positive. Thus $\mathbb{E}M_t^2 \leq \mathbb{E}M_0^2 = 0$. \square

Proposition. Let $M \in \mathcal{M}_c^2$. Then $M^2 - \langle M \rangle$ is a uniformly integrable martingale, and

$$\|M - M_0\|_{\mathcal{M}^2} = (\mathbb{E}\langle M \rangle_\infty)^{1/2}.$$

Proof. We will show that $\langle M \rangle_\infty \in L^1$. This then implies

$$|M_t^2 - \langle M \rangle_t| \leq \sup_{t \geq 0} M_t^2 + \langle M \rangle_\infty.$$

Then the right hand side is in L^1 . Since $M^2 - \langle M \rangle$ is a local martingale, this implies that it is in fact a uniformly integrable martingale.

To show $\langle M \rangle_\infty \in L^1$, we let

$$S_n = \inf\{t \geq 0 : \langle M \rangle_t \geq n\}.$$

Then $S_n \rightarrow \infty$, S_n is a stopping time and moreover $\langle M \rangle_{t \wedge S_n} \leq n$. So we have

$$M_{t \wedge S_n}^2 - \langle M \rangle_{t \wedge S_n} \leq n + \sup_{t \geq 0} M_t^2,$$

and the second term is in L^1 . So $M_{t \wedge S_n}^2 - \langle M \rangle_{t \wedge S_n}$ is a true martingale.

So

$$\mathbb{E}M_{t \wedge S_n}^2 - \mathbb{E}M_0^2 = \mathbb{E}\langle M \rangle_{t \wedge S_n}.$$

Taking the limit $t \rightarrow \infty$, we know $\mathbb{E}M_{t \wedge S_n}^2 \rightarrow \mathbb{E}M_{S_n}^2$ by dominated convergence. Since $\langle M \rangle_{t \wedge S_n}$ is increasing, we also have $\mathbb{E}\langle M \rangle_{t \wedge S_n} \rightarrow \mathbb{E}\langle M \rangle_{S_n}$ by *monotone* convergence. We can take $n \rightarrow \infty$, and by the same justification, we have

$$\mathbb{E}\langle M \rangle \leq \mathbb{E}M_\infty^2 - \mathbb{E}M_0^2 = \mathbb{E}(M_\infty - M_0)^2 < \infty. \quad \square$$

2.5 Covariation

We know \mathcal{M}_c^2 not only has a norm, but also an inner product. This can also be reflected in the bracket by the polarization identity, and it is natural to define

Definition (Covariation). Let M, N be two continuous local martingales. Define the *covariation* (or simply the *bracket*) between M and N to be process

$$\langle M, N \rangle_t = \frac{1}{4}(\langle M + N \rangle_t - \langle M - N \rangle_t).$$

Then if in fact $M, N \in \mathcal{M}_c^2$, then putting $t = \infty$ gives the inner product.

Proposition.

- (i) $\langle M, N \rangle$ is the unique (up to indistinguishability) finite variation process such that $M_t N_t - \langle M, N \rangle_t$ is a continuous local martingale.
- (ii) The mapping $(M, N) \mapsto \langle M, N \rangle$ is bilinear and symmetric.
- (iii)

$$\begin{aligned} \langle M, N \rangle_t &= \lim_{n \rightarrow \infty} \langle M, N \rangle_t^{(n)} \text{ u.c.p.} \\ \langle M, N \rangle_t^{(n)} &= \sum_{i=1}^{\lceil 2^n t \rceil} (M_{i2^{-n}} - M_{(i-1)2^{-n}})(N_{i2^{-n}} - N_{(i-1)2^{-n}}). \end{aligned}$$

- (iv) For every stopping time T ,

$$\langle M^T, N^T \rangle_t = \langle M^T, N \rangle_t = \langle M, N \rangle_{t \wedge T}.$$

- (v) If $M, N \in \mathcal{M}_c^2$, then $M_t N_t - \langle M, N \rangle_t$ is a uniformly integrable martingale, and

$$\langle M - M_0, N - N_0 \rangle_{\mathcal{M}^2} = \mathbb{E}\langle M, N \rangle_\infty. \quad \square$$

Example. Let B, B' be two independent Brownian motions (with respect to the same filtration). Then $\langle B, B' \rangle = 0$.

Proof. Assume $B_0 = B'_0 = 0$. Then $X_\pm = \frac{1}{\sqrt{2}}(B \pm B')$ are Brownian motions, and so $\langle X_\pm \rangle = t$. So their difference vanishes. \square

An important result about the covariation is the following Cauchy–Schwarz like inequality:

Proposition (Kunita–Watanabe). Let M, N be continuous local martingales and let H, K be two (previsible) processes. Then almost surely

$$\int_0^\infty |H_s| |K_s| |d\langle M, N \rangle_s| \leq \left(\int_0^\infty H_s^2 d\langle M \rangle_s \right)^{1/2} \left(\int_0^\infty H_s^2 \langle N \rangle_s \right)^{1/2}.$$

In fact, this is Cauchy–Schwarz. All we have to do is to take approximations and take limits and make sure everything works out well.

Proof. For convenience, we write

$$\langle M, N \rangle_s^t = \langle M, N \rangle_t - \langle M, N \rangle_s.$$

Claim. For all $0 \leq s \leq t$, we have

$$|\langle M, N \rangle_s^t| \leq \sqrt{\langle M, M \rangle_s^t} \sqrt{\langle N, N \rangle_s^t}.$$

By continuity, we can assume that s, t are dyadic rationals. Then

$$\begin{aligned} |\langle M, N \rangle_s^t| &= \lim_{n \rightarrow \infty} \left| \sum_{i=2^{n-1}s+1}^{2^n t} (M_{i2^{-n}} - M_{(i-1)2^{-n}})(N_{i2^{-n}} - N_{(i-1)2^{-n}}) \right| \\ &\leq \lim_{n \rightarrow \infty} \left| \sum_{i=2^{n-1}s+1}^{2^n t} (M_{i2^{-n}} - M_{(i-1)2^{-n}})^2 \right|^{1/2} \times \\ &\quad \left| \sum_{i=2^{n-1}s+1}^{2^n t} (N_{i2^{-n}} - N_{(i-1)2^{-n}})^2 \right|^{1/2} \quad (\text{Cauchy–Schwarz}) \\ &= (\langle M, M \rangle_s^t)^{1/2} (\langle N, N \rangle_s^t)^{1/2}, \end{aligned}$$

where all equalities are u.c.p.

Claim. For all $0 \leq s < t$, we have

$$\int_s^t |d\langle M, N \rangle_u| \leq \sqrt{\langle M, M \rangle_s^t} \sqrt{\langle N, N \rangle_s^t}.$$

Indeed, for any subdivision $s = t_0 < t_1 < \dots < t_n = t$, we have

$$\begin{aligned} \sum_{i=1}^n |\langle M, N \rangle_{t_{i-1}}^{t_i}| &\leq \sum_{i=1}^n \sqrt{\langle M, M \rangle_{t_{i-1}}^{t_i}} \sqrt{\langle N, N \rangle_{t_{i-1}}^{t_i}} \\ &\leq \left(\sum_{i=1}^n \langle M, M \rangle_{t_{i-1}}^{t_i} \right)^{1/2} \left(\sum_{i=1}^n \langle N, N \rangle_{t_{i-1}}^{t_i} \right)^{1/2}. \end{aligned} \quad (\text{Cauchy–Schwarz})$$

Taking the supremum over all subdivisions, the claim follows.

Claim. For all bounded Borel sets $B \subseteq [0, \infty)$, we have

$$\int_B |d\langle M, N \rangle_u| \leq \sqrt{\int_B d\langle M \rangle_u} \sqrt{\int_B d\langle N \rangle_u}.$$

We already know this is true if B is an interval. If B is a finite union of intervals, then we apply Cauchy–Schwarz. By a monotone class argument, we can extend to all Borel sets.

Claim. The theorem holds for

$$H = \sum_{\ell=1}^k h_{\ell} \mathbf{1}_{B_{\ell}}, \quad K = \sum_{\ell=1}^n k_{\ell} \mathbf{1}_{B_{\ell}}$$

for $B_{\ell} \subseteq [0, \infty)$ bounded Borel sets with disjoint support.

We have

$$\begin{aligned} \int |H_s K_s| |d\langle M, N \rangle_s| &\leq \sum_{\ell=1}^n |h_{\ell} k_{\ell}| \int_{B_{\ell}} |d\bar{M}, N\rangle_s| \\ &\leq \sum_{\ell=1}^n |h_{\ell} k_{\ell}| \left(\int_{B_{\ell}} d\langle M \rangle_s \right)^{1/2} \left(\int_{B_{\ell}} d\langle N \rangle_s \right)^{1/2} \\ &\leq \left(\sum_{\ell=1}^n h_{\ell}^2 \int_{B_{\ell}} d\langle M \rangle_s \right)^{1/2} \left(\sum_{\ell=1}^n k_{\ell}^2 \int_{B_{\ell}} d\langle N \rangle_s \right)^{1/2} \end{aligned}$$

To finish the proof, approximate general H and K by step functions and take the limit. \square

2.6 Semi-martingale

Definition (Semi-martingale). A (continuous) adapted process X is a (*continuous*) *semi-martingale* if

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

where $X_0 \in \mathcal{F}_0$, M is a continuous local martingale with $M_0 = 0$, and A is a continuous finite variation process with $A_0 = 0$.

This decomposition is unique up to indistinguishables.

Definition (Quadratic variation). Let $X = X_0 + M + A$ and $X' = X'_0 + M' + A'$ be (continuous) semi-martingales. Set

$$\langle X \rangle = \langle M \rangle, \quad \langle X, X' \rangle = \langle M, M' \rangle.$$

This definition makes sense, because continuous finite variation processes do not have quadratic variation.

Exercise. We have

$$\langle X, Y \rangle_t^{(n)} = \sum_{i=1}^{\lceil 2^n t \rceil} (X_{i2^{-n}} - X_{(i-1)2^{-n}})(Y_{i2^{-n}} - Y_{(i-1)2^{-n}}) \rightarrow \langle X, Y \rangle \text{ u.c.p.}$$

3 The stochastic integral

3.1 Simple processes

We now have all the background required to define the stochastic integral, and we can start constructing it. As in the case of the Lebesgue integral, we first define it for simple processes, and then extend to general processes by taking a limit. Recall that we have

Definition (Simple process). The space of *simple processes* \mathcal{E} consists of functions $H : \Omega \times [0, \infty) \rightarrow \mathbb{R}$ that can be written as

$$H_t(\omega) = \sum_{i=1}^n H_{i-1}(\omega) \mathbf{1}_{(t_{i-1}, t_i]}(t)$$

for some $0 \leq t_0 \leq t_1 \leq \dots \leq t_n$ and bounded random variables $H_i \in \mathcal{F}_{t_i}$.

Definition ($H \cdot M$). For $M \in \mathcal{M}^2$ and $H \in \mathcal{E}$, we set

$$\int_0^t H \, dM = (H \cdot M)_t = \sum_{i=1}^n H_{i-1} (M_{t_i \wedge t} - M_{t_{i-1} \wedge t}).$$

If M were of finite variation, then this is the same as what we have previously seen.

Recall that for the Lebesgue integral, extending this definition to general functions required results like monotone convergence. Here we need some similar results that put bounds on how large the integral can be. In fact, we get something better than a bound.

Proposition. If $M \in \mathcal{M}_c^2$ and $H \in \mathcal{E}$, then $H \cdot M \in \mathcal{M}_c^2$ and

$$\|H \cdot M\|_{\mathcal{M}^2}^2 = \mathbb{E} \left(\int_0^\infty H_s^2 \, d\langle M \rangle_s \right). \quad (*)$$

Proof. We first show that $H \cdot M \in \mathcal{M}_c^2$. By linearity, we only have to check it for

$$X_t^i = H_{i-1} (M_{t_i \wedge t} - M_{t_{i-1} \wedge t})$$

We have to check that $\mathbb{E}(X_t^i | \mathcal{F}_s) = 0$ for all $t > s$, and the only non-trivial case is when $t > t_{i-1}$.

$$\mathbb{E}(X_t^i | \mathcal{F}_s) = H_{i-1} \mathbb{E}(M_{t_i \wedge t} - M_{t_{i-1} \wedge t} | \mathcal{F}_s) = 0.$$

We also check that

$$\|X^i\|_{\mathcal{M}^2} \leq 2\|H\|_\infty \|M\|_{\mathcal{M}^2}.$$

So it is bounded. So $H \cdot M \in \mathcal{M}_c^2$.

To prove (*), we note that the X^i are orthogonal and that

$$\langle X^i \rangle_t = H_{i-1}^2 (\langle M \rangle_{t_i \wedge t} - \langle M \rangle_{t_{i-1} \wedge t}).$$

So we have

$$\langle H \cdot M, H \cdot M \rangle = \sum \langle X^i, X^i \rangle = \sum H_{i-1}^2 (\langle M \rangle_{t_i \wedge t} - \langle M \rangle_{t_{i-1} \wedge t}) = \int_0^t H_s^2 \, d\langle M \rangle_s.$$

In particular,

$$\|H \cdot M\|_{\mathcal{M}^2}^2 = \mathbb{E}\langle H \cdot M \rangle_\infty = \mathbb{E}\left(\int_0^\infty H_s^2 d\langle M \rangle_s\right). \quad \square$$

Proposition. Let $M \in \mathcal{M}_c^2$ and $H \in \mathcal{E}$. Then

$$\langle H \cdot M, N \rangle = H \cdot \langle M, N \rangle$$

for all $N \in \mathcal{M}^2$.

In other words, the stochastic integral commutes with the bracket.

Proof. Write $H \cdot M = \sum X^i = \sum H_{i-1}(M_{t_i \wedge t} - M_{t_{i-1} \wedge t})$ as before. Then

$$\langle X^i, N \rangle_t = H_{i-1} \langle M_{t_i \wedge t} - M_{t_{i-1} \wedge t}, N \rangle = H_{i-1} (\langle M, N \rangle_{t_i \wedge t} - \langle M, N \rangle_{t_{i-1} \wedge t}). \quad \square$$

3.2 Itô isometry

We now try to extend the above definition to something more general than simple processes.

Definition ($L^2(M)$). Let $M \in \mathcal{M}_c^2$. Define $L^2(M)$ to be the space of (equivalence classes of) previsible $H : \Omega \times [0, \infty) \rightarrow \mathbb{R}$ such that

$$\|H\|_{L^2(M)} = \|H\|_{\mathcal{M}} = \mathbb{E}\left(\int_0^\infty H_s^2 d\langle M \rangle_s\right)^{1/2} < \infty.$$

For $H, K \in L^2(M)$, we set

$$(H, K)_{L^2(M)} = \mathbb{E}\left(\int_0^\infty H_s K_s d\langle M \rangle_s\right).$$

In fact, $L^2(M)$ is equal to $L^2(\Omega \times [0, \infty), \mathcal{P}, dP d\langle M \rangle)$, where \mathcal{P} is the previsible σ -algebra, and in particular is a Hilbert space.

Proposition. Let $M \in \mathcal{M}_c^2$. Then \mathcal{E} is dense in $L^2(M)$.

Proof. Since $L^2(M)$ is a Hilbert space, it suffices to show that if $(K, H) = 0$ for all $H \in \mathcal{E}$, then $K = 0$.

So assume that $(K, H) = 0$ for all $H \in \mathcal{E}$ and

$$X_t = \int_0^t K_s d\langle M \rangle_s,$$

Then X is a well-defined finite variation process, and $X_t \in L^1$ for all t . It suffices to show that $X_t = 0$ for all t , and we shall show that X_t is a continuous martingale.

Let $0 \leq s < t$ and $F \in \mathcal{F}_s$ bounded. We let $H = F1_{(s,t]} \in \mathcal{E}$. By assumption, we know

$$0 = (K, H) = \mathbb{E}\left(F \int_s^t K_u d\langle M \rangle_u\right) = \mathbb{E}(F(X_t - X_s)).$$

Since this holds for all \mathcal{F}_s measurable F , we have shown that

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

So X is a (continuous) martingale, and we are done. \square

Theorem. Let $M \in \mathcal{M}_c^2$. Then

- (i) The map $H \in \mathcal{E} \mapsto H \cdot M \in \mathcal{M}_c^2$ extends uniquely to an isometry $L^2(M) \rightarrow \mathcal{M}_c^2$, called the *Itô isometry*.
- (ii) For $H \in L^2(M)$, $H \cdot M$ is the unique martingale in \mathcal{M}_c^2 such that

$$\langle H \cdot M, N \rangle = H \cdot \langle M, N \rangle$$

for all $N \in \mathcal{M}_c^2$, where the integral on the LHS is the stochastic integral (as above) and the RHS is the finite variation integral.

- (iii) If T is a stopping time, then $(1_{[0,T]}H) \cdot M = (H \cdot M)^T = H \cdot M^T$.

Definition (Stochastic integral). $H \cdot M$ is the *stochastic integral* of H with respect to M and we also write

$$(H \cdot M)_t = \int_0^t H_s \, dM_s.$$

It is important that the integral of martingale is still a martingale. After proving Itô's formula, we will use this fact to show that a lot of things are in fact martingales in a rather systematic manner. For example, it will be rather effortless to show that $B_t^2 - t$ is a martingale when B_t is a standard Brownian motion.

Proof.

- (i) We have already shown that this map is an isometry when restricted to \mathcal{E} . So extend by completeness of \mathcal{M}_c^2 and denseness of \mathcal{E} .
- (ii) Again the equation to show is known for simple H , and we want to show it is preserved under taking limits. Suppose $H^n \rightarrow H$ in $L^2(M)$ with $H^n \in L^2(M)$. Then $H^n \cdot M \rightarrow H \cdot M$ in \mathcal{M}_c^2 . We want to show that

$$\begin{aligned} \langle H \cdot M, N \rangle_\infty &= \lim_{n \rightarrow \infty} \langle H^n \cdot M, N \rangle_\infty \text{ in } L^1. \\ H \cdot \langle M, N \rangle &= \lim_{n \rightarrow \infty} H^n \cdot \langle M, N \rangle \text{ in } L^1. \end{aligned}$$

for all $N \in \mathcal{M}_c^2$.

To show the first holds, we use the Kunita–Watanabe inequality to get

$$\mathbb{E}|\langle H \cdot M - H^n \cdot M, N \rangle_\infty| \leq \mathbb{E}(\langle H \cdot M - H^n \cdot M \rangle_\infty)^{1/2} (\mathbb{E}\langle N \rangle_\infty)^{1/2},$$

and the first factor is $\|H \cdot M - H^n \cdot M\|_{\mathcal{M}^2} \rightarrow 0$, while the second is finite since $N \in \mathcal{M}_c^2$. The second follows from

$$\mathbb{E}|((H - H^n) \cdot \langle M, N \rangle)_\infty| \leq \|H - H^n\|_{L^2(M)} \|N\|_{\mathcal{M}^2} \rightarrow 0.$$

So we know that $\langle H \cdot M, N \rangle_\infty = (H \cdot \langle M, N \rangle)_\infty$. We can then replace N by the stopped process N^t to get $\langle H \cdot M, N \rangle_t = (H \cdot \langle M, N \rangle)_t$.

To see uniqueness, suppose $X \in \mathcal{M}_c^2$ is another such martingale. Then we have $\langle X - H \cdot M, N \rangle = 0$ for all N . Take $N = X - H \cdot M$, and then we are done.

(iii) For $N \in \mathcal{M}_c^2$, we have

$$\langle (H \cdot M)^T, N \rangle_t = \langle H \cdot M, N \rangle_{t \wedge T} = H \cdot \langle M, N \rangle_{t \wedge T} = (H1_{[0,T]} \cdot \langle M, N \rangle)_t$$

for every N . So we have shown that

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

by (ii). To prove the second equality, we have

$$\langle H \cdot M^T, N \rangle_t = H \cdot \langle M^T, N \rangle_t = H \cdot \langle M, N \rangle_{t \wedge T} = ((H1_{[0,T]} \cdot \langle M, N \rangle)_t). \quad \square$$

Note that (ii) can be written as

$$\left\langle \int_0^{(-)} H_s \, dM_s, N \right\rangle_t = \int_0^t H_s \, d\langle M, N \rangle_s.$$

Corollary.

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

In other words,

$$\left\langle \int_0^{(-)} H_s \, dM_s, \int_0^{(-)} K_s \, dN_s \right\rangle_t = \int_0^t H_s K_s \, d\langle M, N \rangle_s. \quad \square$$

Corollary. Since $H \cdot M$ and $(H \cdot M)(K \cdot N) - \langle H \cdot M, K \cdot N \rangle$ are martingales starting at 0, we have

$$\begin{aligned} \mathbb{E} \left(\int_0^t H \, dM_s \right) &= 0 \\ \mathbb{E} \left(\left(\int_0^t H_s \, dM_s \right) \left(\int_0^t K_s \, dN_s \right) \right) &= \int_0^t H_s K_s \, d\langle M, N \rangle_s. \quad \square \end{aligned}$$

Corollary. Let $H \in L^2(M)$, then $HK \in L^2(M)$ iff $K \in L^2(H \cdot M)$, in which case

$$(KH) \cdot M = K \cdot (H \cdot M).$$

Proof. We have

$$\mathbb{E} \left(\int_0^\infty K_s^2 H_s^2 \, d\langle M_s \rangle \right) = \mathbb{E} \left(\int_0^\infty K_s^2 \langle H \cdot M \rangle_s \right),$$

so $\|K\|_{L^2(H \cdot M)} = \|HK\|_{L^2(M)}$. For $N \in \mathcal{M}_c^2$, we have

$$\langle (KH) \cdot M, N \rangle_t = (KH \cdot \langle M, N \rangle)_t = (K \cdot (H \cdot \langle M, N \rangle))_t = (K \cdot \langle H \cdot M, N \rangle)_t. \quad \square$$

3.3 Extension to local martingales

We have now defined the stochastic integral for continuous martingales. We next go through some formalities to extend this to local martingales, and ultimately to semi-martingales. We are not doing this just for fun. Rather, when we later prove results like Itô's formula, even when we put in continuous (local) martingales, we usually end up with some semi-martingales. So it is useful to be able to deal with semi-martingales in general.

Definition ($L_{bc}^2(M)$). Let $L_{bc}^2(M)$ be the space of previsible H such that

$$\int_0^t H_s^2 d\langle M \rangle_s < \infty \text{ a.s.}$$

for all finite $t > 0$.

Theorem. Let M be a continuous local martingale.

- (i) For every $H \in L_{bc}^2(M)$, there is a unique continuous local martingale $H \cdot M$ with $(H \cdot M)_0 = 0$ and

$$\langle H \cdot M, N \rangle = H \cdot \langle M, N \rangle$$

for all N, M .

- (ii) If T is a stopping time, then

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

- (iii) If $H \in L_{loc}^2(M)$, K is previsible, then $K \in L_{loc}^2(H \cdot M)$ iff $HK \in L_{loc}^2(M)$, and then

$$K \cdot (H \cdot M) = (KH) \cdot M.$$

- (iv) Finally, if $M \in \mathcal{M}_c^2$ and $H \in L^2(M)$, then the definition is the same as the previous one.

Proof. Assume $M_0 = 0$, and that $\int_0^t H_s^2 d\langle M \rangle_s < \infty$ for all $\omega \in \Omega$ (by setting $H = 0$ when this fails). Set

$$S_n = \inf \left\{ t \geq 0 : \int_0^t (1 + H_s^2) d\langle M \rangle_s \geq n \right\}.$$

These S_n are stopping times that tend to infinity. Then

$$\langle M^{S_n}, M^{S_n} \rangle_t = \langle M, M \rangle_{t \wedge S_n} \leq n.$$

So $M^{S_n} \in \mathcal{M}_c^2$. Also,

$$\int_0^\infty H_s d\langle M^{S_n} \rangle_s = \int_0^{S_n} H_s^2 d\langle M \rangle_s \leq n.$$

So $H \in L^2(M^{S_n})$, and we have already defined what $H \cdot M^{S_n}$ is. Now notice that

$$H \cdot M^{S_n} = (H \cdot M^{S_m})^{S_n} \text{ for } m \geq n.$$

So it makes sense to define

$$H \cdot M = \lim_{n \rightarrow \infty} H \cdot M^{S_n}.$$

This is the unique process such that $(H \cdot M)^{S_n} = H \cdot M^{S_n}$. We see that $H \cdot M$ is a continuous adapted local martingale with reducing sequence S_n .

Claim. $\langle H \cdot M, N \rangle = H \cdot \langle M, N \rangle$.

Indeed, assume that $N_0 = 0$. Set $S'_n = \inf\{t \geq 0 : |N_t| \geq n\}$. Set $T_n = S_n \wedge S'_n$. Observe that $N^{S'_n} \in \mathcal{M}_c^2$. Then

$$\langle H \cdot M, N \rangle^{T_n} = \langle H \cdot M^{S_n}, N^{S'_n} \rangle = H \cdot \langle M^{S_n}, N^{S'_n} \rangle = H \cdot \langle M, N \rangle^{T_n}.$$

Taking the limit $n \rightarrow \infty$ gives the desired result.

The proofs of the other claims are the same as before, since they only use the characterizing property $\langle H \cdot M, N \rangle = H \cdot \langle M, N \rangle$. \square

3.4 Extension to semi-martingales

Definition (Locally bounded previsible process). A previsible process H is *locally bounded* if for all $t \geq 0$, we have

$$\sup_{s \leq t} |H_s| < \infty \text{ a.s.}$$

Fact.

- (i) Any adapted continuous process is locally bounded.
- (ii) If H is locally bounded and A is a finite variation process, then for all $t \geq 0$, we have

$$\int_0^t |H_s| |dA_s| < \infty \text{ a.s.}$$

Now if $X = X_0 + M + A$ is a semi-martingale, where $X_0 \in \mathcal{F}_0$, M is a continuous local martingale and A is a finite variation process, we want to define $\int H_s dX_s$. We already know what it means to define integration with respect to dM_s and dA_s , using the Itô integral and the finite variation integral respectively, and X_0 doesn't change, so we can ignore it.

Definition (Stochastic integral). Let $X = X_0 + M + A$ be a continuous semi-martingale, and H a locally bounded previsible process. Then the *stochastic integral* $H \cdot X$ is the continuous semi-martingale defined by

$$H \cdot X = H \cdot M + H \cdot A,$$

and we write

$$(H \cdot X)_t = \int_0^t H_s dX_s.$$

Proposition.

- (i) $(H, X) \mapsto H \cdot X$ is bilinear.
- (ii) $H \cdot (K \cdot X) = (HK) \cdot X$ if H and K are locally bounded.
- (iii) $(H \cdot X)^T = H \mathbf{1}_{[0, T]} \cdot X = H \cdot X^T$ for every stopping time T .
- (iv) If X is a continuous local martingale (resp. a finite variation process), then so is $H \cdot X$.
- (v) If $H = \sum_{i=1}^n H_{i-1} \mathbf{1}_{(t_{i-1}, t_i]}$ and $H_{i-1} \in \mathcal{F}_{t_{i-1}}$ (not necessarily bounded), then

$$(H \cdot X)_t = \sum_{i=1}^n H_{i-1} (X_{t_i \wedge t} - X_{t_{i-1} \wedge t}).$$

Proof. (i) to (iv) follow from analogous properties for $H \cdot M$ and $H \cdot A$. The last part is also true by definition if the H_i are uniformly bounded. If H_i is not bounded, then the finite variation part is still fine, since for each fixed $\omega \in \Omega$, $H_i(\omega)$ is a fixed number. For the martingale part, set

$$T_n = \inf\{t \geq 0 : |H_t| \geq n\}.$$

Then T_n are stopping times, $T_n \rightarrow \infty$, and $H1_{[0, T_n]} \in \mathcal{E}$. Thus

$$(H \cdot M)_{t \wedge T_n} = \sum_{i=1}^n H_{i-1} T_{[0, T_n]}(X_{t_i \wedge t} - X_{t_{i-1} \wedge t}).$$

Then take the limit $n \rightarrow \infty$. \square

Before we get to Itô's formula, we need a few more useful properties:

Proposition (Stochastic dominated convergence theorem). Let X be a continuous semi-martingale. Let H, H_s be previsible and locally bounded, and let K be previsible and non-negative. Let $t > 0$. Suppose

- (i) $H_s^n \rightarrow H_s$ as $n \rightarrow \infty$ for all $s \in [0, t]$.
- (ii) $|H_s^n| \leq K_s$ for all $s \in [0, t]$ and $n \in \mathbb{N}$.
- (iii) $\int_0^t K_s^2 d\langle M \rangle_s < \infty$ and $\int_0^t K_s |dA_s| < \infty$ (note that both conditions are okay if K is locally bounded).

Then

$$\int_0^t H_s^n dX_s \rightarrow \int_0^t H_s dX_s \text{ in probability.}$$

Proof. For the finite variation part, the convergence follows from the usual dominated convergence theorem. For the martingale part, we set

$$T_m = \inf \left\{ t \geq 0 : \int_0^t K_s^2 d\langle M \rangle_s \geq m \right\}.$$

So we have

$$\mathbb{E} \left(\left(\int_0^{T_m \wedge t} H_s^n dM_s - \int_0^{T_m \wedge t} H_s dM_s \right)^2 \right) \leq \mathbb{E} \left(\int_0^{T_m \wedge t} (H_s^n - H_s)^2 d\langle M \rangle_s \right) \rightarrow 0.$$

using the usual dominated convergence theorem, since $\int_0^{T_m \wedge t} K_s^2 d\langle M \rangle_s \leq m$.

Since $T_n \wedge t = t$ eventually as $n \rightarrow \infty$ almost surely, hence in probability, we are done. \square

Proposition. Let X be a continuous semi-martingale, and let H be an adapted bounded left-continuous process. Then for every subdivision $0 < t_0^{(m)} < t_1^{(m)} < \dots < t_{n_m}^{(m)}$ of $[0, t]$ with $\max_i |t_i^{(m)} - t_{i-1}^{(m)}| \rightarrow 0$, then

$$\int_0^t H_s dX_s = \lim_{m \rightarrow \infty} \sum_{i=1}^{n_m} H_{t_{i-1}^{(m)}} (X_{t_i^{(m)}} - X_{t_{i-1}^{(m)}})$$

in probability.

Proof. We have already proved this for the Lebesgue–Stieltjes integral, and all we used was dominated convergence. So the same proof works using stochastic dominated convergence theorem. \square

3.5 Itô formula

We now prove the equivalent of the integration by parts and the chain rule, i.e. Itô's formula. Compared to the world of usual integrals, the difference is that the quadratic variation, i.e. "second order terms" will crop up quite a lot, since they are no longer negligible.

Theorem (Integration by parts). Let X, Y be a continuous semi-martingale. Then almost surely,

$$X_t Y_t - X_0 Y_0 = \int_0^t X_s dY_s + \int_0^t Y_s dX_s + \langle X, Y \rangle_t$$

The last term is called the *Itô correction*.

Note that if X, Y are martingales, then the first two terms on the right are martingales, but the last is not. So we are forced to think about semi-martingales.

Observe that in the case of finite variation integrals, we don't have the correction.

Proof. We have

$$X_t Y_t - X_s Y_s = X_s (Y_t - Y_s) + (X_t - X_s) Y_s + (X_t - X_s)(Y_t - Y_s).$$

When doing usual calculus, we can drop the last term, because it is second order. However, the quadratic variation of martingales is in general non-zero, and so we must keep track of this. We have

$$\begin{aligned} X_{k2^{-n}} Y_{k2^{-n}} - X_0 Y_0 &= \sum_{i=1}^k (X_{i2^{-n}} Y_{i2^{-n}} - X_{(i-1)2^{-n}} Y_{(i-1)2^{-n}}) \\ &= \sum_{i=1}^n \left(X_{(i-1)2^{-n}} (Y_{i2^{-n}} - Y_{(i-1)2^{-n}}) \right. \\ &\quad \left. + Y_{(i-1)2^{-n}} (X_{i2^{-n}} - X_{(i-1)2^{-n}}) \right. \\ &\quad \left. + (X_{i2^{-n}} - X_{(i-1)2^{-n}}) (Y_{i2^{-n}} - Y_{(i-1)2^{-n}}) \right) \end{aligned}$$

Taking the limit $n \rightarrow \infty$ with $k2^{-n}$ fixed, we see that the formula holds for t a dyadic rational. Then by continuity, it holds for all t . \square

The really useful formula is the following:

Theorem (Itô's formula). Let X^1, \dots, X^p be continuous semi-martingales, and let $f : \mathbb{R}^p \rightarrow \mathbb{R}$ be C^2 . Then, writing $X = (X^1, \dots, X^p)$, we have, almost surely,

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

In particular, $f(X)$ is a semi-martingale.

The proof is long but not hard. We first do it for polynomials by explicit computation, and then use Weierstrass approximation to extend it to more general functions.

Proof.

Claim. Itô's formula holds when f is a polynomial.

It clearly does when f is a constant! We then proceed by induction. Suppose Itô's formula holds for some f . Then we apply integration by parts to

$$g(x) = x^k f(x).$$

where x^k denotes the k th component of x . Then we have

$$g(X_t) = g(X_0) + \int_0^t X_s^k df(X_s) + \int_0^t f(X_s) dX_s^k + \langle X^k, f(X) \rangle_t$$

We now apply Itô's formula for f to write

$$\begin{aligned} \int_0^t X_s^k df(X_s) &= \sum_{i=1}^p \int_0^t X_s^k \frac{\partial f}{\partial x^i}(X_s) dX_s^i \\ &\quad + \frac{1}{2} \sum_{i,j=1}^p \int_0^t X_s^k \frac{\partial^2 f}{\partial x^i \partial x^j}(X_s) d\langle X^i, X^j \rangle_s. \end{aligned}$$

We also have

$$\langle X^k, f(X) \rangle_t = \sum_{i=1}^p \int_0^t \frac{\partial f}{\partial x^i}(X_s) d\langle X^k, X^i \rangle_s.$$

So we have

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

So by induction, Itô's formula holds for all polynomials.

Claim. Itô's formula holds for all $f \in C^2$ if $|X_t(\omega)| \leq n$ and $\int_0^t |dA_s| \leq n$ for all (t, ω) .

By the Weierstrass approximation theorem, there are polynomials p_k such that

$$\sup_{|x| \leq k} \left(|f(x) - p_k(x)| + \max_i \left| \frac{\partial f}{\partial x^i} - \frac{\partial p_k}{\partial x^i} \right| + \max_{i,j} \left| \frac{\partial^2 f}{\partial x^i \partial x^j} - \frac{\partial^2 p_k}{\partial x^i \partial x^j} \right| \right) \leq \frac{1}{k}.$$

By taking limits, in probability, we have

$$\begin{aligned} f(X_t) - f(X_0) &= \lim_{k \rightarrow \infty} (p_k(X_t) - p_k(X_0)) \\ \int_0^t \frac{\partial f}{\partial x^i}(X_s) dX_s^i &= \lim_{k \rightarrow \infty} \int_0^t \frac{\partial p_k}{\partial x^i}(X_s) dX_s^i \end{aligned}$$

by stochastic dominated convergence theorem, and by the regular dominated convergence, we have

$$\int_0^t \frac{\partial f}{\partial x^i \partial x^j} d\langle X^i, X^j \rangle_s = \lim_{k \rightarrow \infty} \int_0^t \frac{\partial^2 p_k}{\partial x^i \partial x^j} d\langle X^i, X^j \rangle_s.$$

Claim. Itô's formula holds for all X .

Let

$$T_n = \inf \left\{ t \geq 0 : |X_t| \geq n \text{ or } \int_0^t |dA_s| \geq n \right\}$$

Then by the previous claim, we have

$$\begin{aligned} f(X_t^{T_n}) &= f(X_0) + \sum_{i=1}^p \int_0^t \frac{\partial f}{\partial x^i}(X_s^{T_n}) d(X_i)_s^{T_n} \\ &\quad + \frac{1}{2} \sum_{i,j} \int_0^t \frac{\partial^2 f}{\partial x^i \partial x^j}(X_s^{T_n}) d\langle (X_i)^{T_n}, (X^j)^{T_n} \rangle_s \\ &= f(X_0) + \sum_{i=1}^p \int_0^{t \wedge T_n} \frac{\partial f}{\partial x^i}(X_s) d(X_i)_s \\ &\quad + \frac{1}{2} \sum_{i,j} \int_0^{t \wedge T_n} \frac{\partial^2 f}{\partial x^i \partial x^j}(X_s) d\langle (X_i), (X^j) \rangle_s. \end{aligned}$$

Then take $T_n \rightarrow \infty$. □

Example. Let B be a standard Brownian motion, $B_0 = 0$ and $f(x) = x^2$. Then

$$B_t^2 = 2 \int_0^t B_s dB_s + t.$$

In other words,

$$B_t^2 - t = 2 \int_0^t B_s dB_s.$$

In particular, this is a continuous local martingale.

Example. Let $B = (B^1, \dots, B^d)$ be a d -dimensional Brownian motion. Then we apply Itô's formula to the semi-martingale $X = (t, B^1, \dots, B^d)$. Then we find that

$$f(t, B_t) - f(0, B_0) - \int_0^t \left(\frac{\partial}{\partial s} + \frac{1}{2} \Delta \right) f(s, B_s) ds = \sum_{i=1}^d \int_0^t \frac{\partial}{\partial x^i} f(s, B_s) dB_s^i$$

is a continuous local martingale.

There are some syntactic tricks that make stochastic integrals easier to manipulate, namely by working in differential form. We can state Itô's formula in differential form

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

which we can think of as the chain rule. For example, in the case of Brownian motion, we have

$$df(B_t) = f'(B_t) dB_t + \frac{1}{2} f''(B_t) dt.$$

Formally, one expands f using that that “ $(dt)^2 = 0$ ” but “ $(dB)^2 = dt$ ”. The following formal rules hold:

$$\begin{aligned} Z_t - Z_0 &= \int_0^t H_s dX_s \iff dZ_t = H_t dX_t \\ Z_t &= \langle X, Y \rangle_t = \int_0^t d\langle X, Y \rangle_t \iff dZ_t = dX_t dY_t. \end{aligned}$$

Then we have rules such as

$$\begin{aligned} H_t(K_t dX_t) &= (H_t K_t) dX_t \\ H_t(dX_t dY_t) &= (H_t dX_t) dY_t \\ d(X_t Y_t) &= X_t dY_t + Y_t dX_t + dX_t dY_t \\ df(X_t) &= f'(X_t) dX_t + \frac{1}{2} f''(X_t) dX_t dX_t. \end{aligned}$$

3.6 The Lévy characterization

A more major application of the stochastic integral is the following convenient characterization of Brownian motion:

Theorem (Lévy’s characterization of Brownian motion). Let (X^1, \dots, X^d) be continuous local martingales. Suppose that $X_0 = 0$ and that $\langle X^i, X^j \rangle_t = \delta_{ij}t$ for all $i, j = 1, \dots, d$ and $t \geq 0$. Then (X^1, \dots, X^d) is a standard d -dimensional Brownian motion.

This might seem like a rather artificial condition, but it turns out to be quite useful in practice (though less so in this course). The point is that we know that $\langle H \cdot M \rangle_t = H_t^2 \cdot \langle M \rangle_t$, and in particular if we are integrating things with respect to Brownian motions of some sort, we know $\langle B_t \rangle_t = t$, and so we are left with some explicit, familiar integral to do.

Proof. Let $0 \leq s < t$. It suffices to check that $X_t - X_s$ is independent of \mathcal{F}_s and $X_t - X_s \sim N(0, (t-s)I)$.

Claim. $\mathbb{E}(e^{i\theta \cdot (X_t - X_s)} \mid \mathcal{F}_s) = e^{-\frac{1}{2}|\theta|^2(t-s)}$ for all $\theta \in \mathbb{R}^d$ and $s < t$.

This is sufficient, since the right-hand side is independent of \mathcal{F}_s , hence so is the left-hand side, and the Fourier transform characterizes the distribution.

To check this, for $\theta \in \mathbb{R}^d$, we define

$$Y_t = \theta \cdot X_t = \sum_{i=1}^d \theta^i X_t^i.$$

Then Y is a continuous local martingale, and we have

$$\langle Y \rangle_t = \langle Y, Y \rangle_t = \sum_{i,j=1}^d \theta^j \theta^k \langle X^j, X^k \rangle_t = |\theta|^2 t.$$

by assumption. Let

$$Z_t = e^{iY_t + \frac{1}{2}\langle Y \rangle_t} = e^{i\theta \cdot X_t + \frac{1}{2}|\theta|^2 t}.$$

By Itô's formula, with $X = iY + \frac{1}{2}\langle Y \rangle_t$ and $f(x) = e^x$, we get

$$dZ_t = Z_t \left(idY_t - \frac{1}{2}d\langle Y \rangle_t + \frac{1}{2}d\langle Y \rangle_t \right) = iZ_t dY_t.$$

So this implies Z is a continuous local martingale. Moreover, since Z is bounded on bounded intervals of t , we know Z is in fact a martingale, and $Z_0 = 1$. Then by definition of a martingale, we have

$$\mathbb{E}(Z_t | \mathcal{F}_s) = Z_s,$$

And unwrapping the definition of Z_t shows that the result follows. \square

In general, the quadratic variation of a process doesn't have to be linear in t . It turns out if the quadratic variation increases to infinity, then the martingale is still a Brownian motion up to reparametrization.

Theorem (Dubins–Schwarz). Let M be a continuous local martingale with $M_0 = 0$ and $\langle M \rangle_\infty = \infty$. Let

$$T_s = \inf\{t \geq 0 : \langle M \rangle_t > s\},$$

the right-continuous inverse of $\langle M \rangle_t$. Let $B_s = M_{T_s}$ and $\mathcal{G}_s = \mathcal{F}_{T_s}$. Then T_s is a (\mathcal{F}_t) stopping time, $\langle M \rangle_{T_s} = s$ for all $s \geq 0$, B is a (\mathcal{G}_s) -Brownian motion, and

$$M_t = B_{\langle M \rangle_t}.$$

Proof. Since $\langle M \rangle$ is continuous and adapted, and $\langle M \rangle_\infty = \infty$, we know T_s is a stopping time and $T_s < \infty$ for all $s \geq 0$.

Claim. (\mathcal{G}_s) is a filtration obeying the usual conditions, and $\mathcal{G}_\infty = \mathcal{F}_\infty$

Indeed, if $A \in \mathcal{G}_s$ and $s < t$, then

$$A \cap \{T_t \leq u\} = A \cap \{T_s \leq u\} \cap \{T_t \leq u\} \in \mathcal{F}_u,$$

using that $A \cap \{T_s \leq u\} \in \mathcal{F}_u$ since $A \in \mathcal{G}_s$. Then right-continuity follows from that of (\mathcal{F}_t) and the right-continuity of $s \mapsto T_s$.

Claim. B is adapted to (\mathcal{G}_s) .

In general, if X is càdlàg and T is a stopping time, then $X_T \mathbf{1}_{\{T < \infty\}} \in \mathcal{F}_T$. Apply this with $X = M$, $T = T_s$ and $\mathcal{F}_T = \mathcal{G}_s$. Thus $B_s \in \mathcal{G}_s$.

Claim. B is continuous.

Here this is actually something to verify, because $s \mapsto T_s$ is only right continuous, not necessarily continuous. Thus, we only know B_s is right continuous, and we have to check it is left continuous.

Now B is left-continuous at s iff $B_s = B_{s-}$, iff $M_{T_s} = M_{T_{s-}}$. Now we have

$$T_{s-} = \inf\{t \geq 0 : \langle M \rangle_t \geq s\}.$$

If $T_s = T_{s-}$, then there is nothing to show. Thus, we may assume $T_s > T_{s-}$. Then we have $\langle M \rangle_{T_s} = \langle M \rangle_{T_{s-}}$. Since $\langle M \rangle_t$ is increasing, it means $\langle M \rangle_t$ is constant in $[T_{s-}, T_s]$. We will later prove that

Lemma. M is constant on $[a, b]$ iff $\langle M \rangle$ being constant on $[a, b]$.

So we know that if $T_s > T_{s-}$, then $M_{T_s} = M_{T_{s-}}$. So B is left continuous.

We then have to show that B is a martingale.

Claim. $(M^2 - \langle M \rangle)^{T_s}$ is a uniformly integrable martingale.

To see this, observe that $\langle M^{T_s} \rangle_\infty = \langle M \rangle_{T_s} = s$, and so M^{T_s} is bounded. So $(M^2 - \langle M \rangle)^{T_s}$ is a uniformly integrable martingale.

We now apply the optional stopping theorem, which tells us

$$\mathbb{E}(B_s | \mathcal{G}_r) = \mathbb{E}(M_\infty^{T_s} | \mathcal{G}_s) = M_{T_t} = B_t.$$

So B_t is a martingale. Moreover,

$$\mathbb{E}(B_s^2 - s | \mathcal{G}_r) = \mathbb{E}((M^2 - \langle M \rangle)^{T_s} | \mathcal{F}_{T_r}) = M_{T_r}^2 - \langle M \rangle_{T_r} = B_r^2 - r.$$

So $B_t^2 - t$ is a martingale, so by the characterizing property of the quadratic variation, $\langle B \rangle_t = t$. So by Lévy's criterion, this is a Brownian motion in one dimension. \square

The theorem is only true for martingales in one dimension. In two dimensions, this need not be true, because the time change needed for the horizontal and vertical may not agree. However, in the example sheet, we see that the holomorphic image of a Brownian motion is still a Brownian motion up to a time change.

Lemma. M is constant on $[a, b]$ iff $\langle M \rangle$ being constant on $[a, b]$.

Proof. It is clear that if M is constant, then so is $\langle M \rangle$. To prove the converse, by continuity, it suffices to prove that for any fixed $a < b$,

$$\{M_t = M_a \text{ for all } t \in [a, b]\} \supseteq \{\langle M \rangle_b = \langle M \rangle_a\} \text{ almost surely.}$$

We set $N_t = M_t - M_t \wedge a$. Then $\langle N \rangle_t = \langle M \rangle_t - \langle M \rangle_{t \wedge a}$. Define

$$T_\varepsilon = \inf\{t \geq 0 : \langle N \rangle_t \geq \varepsilon\}.$$

Then since $N^2 - \langle N \rangle$ is a local martingale, we know that

$$\mathbb{E}(N_{t \wedge T_\varepsilon}^2) = \mathbb{E}(\langle N \rangle_{t \wedge T_\varepsilon}) \leq \varepsilon.$$

Now observe that on the event $\{\langle M \rangle_b = \langle M \rangle_a\}$, we have $\langle N \rangle_b = 0$. So for $t \in [a, b]$, we have

$$\mathbb{E}(1_{\{\langle M \rangle_b = \langle M \rangle_a\}} N_t^2) = \mathbb{E}(1_{\{\langle M \rangle_b = \langle M \rangle_a\}} N_{t \wedge T_\varepsilon}^2) = \mathbb{E}(\langle N \rangle_{t \wedge T_\varepsilon}) = 0. \quad \square$$

3.7 Girsanov's theorem

Girsanov's theorem tells us what happens to our (semi)-martingales when we change the measure of our space. We first look at a simple example when we perform a shift.

Example. Let $X \sim N(0, C)$ be an n -dimensional centered Gaussian with positive definite covariance $C = (C_{ij})_{i,j=1}^n$. Put $M = C^{-1}$. Then for any function f , we have

$$\mathbb{E}f(X) = \left(\det \frac{M}{2\pi}\right)^{1/2} \int_{\mathbb{R}^n} f(x) e^{-\frac{1}{2}(x, Mx)} dx.$$

Now fix an $a \in \mathbb{R}^n$. The distribution of $X + a$ then satisfies

$$\mathbb{E}f(X + a) = \left(\det \frac{M}{2\pi}\right)^{1/2} \int_{\mathbb{R}^n} f(x) e^{-\frac{1}{2}(x-a, M(x-a))} dx = \mathbb{E}[Zf(X)],$$

where

$$Z = Z(x) = e^{-\frac{1}{2}(a, Ma) + (x, Ma)}.$$

Thus, if \mathbb{P} denotes the distribution of X , then the measure \mathbb{Q} with

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = Z$$

is that of $N(a, C)$ vector.

Example. We can extend the above example to Brownian motion. Let B be a Brownian motion with $B_0 = 0$, and $h : [0, \infty) \rightarrow \mathbb{R}$ a deterministic function. We then want to understand the distribution of $B_t + h$.

Fix a finite sequence of times $0 = t_0 < t_1 < \dots < t_n$. Then we know that $(B_{t_i})_{i=1}^n$ is a centered Gaussian random variable. Thus, if $f(B) = f(B_{t_1}, \dots, B_{t_n})$ is a function, then

$$\mathbb{E}(f(B)) = c \cdot \int_{\mathbb{R}^n} f(x) e^{-\frac{1}{2} \sum_{i=1}^n \frac{(x_i - x_{i-1})^2}{t_i - t_{i-1}}} dx_1 \cdots dx_n.$$

Thus, after a shift, we get

$$\begin{aligned} \mathbb{E}(f(B + h)) &= \mathbb{E}(Zf(B)), \\ Z &= \exp \left(-\frac{1}{2} \sum_{i=1}^n \frac{(h_{t_i} - h_{t_{i-1}})^2}{t_i - t_{i-1}} + \sum_{i=1}^n \frac{(h_{t_i} - h_{t_{i-1}})(B_{t_i} - B_{t_{i-1}})}{t_i - t_{i-1}} \right). \end{aligned}$$

In general, we are interested in what happens when we change the measure by an exponential:

Definition (Stochastic exponential). Let M be a continuous local martingale. Then the *stochastic exponential* (or *Doléans–Dade exponential*) of M is

$$\mathcal{E}(M)_t = e^{M_t - \frac{1}{2}\langle M \rangle_t}$$

The point of introducing that quadratic variation term is

Proposition. Let M be a continuous local martingale with $M_0 = 0$. Then $\mathcal{E}(M) = Z$ satisfies

$$dZ_t = Z_t dM,$$

i.e.

$$Z_t = 1 + \int_0^t Z_s dM_s.$$

In particular, $\mathcal{E}(M)$ is a continuous local martingale. Moreover, if $\langle M \rangle$ is uniformly bounded, then $\mathcal{E}(M)$ is a uniformly integrable martingale.

There is a more general condition for the final property, namely Novikov's condition, but we will not go into that.

Proof. By Itô's formula with $X = M - \frac{1}{2}\langle M \rangle$, we have

$$dZ_t = Z_t d\left(M_t - \frac{1}{2}d\langle M \rangle_t\right) + \frac{1}{2}Z_t d\langle M \rangle_t = Z_t dM_t.$$

Since M is a continuous local martingale, so is $\int Z_s dM_s$. So Z is a continuous local martingale.

Now suppose $\langle M \rangle_\infty \leq b < \infty$. Then

$$\mathbb{P}\left(\sup_{t \geq 0} M_t \geq a\right) = \mathbb{P}\left(\sup_{t \geq 0} M_t \geq a, \langle M \rangle_\infty \leq b\right) \leq e^{-a^2/2b},$$

where the final equality is an exercise on the third example sheet, which is true for general continuous local martingales. So we get

$$\begin{aligned} \mathbb{E}\left(\exp\left(\sup_t M_t\right)\right) &= \int_0^\infty \mathbb{P}(\exp(\sup M_t) \geq \lambda) d\lambda \\ &= \int_0^\infty \mathbb{P}(\sup M_t \geq \log \lambda) d\lambda \\ &\leq 1 + \int_1^\infty e^{-(\log \lambda)^2/2b} d\lambda < \infty. \end{aligned}$$

Since $\langle M \rangle \geq 0$, we know that

$$\sup_{t \geq 0} \mathcal{E}(M)_t \leq \exp(\sup M_t),$$

So $\mathcal{E}(M)$ is a uniformly integrable martingale. \square

Theorem (Girsanov's theorem). Let M be a continuous local martingale with $M_0 = 0$. Suppose that $\mathcal{E}(M)$ is a uniformly integrable martingale. Define a new probability measure

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \mathcal{E}(M)_\infty$$

Let X be a continuous local martingale with respect to \mathbb{P} . Then $X - \langle X, M \rangle$ is a continuous local martingale with respect to \mathbb{Q} .

Proof. Define the stopping time

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

and $\mathbb{P}(T_n \rightarrow \infty) = 1$ by continuity. Since \mathbb{Q} is absolutely continuous with respect to \mathbb{P} , we know that $\mathbb{Q}(T_n \rightarrow \infty) = 1$. Thus it suffices to show that $X^{T_n} - \langle X^{T_n}, M \rangle$ is a continuous martingale for any n . Let

$$Y = X^{T_n} - \langle X^{T_n}, M \rangle, \quad Z = \mathcal{E}(M).$$

Claim. ZY is a continuous local martingale with respect to \mathbb{P} .

We use the product rule to compute

$$\begin{aligned} d(ZY) &= Y_t dZ_t + Z_t dY_t + d\langle Y, Z \rangle_t \\ &= Y Z_t dM_t + Z_t (dX^{T_n} - d\langle X^{T_n}, M \rangle_t) + Z_t d\langle M, X^{T_n} \rangle \\ &= Y Z_t dM_t + Z_t dX^{T_n} \end{aligned}$$

So we see that ZY is a stochastic integral with respect to a continuous local martingale. Thus ZY is a continuous local martingale.

Claim. ZY is uniformly integrable.

Since Z is a uniformly integrable martingale, $\{Z_T : T \text{ is a stopping time}\}$ is uniformly integrable. Since Y is bounded, $\{Z_T Y_T : T \text{ is a stopping time}\}$ is also uniformly integrable. So ZY is a true martingale (with respect to \mathbb{P}).

Claim. Y is a martingale with respect to \mathbb{Q} .

We have

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}}(Y_t - Y_s \mid \mathcal{F}_s) &= \mathbb{E}^{\mathbb{P}}(Z_{\infty} Y_t - Z_{\infty} Y_s \mid \mathcal{F}_s) \\ &= \mathbb{E}^{\mathbb{P}}(Z_t Y_t - Z_s Y_s \mid \mathcal{F}_s) = 0. \end{aligned} \quad \square$$

Note that the quadratic variation does not change since

$$\langle X - \langle X, M \rangle \rangle = \langle X \rangle_t = \lim_{n \rightarrow \infty} \sum_{i=1}^{\lfloor 2^n t \rfloor} (X_{i2^{-n}} - X_{(i-1)2^{-n}})^2 \text{ a.s.}$$

along a subsequence.

4 Stochastic differential equations

4.1 Existence and uniqueness of solutions

After all this work, we can return to the problem we described in the introduction. We wanted to make sense of equations of the form

$$\dot{x}(t) = F(x(t)) + \eta(t),$$

where $\eta(t)$ is Gaussian white noise. We can now interpret this equation as saying

$$dX_t = F(X_t) dt + dB_t,$$

or equivalently, in integral form,

$$X_t - X_0 = \int_0^t F(X_s) ds + B_t.$$

In general, we can make the following definition:

Definition (Stochastic differential equation). Let $d, m \in \mathbb{N}$, $b : \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}^d$, $\sigma : \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times m}$ be locally bounded (and measurable). A solution to the stochastic differential equation $E(\sigma, b)$ given by

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

consists of

- (i) a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ obeying the usual conditions;
- (ii) an m -dimensional Brownian motion B with $B_0 = 0$; and
- (iii) an (\mathcal{F}_t) -adapted continuous process X with values in \mathbb{R}^d such that

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

If $X_0 = x \in \mathbb{R}^d$, then we say X is a (*weak*) *solution* to $E_x(\sigma, b)$. It is a *strong* solution if it is adapted with respect to the canonical filtration of B .

Our goal is to prove existence and uniqueness of solutions to a general class of SDEs. We already know what it means for solutions to be unique, and in general there can be multiple notions of uniqueness:

Definition (Uniqueness of solutions). For the stochastic differential equation $E(\sigma, b)$, we say there is

- *uniqueness in law* if for every $x \in \mathbb{R}^d$, all solutions to $E_x(\sigma, b)$ have the same distribution.
- *pathwise uniqueness* if when $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ and B are fixed, any two solutions X, X' with $X_0 = X'_0$ are indistinguishable.

These two notions are not equivalent, as the following example shows:

Example (Tanaka). Consider the stochastic differential equation

$$dX_t = \operatorname{sgn}(X_t) dB_t, \quad X_0 = x,$$

where

$$\operatorname{sgn}(x) = \begin{cases} +1 & x > 0 \\ -1 & x \leq 0 \end{cases}.$$

This has a weak solution which is unique in law, but pathwise uniqueness fails.

To see the existence of solutions, let X be a one-dimensional Brownian motion with $X_0 = x$, and set

$$B_t = \int_0^t \operatorname{sgn}(X_s) dX_s,$$

which is well-defined because $\operatorname{sgn}(X_s)$ is previsible and left-continuous. Then we have

$$x + \int_0^t \operatorname{sgn}(X_s) dB_s = x + \int_0^t \operatorname{sgn}(X_s)^2 dX_s = x + X_t - X_0 = X_t.$$

So it remains to show that B is a Brownian motion. We already know that B is a continuous local martingale, so by Lévy's characterization, it suffices to show its quadratic variation is t . We simply compute

$$\langle B, B \rangle_t = \int_0^t d\langle X_s, X_s \rangle = t.$$

So there is weak existence. The same argument shows that any solution is a Brownian motion, so we have uniqueness in law.

Finally, observe that if $x = 0$ and X is a solution, then $-X$ is also a solution with the same Brownian motion. Indeed,

$$-X_t = \int_0^t \operatorname{sgn}(X_s) dB_s = \int_0^t \operatorname{sgn}(-X_s) dB_s + 2 \int_0^t \mathbf{1}_{X_s=0} dB_s,$$

where the second term vanishes, since it is a continuous local martingale with quadratic variation $\int_0^t \mathbf{1}_{X_s=0} ds = 0$. So pathwise uniqueness does not hold.

In the other direction, however, it turns out pathwise uniqueness implies uniqueness in law.

Theorem (Yamada–Watanabe). Assume weak existence and pathwise uniqueness holds. Then

- (i) Uniqueness in law holds.
- (ii) For every $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ and B and any $x \in \mathbb{R}^d$, there is a unique strong solution to $E_x(a, b)$. \square

We will not prove this, since we will not actually need it.

The key, important theorem we are now heading for is the existence and uniqueness of solutions to SDEs, assuming reasonable conditions. As in the case of ODEs, we need the following Lipschitz conditions:

Definition (Lipschitz coefficients). The coefficients $b : \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}^d$, $\sigma : \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times m}$ are Lipschitz in x if there exists a constant $K > 0$ such that for all $t \geq 0$ and $x, y \in \mathbb{R}^d$, we have

$$\begin{aligned} |b(t, x) - b(t, y)| &\leq K|x - y| \\ |\sigma(t, x) - \sigma(t, y)| &\leq |x - y| \end{aligned}$$

Theorem. Assume b, σ are Lipschitz in x . Then there is pathwise uniqueness for the $E(\sigma, b)$ and for every $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ satisfying the usual conditions and every (\mathcal{F}_t) -Brownian motion B , for every $x \in \mathbb{R}^d$, there exists a unique strong solution to $E_x(\sigma, b)$.

Proof. To simplify notation, we assume $m = d = 1$.

We first prove pathwise uniqueness. Suppose X, X' are two solutions with $X_0 = X'_0$. We will show that $\mathbb{E}[(X_t - X'_t)^2] = 0$. We will actually put some bounds to control our variables. Define the stopping time

$$S = \inf\{t \geq 0 : |X_t| \geq n \text{ or } |X'_t| \geq n\}.$$

By continuity, $S \rightarrow \infty$ as $n \rightarrow \infty$. We also fix a deterministic time $T > 0$. Then whenever $t \in [0, T]$, we can bound, using the identity $(a + b)^2 \leq 2a^2 + 2b^2$,

$$\begin{aligned} \mathbb{E}((X_{t \wedge S} - X'_{t \wedge S})^2) &\leq 2\mathbb{E} \left(\left(\int_0^{t \wedge S} (\sigma(s, X_s) - \sigma(s, X'_s)) dB_s \right)^2 \right) \\ &\quad + 2\mathbb{E} \left(\left(\int_0^{t \wedge S} (b(s, X_s) - b(s, X'_s)) ds \right)^2 \right). \end{aligned}$$

We can apply the Lipschitz bound to the second term immediately, while we can simplify the first term using the (corollary of the) Itô isometry

$$\mathbb{E} \left(\left(\int_0^{t \wedge S} (\sigma(s, X_s) - \sigma(s, X'_s)) dB_s \right)^2 \right) = \mathbb{E} \left(\int_0^{t \wedge S} (\sigma(s, X_s) - \sigma(s, X'_s))^2 ds \right).$$

So using the Lipschitz bound, we have

$$\begin{aligned} \mathbb{E}((X_{t \wedge S} - X'_{t \wedge S})^2) &\leq 2K^2(1 + T)\mathbb{E} \left(\int_0^{t \wedge S} |X_s - X'_s|^2 ds \right) \\ &\leq 2K^2(1 + T) \int_0^t \mathbb{E}(|X_{s \wedge S} - X'_{s \wedge S}|^2) ds. \end{aligned}$$

We now use Grönwall's lemma:

Lemma. Let $h(t)$ be a function such that

$$h(t) \leq c \int_0^t h(s) ds$$

for some constant c . Then

$$h(t) \leq h(0)e^{ct}. \quad \square$$

Applying this to

$$h(t) = \mathbb{E}((X_{t \wedge S} - X'_{t \wedge S})^2),$$

we deduce that $h(t) \leq h(0)e^{ct} = 0$. So we know that

$$\mathbb{E}(|X_{t \wedge S} - X'_{t \wedge S}|^2) = 0$$

for every $t \in [0, T]$. Taking $n \rightarrow \infty$ and $T \rightarrow \infty$ gives pathwise uniqueness.

We next prove existence of solutions. We fix $(\Omega, \mathcal{F}, (\mathcal{F}_t)_t)$ and B , and define

$$F(X)_t = X_0 + \int_0^t \sigma(s, X_s) dB_s + \int_0^t b(s, X_s) ds.$$

Then X is a solution to $E_x(a, b)$ iff $F(X) = X$ and $X_0 = x$. To find a fixed point, we use Picard iteration. We fix $T > 0$, and define the T -norm of a continuous adapted process X as

$$\|X\|_T = \mathbb{E} \left(\sup_{t \leq T} |X_t|^2 \right)^{1/2}.$$

In particular, if X is a martingale, then this is the same as the norm on the space of L^2 -bounded martingales by Doob's inequality. Then

$$B = \{X : \Omega \times [0, T] \rightarrow \mathbb{R} : \|X\|_T < \infty\}$$

is a Banach space.

Claim. $\|F(0)\|_T < \infty$, and

$$\|F(X) - F(Y)\|_T^2 \leq (2T + 8)K^2 \int_0^T \|X - Y\|_t^2 dt.$$

We first see how this claim implies the theorem. First observe that the claim implies F indeed maps B into itself. We can then define a sequence of processes X_t^i by

$$X_t^0 = x, \quad X^{i+1} = F(X^i).$$

Then we have

$$\|X^{i+1} - X^i\|_T^2 \leq CT \int_0^T \|X^i - X^{i-1}\|_t^2 dt \leq \dots \leq \|X^1 - X^0\|_T^2 \left(\frac{CT^i}{i!} \right).$$

So we find that

$$\sum_{i=1}^{\infty} \|X^i - X^{i-1}\|_T^2 < \infty$$

for all T . So X^i converges to X almost surely and uniformly on $[0, T]$, and $F(X) = X$. We then take $T \rightarrow \infty$ and we are done.

To prove the claim, we write

$$\|F(0)\|_T \leq |X_0| + \left\| \int_0^t b(s, 0) ds \right\| + \left\| \int_0^t \sigma(s, 0) dB_s \right\|_T.$$

The first two terms are constant, and we can bound the last by Doob's inequality and the Itô isometry:

$$\left\| \int_0^t \sigma(s, 0) dB_s \right\|_T \leq 2\mathbb{E} \left(\left| \int_0^T \sigma(s, 0) dB_s \right|^2 \right) = 2 \int_0^T \sigma(s, 0)^2 ds.$$

To prove the second part, we use

$$\begin{aligned} \|F(X) - F(Y)\|^2 &\leq 2\mathbb{E} \left(\sup_{t \leq T} \left| \int_0^t b(s, X - s) - b(s, Y_s) ds \right|^2 \right) \\ &\quad + 2\mathbb{E} \left(\sup_{t \leq T} \left| \int_0^t (\sigma(s, X_s) - \sigma(s, Y_s)) dB_s \right|^2 \right). \end{aligned}$$

We can bound the first term with Cauchy-Schwartz by

$$T\mathbb{E} \left(\int_0^T |b(s, X_s) - b(s, Y_s)|^2 ds \right) \leq TK^2 \int_0^T \|X - Y\|_t^2 dt,$$

and the second term with Doob's inequality by

$$\mathbb{E} \left(\int_0^T |\sigma(s, X_s) - \sigma(s, Y_s)|^2 ds \right) \leq 4K^2 \int_0^T \|X - Y\|_t^2 dt. \quad \square$$

4.2 Examples of stochastic differential equations

Example (The *Ornstein-Uhlenbeck process*). Let $\lambda > 0$. Then the Ornstein-Uhlenbeck process is the solution to

$$dX_t = -\lambda X_t dt + dB_t.$$

The solution exists by the previous theorem, but we can also explicitly find one.

By Itô's formula applied to $e^{\lambda t} X_t$, we get

$$d(e^{\lambda t} X_t) = e^{\lambda t} dX_t + \lambda e^{\lambda t} X_t dt = dB_t.$$

So we find that

$$X_t = e^{-\lambda t} X_0 + \int_0^t e^{-\lambda(t-s)} dB_s.$$

Observe that the integrand is deterministic. So we can in fact interpret this as an Wiener integral.

Fact. If $X_0 = x \in \mathbb{R}$ is fixed, then (X_t) is a Gaussian process, i.e. $(X_{t_i})_{i=1}^n$ is jointly Gaussian for all $t_1 < \dots < t_n$. Any Gaussian process is determined by the mean and covariance, and in this case, we have

$$\mathbb{E}X_t = e^{-\lambda t} x, \quad \text{cov}(X_t, X_s) = \frac{1}{2\lambda} \left(e^{-\lambda|t-s|} - e^{-\lambda(t+s)} \right)$$

Proof. We only have to compute the covariance. By the Itô isometry, we have

$$\begin{aligned} \mathbb{E}((X_t - \mathbb{E}X_t)(X_s - \mathbb{E}X_s)) &= \mathbb{E} \left(\int_0^t e^{-\lambda(t-u)} dB_u \int_0^s e^{-\lambda(s-u)} dB_u \right) \\ &= e^{-\lambda(t+s)} \int_0^{t \wedge s} e^{\lambda u} du. \quad \square \end{aligned}$$

In particular,

$$X_t \sim N\left(e^{-\lambda t}x, \frac{1 - e^{-2\lambda t}}{2\lambda}\right) \rightarrow N\left(0, \frac{1}{2\lambda}\right).$$

Fact. If $X_0 \sim N(0, \frac{1}{2\lambda})$, then (X_t) is a centered Gaussian process with stationary covariance, i.e. the covariance depends only on time differences:

$$\text{cov}(X_t, X_s) = \frac{1}{2\lambda} e^{-\lambda|t-s|}.$$

The difference is that in the deterministic case, the $\mathbb{E}X_t$ cancels the first $e^{-\lambda t}X_0$ term, while in the non-deterministic case, it doesn't.

This is a very nice example where we can explicitly understand the long-time behaviour of the SDE. In general, this is non-trivial.

Dyson Brownian motion

Let \mathcal{H}_N be an inner product space of real symmetric $N \times N$ matrices with inner product $N \text{Tr}(HK)$ for $H, K \in \mathcal{H}_N$. Let $H^1, \dots, H^{\dim(\mathcal{H}_N)}$ be an orthonormal basis for \mathcal{H}_N .

Definition (Gaussian orthogonal ensemble). The *Gaussian Orthogonal Ensemble* GOE_N is the standard Gaussian measure on \mathcal{H}_N , i.e. $H \sim \text{GOE}_N$ if

$$H = \sum_{r=1}^{\dim \mathcal{H}_n} H^r X^r$$

where each X^i are iid standard normals.

We now replace each X^i by a Ornstein–Uhlenbeck process with $\lambda = \frac{1}{2}$. Then GOE_N is invariant under the process.

Theorem. The eigenvalues $\lambda_1(t) \leq \dots \leq \lambda_N(t)$ satisfies

$$d\lambda_t^i = \left(-\frac{\lambda^i}{2} + \frac{1}{N} \sum_{j \neq i} \frac{1}{\lambda_i - \lambda_j} \right) dt + \sqrt{\frac{2}{N\beta}} dB^i.$$

Here $\beta = 1$, but if we replace symmetric matrices by Hermitian ones, we get $\beta = 2$; if we replace symmetric matrices by symplectic ones, we get $\beta = 4$.

This follows from Itô's formula and formulas for derivatives of eigenvalues.

Example (Geometric Brownian motion). Fix $\sigma > 0$ and $t \in \mathbb{R}$. Then geometric Brownian motion is given by

$$dX_t = \sigma X_t dB_t + r X_t dt.$$

We apply Itô's formula to $\log X_t$ to find that

$$X_t = X_0 \exp\left(\sigma B_t + \left(r - \frac{\sigma^2}{2}\right)t\right).$$

Example (Bessel process). Let $B = (B^1, \dots, B^d)$ be a d -dimensional Brownian motion. Then

$$X_t = |B_t|$$

satisfies the stochastic differential equation

$$dX_t = \frac{d-1}{2X_t} dt + dB_t$$

if $t < \inf\{t \geq 0 : X_t = 0\}$.

4.3 Representations of solutions to PDEs

Recall that in Advanced Probability, we learnt that we can represent the solution to Laplace's equation via Brownian motion, namely if D is a suitably nice domain and $g : \partial D \rightarrow \mathbb{R}$ is a function, then the solution to the Laplace's equation on D with boundary conditions g is given by

$$u(\mathbf{x}) = \mathbb{E}_{\mathbf{x}}[g(B_T)],$$

where T is the first hitting time of the boundary ∂D .

A similar statement we can make is that if we want to solve the heat equation

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

with initial conditions $u(x, 0) = u_0(x)$, then we can write the solution as

$$u(\mathbf{x}, t) = \mathbb{E}_{\mathbf{x}}[u_0(\sqrt{2}B_t)]$$

This is just a fancy way to say that the Green's function for the heat equation is a Gaussian, but is a good way to think about it nevertheless.

In general, we would like to associate PDEs to certain stochastic processes. Recall that a stochastic PDE is generally of the form

$$dX_t = b(X_t) dt + \sigma(X_t) dB_t$$

for some $b : \mathbb{R}^d \rightarrow \mathbb{R}$ and $\sigma : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times m}$ which are measurable and locally bounded. Here we assume these functions do not have time dependence. We can then associate to this a differential operator L defined by

$$L = \frac{1}{2} \sum_{i,j} a_{ij} \partial_i \partial_j + \sum_i b_i \partial_i.$$

where $a = \sigma \sigma^T$.

Example. If $b = 0$ and $\sigma = \sqrt{2}I$, then $L = \Delta$ is the standard Laplacian.

The basic computation is the following result, which is a standard application of the Itô formula:

Proposition. Let $x \in \mathbb{R}^d$, and X a solution to $E_x(\sigma, b)$. Then for every $f : \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}$ that is C^1 in \mathbb{R}_+ and C^2 in \mathbb{R}^d , the process

$$M_t^f = f(t, X_t) - f(0, X_0) - \int_0^t \left(\frac{\partial}{\partial s} + L \right) f(s, X_s) ds$$

is a continuous local martingale. □

We first apply this to the *Dirichlet–Poisson problem*, which is essentially to solve $-Lu = f$. To be precise, let $U \subseteq \mathbb{R}^d$ be non-empty, bounded and open; $f \in C_b(U)$ and $g \in C_b(\partial U)$. We then want to find a $u \in C^2(\bar{U}) = C^2(U) \cap C(\bar{U})$ such that

$$\begin{aligned} -Lu(x) &= f(x) & \text{for } x \in U \\ u(x) &= g(x) & \text{for } x \in \partial U. \end{aligned}$$

If $f = 0$, this is called the *Dirichlet problem*; if $g = 0$, this is called the *Poisson problem*.

We will have to impose the following technical condition on a :

Definition (Uniformly elliptic). We say $a : \bar{U} \rightarrow \mathbb{R}^{d \times d}$ is *uniformly elliptic* if there is a constant $c > 0$ such that for all $\xi \in \mathbb{R}^d$ and $x \in \bar{U}$, we have

$$\xi^T a(x) \xi \geq c|\xi|^2.$$

If a is symmetric (which it is in our case), this is the same as asking for the smallest eigenvalue of a to be bounded away from 0.

It would be very nice if we can write down a solution to the Dirichlet–Poisson problem using a solution to $E_x(\sigma, b)$, and then simply check that it works. We can indeed do that, but it takes a bit more time than we have. Instead, we shall prove a slightly weaker result that if we happen to have a solution, it must be given by our formula involving the SDE. So we first note the following theorem without proof:

Theorem. Assume U has a smooth boundary (or satisfies the exterior cone condition), a, b are Hölder continuous and a is uniformly elliptic. Then for every Hölder continuous $f : \bar{U} \rightarrow \mathbb{R}$ and any continuous $g : \partial U \rightarrow \mathbb{R}$, the Dirichlet–Poisson process has a solution. \square

The main theorem is the following:

Theorem. Let σ and b be bounded measurable and $\sigma\sigma^T$ uniformly elliptic, $U \subseteq \mathbb{R}^d$ as above. Let u be a solution to the Dirichlet–Poisson problem and X a solution to $E_x(\sigma, b)$ for some $x \in \mathbb{R}^d$. Define the stopping time

$$T_U = \inf\{t \geq 0 : X_t \notin U\}.$$

Then $\mathbb{E}T_U < \infty$ and

$$u(x) = \mathbb{E}_x \left(g(X_{T_U}) + \int_0^{T_U} f(X_s) ds \right).$$

In particular, the solution to the PDE is unique.

Proof. Our previous proposition applies to functions defined on all of \mathbb{R}^n , while u is just defined on U . So we set

$$U_n = \left\{ x \in U : \text{dist}(x, \partial U) > \frac{1}{n} \right\}, \quad T_n = \inf\{t \geq 0 : X_t \notin U_n\},$$

and pick $u_n \in C_b^2(\mathbb{R}^d)$ such that $u|_{U_n} = u_n|_{U_n}$. Recalling our previous notation, let

$$M_t^n = (M^{u_n})_t^{T_n} = u_n(X_{t \wedge T_n}) - u_n(X_0) - \int_0^{t \wedge T_n} Lu_n(X_s) ds.$$

Then this is a continuous local martingale that is bounded by the proposition, and is bounded, hence a true martingale. Thus for $x \in U$ and n large enough, the martingale property implies

$$\begin{aligned} u(x) = u_n(x) &= \mathbb{E} \left(u(X_{t \wedge T_n}) - \int_0^{t \wedge T_n} Lu(X_s) \, ds \right) \\ &= \mathbb{E} \left(u(X_{t \wedge T_n}) + \int_0^{t \wedge T_n} f(X_s) \, ds \right). \end{aligned}$$

We would be done if we can take $n \rightarrow \infty$. To do so, we first show that $\mathbb{E}[T_U] < \infty$.

Note that this does not depend on f and g . So we can take $f = 1$ and $g = 0$, and let v be a solution. Then we have

$$\mathbb{E}(t \wedge T_n) = \mathbb{E} \left(- \int_0^{t \wedge T_n} Lv(X_s) \, ds \right) = v(x) - \mathbb{E}(v(X_{t \wedge T_n})).$$

Since v is bounded, by dominated/monotone convergence, we can take the limit to get

$$\mathbb{E}(T_U) < \infty.$$

Thus, we know that $t \wedge T_n \rightarrow T_U$ as $t \rightarrow \infty$ and $n \rightarrow \infty$. Since

$$\mathbb{E} \left(\int_0^{T_U} |f(X_s)| \, ds \right) \leq \|f\|_\infty \mathbb{E}[T_U] < \infty,$$

the dominated convergence theorem tells us

$$\mathbb{E} \left(\int_0^{t \wedge T_n} f(X_s) \, ds \right) \rightarrow \mathbb{E} \left(\int_0^{T_U} f(X_s) \, ds \right).$$

Since u is continuous on \bar{U} , we also have

$$\mathbb{E}(u(X_{t \wedge T_n})) \rightarrow \mathbb{E}(u(T_U)) = \mathbb{E}(g(T_U)). \quad \square$$

We can use SDE's to solve the *Cauchy problem* for parabolic equations as well, just like the heat equation. The problem is as follows: for $f \in C_b^2(\mathbb{R}^d)$, we want to find $u : \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}$ that is C^1 in \mathbb{R}_+ and C^2 in \mathbb{R}^d such that

$$\begin{aligned} \frac{\partial u}{\partial t} &= Lu && \text{on } \mathbb{R}_+ \times \mathbb{R}^d \\ u(0, \cdot) &= f && \text{on } \mathbb{R}^d \end{aligned}$$

Again we will need the following theorem:

Theorem. For every $f \in C_b^2(\mathbb{R}^d)$, there exists a solution to the Cauchy problem. \square

Theorem. Let u be a solution to the Cauchy problem. Let X be a solution to $E_x(\sigma, b)$ for $x \in \mathbb{R}^d$ and $0 \leq s \leq t$. Then

$$\mathbb{E}_x(f(X_t) \mid \mathcal{F}_s) = u(t - s, X_s).$$

In particular,

$$u(t, x) = \mathbb{E}_x(f(X_t)).$$

In particular, this implies X_t is a continuous Markov process.

Proof. The martingale has $\frac{\partial}{\partial t} + L$, but the heat equation has $\frac{\partial}{\partial t} - L$. So we set $g(s, x) = u(t - s, x)$. Then

$$\left(\frac{\partial}{\partial s} + L\right)g(s, x) = -\frac{\partial}{\partial t}u(t - s, x) + Lu(t - s, x) = 0.$$

So $g(s, X_s) - g(0, X_0)$ is a martingale (boundedness is an exercise), and hence

$$u(t - s, X_s) = g(s, X_s) = \mathbb{E}(g(t, X_t) \mid \mathcal{F}_s) = \mathbb{E}(u(0, X_t) \mid \mathcal{F}_s) = \mathbb{E}(f(X_t) \mid X_s). \quad \square$$

There is a generalization to the *Feynman–Kac formula*.

Theorem (Feynman–Kac formula). Let $f \in C_b^2(\mathbb{R}^d)$ and $V \in C_b(\mathbb{R}^d)$ and suppose that $u : \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}$ satisfies

$$\begin{aligned} \frac{\partial u}{\partial t} &= Lu + Vu && \text{on } \mathbb{R}_+ \times \mathbb{R}^d \\ u(0, \cdot) &= f && \text{on } \mathbb{R}^d, \end{aligned}$$

where $Vu = V(x)u(x)$ is given by multiplication.

Then for all $t > 0$ and $x \in \mathbb{R}^d$, and X a solution to $\mathbb{E}_x(\sigma, b)$. Then

$$u(t, x) = \mathbb{E}_x \left(f(X_t) \exp \left(\int_0^t V(X_s) \, ds \right) \right). \quad \square$$

If L is the Laplacian, then this is Schrödinger equation, which is why Feynman was thinking about this.

Index

- $(H \cdot A)_t$, 9
- $H \cdot X$, 30
- $L^2(M)$, 26
- $L^2_{bc}(M)$, 29
- X^T , 11
- \mathcal{E} , 10
- \mathcal{M}^2 , 16

- Bessel process, 47
- bounded variation, 7
- bracket, 22
- BV, 7

- càdlàg, 7
- càdlàg adapted process, 9
- Cauchy problem, 49
- covariation, 22

- Dirichlet problem, 48
- Dirichlet–Poisson problem, 48
- Doléans–Dade exponential, 38
- Doob’s inequality, 16
- Dubins–Schwarz theorem, 36
- Dyson Brownian motion, 46

- Feynman–Kac formula, 50
- finite variation, 8
- Finite variation process, 9

- Gaussian orthogonal ensemble, 46
- Gaussian space, 4
- Gaussian white noise, 4
- Geometric Brownian motion, 46
- Girsanov’s theorem, 39

- Hahn decomposition, 6

- integration by parts, 32
- Itô correction, 32
- Itô isometry, 27
- Itô’s formula, 32

- Kunita–Watanabe, 23

- Lévy’s characterization of Brownian motion, 35

- Lebesgue–Stieltjes integral, 8
- Lipschitz coefficients, 43
- local martingale, 12
- locally bounded previsible process, 30

- Ornstein–Uhlenbeck process, 45

- pathwise uniqueness, 41
- Poisson problem, 48
- previsible σ -algebra, 10
- previsible process, 10

- quadratic variation, 18
 - semi-martingale, 24

- reduces, 12

- semi-martingale, 24
- signed measure, 6
- simple process, 10, 25
- stochastic differential equation, 41
- stochastic dominated convergence theorem, 31
- stochastic exponential, 38
- stochastic integral, 27, 30
- stopped process, 11
- strong solution, 41

- Tanaka equation, 42
- total variation, 6, 7
- total variation process, 9

- u.c.p., 17
- uniformly elliptic, 48
- uniformly on compact sets in probability, 17
- uniqueness in law, 41
- usual conditions, 11

- Vitali theorem, 13

- weak solution, 41
- Yamada–Watanabe, 42