

Chapter 10

★ Hilbert spaces

This chapter contains some non-examinable, though useful, foundational material on Hilbert spaces. Hilbert spaces play an important rôle in the rigorous construction of the Ito integral in the next chapter. Some basic knowledge of them is useful for quantitative finance (and indeed for all kinds of applied mathematics) but we cannot cover their theory in any detail due to lack of time. A basic example of a Hilbert space is the set of random variables on some given probability space $(\Omega, \mathcal{F}, \mathbb{P})$ which have finite variance. In analysis, this space is known as the space of square integrable functions, $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Hilbert space terminology permeates all of modern mathematics, from the theory of differential equations (where it originated) to probability theory, statistics and mathematical finance.

10.1 Pre-Hilbert and Hilbert spaces

A Hilbert space is basically an infinite-dimensional generalization of the familiar Euclidean spaces \mathbb{R}^n (we will only be concerned with what are called *real* Hilbert spaces). We give a slightly informal and descriptive – rather than formal – definition, for which we refer to the mathematical literature [cite something!].

Definition 10.1. A *pre-Hilbert space* H is a real vector space, provided with an *inner product*, which is a map from $H \times H$ to \mathbb{R} , sending pairs of elements x, y to a number $(x, y)_H \in \mathbb{R}$, such that, for $x, y, z \in H$ and $\lambda \in \mathbb{R}$,

$$(x + y, z)_H = (x, z)_H + (y, z)_H, (\lambda \cdot x, y)_H = \lambda(x, y)_H, (x, y)_H = (y, x)_H,$$

and

$$(x, x) \geq 0, \quad \text{with equality iff } x = 0.$$

To say that H is a vector space simply means that it is closed under linear combinations: if $x, y \in H$ and $\lambda, \mu \in \mathbb{R}$, then $\lambda x + \mu y \in H$. We usually simply write the inner product as (x, y) instead of $(x, y)_H$, when no confusion is possible.

Examples 10.2. Examples of pre-Hilbert spaces are as follows.

- (i) \mathbb{R}^n , with inner product $(x, y) = \sum_{j=1}^n x_j y_j$.

(ii) The space of square integrable random functions:

$$L^2(\Omega, \mathcal{F}, \mathbb{P}) = \left\{ X : \Omega \rightarrow \mathbb{R} : X \text{ } \mathcal{F}\text{-measurable, } \int_{\Omega} X(\omega)^2 d\mathbb{P}(\omega) < \infty \right\}, \quad (10.1)$$

which is the same as the vector space of random variables X such that $\mathbb{E}(X^2) < \infty$. The inner product is

$$(X, Y)_{L^2} = \int_{\Omega} X(\omega)Y(\omega) d\mathbb{P}(\omega), \quad (10.2)$$

which can be written in a more simple, and also more probabilistic, way as

$$(X, Y)_{L^2} = \mathbb{E}(XY). \quad (10.3)$$

This space will be extremely important from now on and, for us, will be *the* example of a Hilbert space.

Given a Hilbert space, we define the length or *norm* of an element x by

$$\|x\|_H := \sqrt{(x, x)_H}, \quad (10.4)$$

again often leaving off the subscript H . Clearly, $\|\lambda \cdot x\| = |\lambda| \|x\|$, if $\lambda \in \mathbb{R}$: multiplying x by the scalar λ means multiplying its length by the absolute value $|\lambda|$. If $H = \mathbb{R}^n$, then simply,

$$\|x\| = \sqrt{x_1^2 + \cdots + x_n^2},$$

the Euclidean length of x . In the case of $L^2(\Omega, \mathcal{F}, \mathbb{P})$, the norm is

$$\|X\|_{L^2} := \left(\int_{\Omega} X^2 d\mathbb{P} \right)^{1/2}. \quad (10.5)$$

This is often called the L^2 -norm.

Another important concept is that of orthogonality: we say that $x, y \in$ are *orthogonal* (notation $x \perp y$) if $(x, y) = 0$:

$$x \perp y \Leftrightarrow (x, y) = 0. \quad (10.6)$$

An important general fact about inner products is the so-called *Cauchy–Schwarz inequality*:¹

$$|(x, y)| \leq \|x\| \cdot \|y\|. \quad (10.7)$$

Another important general fact is the *triangle inequality*:

$$\|x + y\| \leq \|x\| + \|y\|.$$

To go from pre-Hilbert spaces to Hilbert spaces we have to explain the concept of completeness, which has to do with convergent sequences in H . We

¹Russian mathematicians often add the name of *Buniatowski* who indeed seems to have been the first to discover it, presumably (as for the other two) in the context of \mathbb{R}^n .

say that a sequence x_0, x_1, x_2, \dots of elements of H converges to an element $x \in H$ if the norm of the difference tends to 0, *i.e.*,

$$\lim_{n \rightarrow \infty} \|x - x_n\|,$$

which we write as

$$x_n \rightarrow x \quad \text{in } H.$$

If $(x_n)_n$ is such a convergent sequence, then the distances $\|x_m - x_n\|$ will tend to 0 as both m and n tend to ∞ simultaneously:

$$\|x_m - x_n\| \rightarrow 0 \quad \text{as } m, n \rightarrow \infty, \quad (10.8)$$

and we summarize this by saying that convergent sequences are *Cauchy sequences*. However, in general the *converse* of this statement need not be true. Hilbert spaces are precisely those pre-Hilbert spaces for which the converse *does* hold.

Definition 10.3. A Hilbert space is a pre-Hilbert space for which every Cauchy sequence has a limit.

This is called the *completeness property* of Hilbert spaces: it allows us to define elements of H by constructing Cauchy sequences. A case in point will be the Itô integral. Two examples might help to clarify this.

Examples 10.4.

- (i) Not every Cauchy sequence in \mathbb{Q} converges in \mathbb{Q} : take, for example, a sequence of rational numbers converging to $\sqrt{2}$. This will be a Cauchy sequence, but its limit, $\sqrt{2}$, is irrational.
- (ii) Slightly more ambitiously, consider the space of continuous functions

$$C[0, 1] = \{f : [0, 1] \rightarrow \mathbb{R}, f \text{ continuous}\},$$

with the L^2 -inner product:

$$(f, g) = \int_0^1 f(x)g(x) dx.$$

This is a pre-Hilbert space, but not a Hilbert space: consider the sequence of continuous functions f_n , $n \geq 1$, defined by

- (1) $f_n = 0$ on $[0, \frac{1}{2} - \frac{1}{n}]$,
- (2) f_n is linear on $[\frac{1}{2} - \frac{1}{n}, \frac{1}{2} + \frac{1}{n}]$, taking 0 at the left end-point and 1 at the right end-point,
- (3) $f_n = 1$ on $[\frac{1}{2} + \frac{1}{n}, 1]$.

Then $(f_n)_n$ converges in the L^2 -norm to the function which is 0 on $[0, \frac{1}{2})$, and 1 on $[\frac{1}{2}, 1]$ (in fact it does not matter what value we give it at point $\frac{1}{2}$). It is therefore a Cauchy sequence of elements in V , but the limit falls ‘outside V ’.

A pre-Hilbert space can have ‘holes’. There exists a general mathematical construction called *completion*, which amounts to ‘filling in all the holes’ corresponding to non-convergent Cauchy sequences. Applied to the vector space V of the previous example, this would lead us to the space $L^2([0, 1], \mathcal{B}([0, 1]), dx)$. More generally, we have the following important theorem.

Theorem 10.5. *The space $L^2(\Omega, \mathcal{F}, \mathbb{P})$ is complete, and therefore a Hilbert space.*

We will omit the proof, which can be found in any text on measure theory. In practice, it is more important to be familiar with its statement, and to know how and when to apply it.

10.2 The projection theorem

We now consider the following situation. Let H be a Hilbert space, and let $V \subset H$ be a subspace of H , that is, a subset V of H such that sums of elements are again in H , as are all products of elements of V by real numbers. Such a subspace is called *closed* if limits of convergent sequences of elements of V are again in V , that is, if $x_n \in V$, $x_n \rightarrow x$ in H implies that $x \in V$ (observe that, *a priori*, we only know that x is in the larger space, H). An example of a subspace which is *not* closed is the space V in Example 10.4(ii), regarded as a subspace of L^2 .

Theorem 10.6 (projection theorem). *Let V be a closed subspace of the Hilbert space H . Then there exists, for each $h \in H$, a unique element $v \in V$ having smallest distance to h :*

$$\|h - v\| = \min_{w \in V} \|h - w\|. \quad (10.9)$$

The element $v \in V$ is called the projection of x onto V , and is characterized by the property that $h - v$ is perpendicular to V , or

$$(h - v, w) = 0, \quad \text{for all } w \in V. \quad (10.10)$$

We again skip the proof, as we will be more concerned with applying this theorem. For the finite-dimensional Euclidean spaces \mathbb{R}^n its statement should be relatively intuitive: make a drawing in \mathbb{R}^3 with V a plane, or a line). In infinite dimensions, some care is needed: Example 10.4(ii) will again show that Theorem 10.6 is false if V is not closed: see the exercises.

Given a closed subspace $V \subset H$, we define the *orthogonal projection* $P_V : H \rightarrow V$ by

$$P_V h = v \text{ if (10.9) (or equivalently, (10.10)) holds.} \quad (10.11)$$

Note that if we let $V^\perp := \{w \in H : (w, v) = 0 \text{ for all } v \in V\}$, then P_V maps V^\perp onto 0. Further basic properties of $P = P_V$ are:

- (1) P_V is a projection, meaning that $P_V^2 = P_V$,
- (2) P_V is orthogonal, meaning that $(h - P_V h, v) = 0$, for all $v \in V$.

We note in passing that an equivalent way of stating (2) can be shown to be: for all $h, g \in H$: $(P_V h, g) = (h, P_V g)$.

10.3 Application: Conditional expectations of finite variance random variables

The following is an important application of this construction. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a σ -algebra, and let $\mathcal{G} \subset \mathcal{F}$ be some smaller σ -algebra, *e.g.*, $\mathcal{G} = \sigma(X)$, the σ -algebra of information which can be gleaned from observing the random variable X . We can show that in this case the space

$$V = L^2(\Omega, \mathcal{G}, \mathbb{P})$$

of square-integrable \mathcal{G} -measurable random variables is a closed subspace of $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Note that for a random variable X to be in V , $X^{-1}((a, b))$ has to be in \mathcal{G} instead of \mathcal{F} , which is a more stringent condition. The *conditional expectation with respect to \mathcal{G}* can now be defined as being the orthogonal projection

$$\mathbb{E}(\cdot | \mathcal{G}) : L^2(\Omega, \mathcal{F}, \mathbb{P}) \rightarrow L^2(\Omega, \mathcal{G}, \mathbb{P}). \quad (10.12)$$

That is, if X is a random variable with finite variance, then its *conditional expectation* $\Phi := \mathbb{E}(X | \mathcal{G})$ is characterized by the two following conditions:

- Φ is \mathcal{G} -measurable, and square-integrable: $\mathbb{E}(\Phi^2) < \infty$,
- for all square integrable and \mathcal{G} -measurable random variables $Y \in L^2(\Omega, \mathcal{G}, \mathbb{P})$,

$$\mathbb{E}(\Phi Y) = \mathbb{E}(XY), \quad (10.13)$$

since this is the same (in a different notation) as

$$(X - \Phi, Y) = 0, \text{ for all } Y \in V = L^2(\Omega, \mathcal{G}, \mathbb{P}),$$

which characterized the orthogonal projection, according to the projection theorem.

One way to think about $\mathbb{E}(X | \mathcal{G})$ is that it represents the best prediction of the random variable X , given that we only dispose of the information \mathcal{G} . Indeed, remembering the equivalence of (10.9) and (10.10) in Theorem 10.6, we may restate (10.13) as a variance-minimizing property (after subtraction of the mean from X):

$$\text{var}(X - \mathbb{E}(X | \mathcal{G})) = \min_{\substack{Y \mathcal{G}\text{-meas.} \\ \mathbb{E}(Y^2) < \infty}} \text{var}(X - Y). \quad (10.14)$$

We will return to conditional expectations in more detail (and from different points of view) later on.

10.4 Exercises to Chapter 10

Exercise 10.7. Let $H = L^2(\Omega, \mathcal{F}, \mathbb{P})$. Show, using the Cauchy–Schwarz–Buniatowski inequality (10.7), that if $X \in H$, then $\mathbb{E}(|X|) < \infty$. Hence the mean $\mathbb{E}(X)$ of X exists (is finite). Conclude that H is the same as the space of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ having finite mean and variance.

Exercise 10.8. Show by an example that Theorem 10.6 is false if V is not closed.

Hint. See Example 10.4(ii).

Exercise 10.9. Assuming the existence of a distance minimizing v , prove the equivalence of (10.9) and (10.10).

Exercise 10.10. Show that if $X \geq 0$ with probability one, then the same holds for its conditional expectation $\mathbb{E}(X \mid \mathcal{G})$.

Hint. A \mathcal{G} -measurable random variable Φ is ≥ 0 with probability one iff $\mathbb{E}(\Phi I_G) \geq 0$ for all $G \in \mathcal{G}$.

Chapter 11

The Itô stochastic integral

We will now return to the issue of defining a stochastic integral,

$$\int_0^T f_t dW_t, \quad (11.1)$$

where W_t is a Brownian motion. We now want to allow f_t to be stochastic as well, and not just a deterministic function of time t , as in the discussion in Chapter 3, and part of the problem is which stochastic f_t we can allow.

11.1 Brownian motion revisited

With our new view of probability, Brownian motion will now consist of a family of random variables

$$W_t : \Omega \rightarrow \mathbb{R},$$

defined on some fixed probability space $(\Omega, \mathcal{F}, \mathbb{P})$ (for example, the one we introduced in Section 6.5, although other probability spaces are possible), and satisfying the usual axioms for Brownian motion:

- $W_0 = 0$,
- $W_t - W_s \sim N(0, t - s)$ for $s \leq t$,
- $W_u, W_t - W_s$ are independent if $u \leq s < t$.

The main advantage of this new viewpoint is that we can now talk about the *sample paths* of Brownian motion, which are the functions

$$t \rightarrow W_t(\omega) : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}, \quad \omega \in \Omega \text{ fixed.} \quad (11.2)$$

Note that there is a sample path for each $\omega \in \Omega$.

We can now also be more precise on the continuity of Brownian motion. In fact, we can (and will) suppose that $(\Omega, \mathcal{F}, \mathbb{P})$ and W_t are such that *with probability one, the sample paths of W_t are continuous functions of t* . More precisely, there exists a null-set or null-event $N \subset \Omega$, $\mathbb{P}(N) = 0$, such that:

$$\text{For all } \omega \in \Omega \setminus N, t \rightarrow W_t(\omega) \text{ is continuous everywhere.} \quad (11.3)$$

It turns out that it is always possible to achieve this, by re-defining W_t suitably: see Remark 8.18.

Theorem 11.1. *Brownian motion is almost surely nowhere differentiable. More precisely, there exists a null-event $N \in \mathcal{F}$ such that, for all $\omega \notin N$, the function $t \rightarrow W_t(\omega)$ is nowhere differentiable (i.e., its graph does not have a tangent at any point).*

The proof consists again of some pretty technical mathematics; see, for example, Billingsley [5].

11.2 Filtrations of σ -algebras

Before starting the construction of the Itô integral, we have to introduce an important new concept, that of a (*continuous*) *filtration of σ -algebras*. This is a family of σ -algebras \mathcal{F}_t , one for each $t \geq 0$, which is *increasing*, in the sense that

$$s \leq t \Rightarrow \mathcal{F}_s \subset \mathcal{F}_t. \quad (11.4)$$

If we think of a σ -algebra as codifying information which can be obtained from making observations/doing statistical experiments, then we are simply dealing with an information set that grows with time. The following is a standard example.

Example 11.2. Let $(W_t)_{t \geq 0}$ be a Brownian motion, as above. Define \mathcal{F}_0^W as being the collection of all null-sets in \mathcal{F} , and \mathcal{F}_t^W as the σ -algebra generated by \mathcal{F}_0^W together with all W_s for $s \leq t$:

$$\mathcal{F}_t^W = \sigma(\{W_s : s \leq t\} \cup \mathcal{F}_0^W). \quad (11.5)$$

Loosely speaking, \mathcal{F}_t^W contains all possible information that can be obtained from observing Brownian motion up to, and including, time t (the null-sets counting as containing no information at all). The filtration $(\mathcal{F}_t^W)_{t \geq 0}$ is called the *Brownian filtration* (on the given probability space $(\Omega, \mathcal{F}, \mathbb{P})$, to be precise).

Informally, \mathcal{F}_t^W -measurable functions can be thought of as functions of some, or all, of the W_{u_s} with $u \leq t$, i.e., random variables Y of the form

$$Y = g(W_{u_1}, \dots, W_{u_k})$$

(where g is a continuous function, for example, and $u_j \leq t$) and every \mathcal{F}_t^W -measurable Y can be shown to be a (pointwise) limit of a sequence of such special Y s. Now recall that, if $u \leq s < t$, then W_u and $W_t - W_s$ are independent. Functions of such W_u will also be independent of $W_t - W_s$, and, in view of the above, we obtain the following important property:

$$\begin{aligned} \text{If } u \leq s < t, \text{ then every } \mathcal{F}_u^W\text{-measurable} \\ \text{function } Y \text{ is independent of } W_t - W_s. \end{aligned} \quad (11.6)$$

11.3 Defining the Itô integral

We now place ourselves in the following situation: besides our Brownian motion $(W_t)_{t \geq 0}$, we dispose of some filtration $(\mathcal{F}_t)_{t \geq 0}$ such that:

$$\text{Each } W_t \text{ is } \mathcal{F}_t\text{-measurable,} \quad (11.7)$$

and such that the analogue of (11.6) holds:

$$\begin{aligned} \text{If } u \leq s < t, \text{ then every } \mathcal{F}_u\text{-measurable} \\ \text{function } Y \text{ is independent of } W_t - W_s. \end{aligned} \quad (11.8)$$

We can always take for \mathcal{F}_t the Brownian filtration \mathcal{F}_t^W , but larger filtrations are also allowed, provided (11.8) is satisfied. This liberty of choice is often convenient. For example, we might have *two* independent Brownian motions $(W_t)_t$ and $(Z_t)_t$. Then, if

$$\mathcal{F}_t = \sigma(W_s, Z_r : s, r \leq t)$$

is the natural filtration generated by the two Brownian motions together, then \mathcal{F}_t satisfies both (11.7) and (11.8), and is strictly bigger than \mathcal{F}_t^W , the filtration generated by W_t only.

The second ingredient for a stochastic integral is a stochastic process $H = (H_t)_{t \geq 0}$, $H_t : \Omega \rightarrow \mathbb{R}$, which will serve as the integrand in our stochastic integral (11.1), and which has to satisfy the following important condition:

$$H_t \text{ is } \mathcal{F}_t\text{-measurable, for each } t \geq 0. \quad (11.9)$$

We will say in this case that the process H_t is *adapted to the filtration* \mathcal{F}_t ($t \geq 0$), and also that it is *non-anticipating*.¹

For such $H = (H_t)_t$ we will now give a sense to the integral

$$I_T(H) = I_T(H)(\omega) = \int_0^T H_t(\omega) dW_t(\omega), \quad \omega \in \Omega, \quad (11.10)$$

even though $t \rightarrow W_t(\omega)$ is, with probability one, nowhere differentiable. Note that $I_T(H)$ will be a *function* on Ω , that is, a *random variable*.

We will construct the integral in two stages, first for a suitable class of *simple* adapted processes H_t , and then for more general H_t , by a process of approximation approximation and passing to the limit.² It is for this last step that the formalism of Hilbert spaces will prove to be very convenient.

11.4 Itô's integral for simple adapted processes

To define a simple process $(H_t)_{t \geq 0}$ we first need a finite partition

$$t_0 = 0 < t_1 < \cdots < t_{N-1} < t_N = T \quad (11.11)$$

of the interval $[0, T]$: the points t_j divide the interval $(0, T]$ up in (typically small) subintervals $(t_{j-1}, t_j]$. As a concrete example you could think of

$$t_j = \frac{jT}{N}, \quad 0 \leq j \leq N,$$

¹The origin of this term lies in the fact that when $\mathcal{F}_t = \mathcal{F}_t^W$, then such an H_t only depends on past-to-present values W_s , $s \leq t$, not on future values.

²It is perhaps good to realize that *any* integral is always the result of some limit process, beginning with the basic integrals you were taught about in calculus!

where N is thought of as a big number. We next choose, for each t_j , an $\mathcal{F}_{t_{j-1}}$ -measurable random variable K_{j-1} and call $(H_t)_{t \geq 0}$ *simple* if it is of the form

$$H_t(\omega) = \sum_{j=1}^n K_{j-1}(\omega) I_{(t_{j-1}, t_j]}(t). \quad (11.12)$$

In other words,

$$H_t(\omega) = \begin{cases} K_0(\omega), & \text{if } 0 = t_0 \leq t < t_1, \\ K_1(\omega), & \text{if } t_1 \leq t < t_2, \\ \vdots & \\ K_{n-1}(\omega), & \text{if } t_{n-1} \leq t < t_n = T. \end{cases}$$

One can visualize such an H_t as a step function on $[0, T]$, whose levels are given by the random numbers $K_j = K_j(\omega)$. Observe that H_t is adapted, since, for $t_{j-1} \leq t < t_j$, say,

$$H_t = K_{j-1} \text{ is } \mathcal{F}_{t_{j-1}}\text{-measurable, and therefore } \mathcal{F}_t\text{-measurable,}$$

since $\mathcal{F}_{t_{j-1}} \subset \mathcal{F}_t$. We next define the Itô integral of such a simple function by

$$I_T(H) = \int_0^T H_t dW_t := \sum_{j=0}^{N-1} H_{t_j}(\omega) (W_{t_{j+1}}(\omega) - W_{t_j}(\omega)). \quad (11.13)$$

Observe that this is a function of ω , and therefore a random variable on Ω (although we usually suppress the variable ω when writing $\int_0^T H_t dW_t$); this is why it is called a *stochastic* integral.

The integral (11.13) has the following important properties.

Lemma 11.3. *If $(H_t)_{t \geq 0}$ is a simple adapted process, and if*

$$I_T = I_T(H) := \int_0^T H_t dW_t,$$

then:

(i) I_T has mean 0,

$$\mathbb{E}(I_T) = 0, \quad (11.14)$$

(ii) I_T has variance

$$\mathbb{E}(I_T^2) = \int_0^T \mathbb{E}(H_t)^2 dt. \quad (11.15)$$

Proof. The proof of (i) is easy:

$$\begin{aligned} \mathbb{E}(I_T(H)) &= \sum_j \mathbb{E}(H_{t_j}(W_{t_{j+1}} - W_{t_j})) \\ &= \sum_j \mathbb{E}(H_{t_j}) \mathbb{E}(W_{t_{j+1}} - W_{t_j}) \\ &= 0, \end{aligned}$$

since H_{t_j} and $W_{t_{j+1}} - W_{t_j}$ are independent, and Brownian motion has mean zero.

The proof of (ii) is a bit more involved:

$$\begin{aligned} I_T^2 &= \left(\sum_j H_{t_j} (W_{t_{j+1}} - W_{t_j}) \right)^2 \\ &= \sum_j \sum_k (H_{t_j} (W_{t_{j+1}} - W_{t_j})) (H_{t_k} (W_{t_{k+1}} - W_{t_k})) \\ &= \sum_j H_{t_j}^2 (W_{t_{j+1}} - W_{t_j})^2 \\ &\quad + \sum_{j \neq k} \sum (H_{t_j} (W_{t_{j+1}} - W_{t_j})) (H_{t_k} (W_{t_{k+1}} - W_{t_k})). \end{aligned}$$

We have to compute the expectation of this expression.

Now, by independence of the (\mathcal{F}_{t_j} -measurable) random variable H_{t_j} and $(W_{t_{j+1}} - W_{t_j})$,

$$\begin{aligned} \mathbb{E}(H_{t_j}^2 (W_{t_{j+1}} - W_{t_j})^2) &= \mathbb{E}(H_{t_j}^2) \mathbb{E}((W_{t_{j+1}} - W_{t_j})^2) \\ &= \mathbb{E}(K_j^2) \cdot (t_{j+1} - t_j). \end{aligned}$$

As for the terms with $j \neq k$, if for example $j < k$, then $j + 1 \leq k$, and

$$H_{t_j} (W_{t_{j+1}} - W_{t_j}) H_{t_k}$$

will be \mathcal{F}_{t_k} -measurable and therefore independent of $W_{t_{k+1}} - W_{t_k}$. Hence,

$$\begin{aligned} \mathbb{E}(H_{t_j} (W_{t_{j+1}} - W_{t_j}) H_{t_k} (W_{t_{k+1}} - W_{t_k})) \\ &= \mathbb{E}(H_{t_j} (W_{t_{j+1}} - W_{t_j}) H_{t_k}) \mathbb{E}(W_{t_{k+1}} - W_{t_k}) \\ &= 0. \end{aligned}$$

It follows that

$$\begin{aligned} \mathbb{E}(I_T^2) &= \sum_j \mathbb{E}(K_j^2) (t_{j+1} - t_j)^2 \\ &= \int_0^T \mathbb{E}(H_t^2) dt, \end{aligned}$$

since the function $t \rightarrow \mathbb{E}(H_t^2)$ is an ordinary step function (with values in \mathbb{R}), which is equal to $\mathbb{E}(K_j^2)$ on $(t_{j-1}, t_j]$. \square

Remark 11.4. If we write out (11.15) in full, we get the statement

$$\int_{\Omega} \left(\int_0^T H_t(\omega) dW_t(\omega) \right)^2 d\mathbb{P}(\omega) = \int_0^T \int_{\Omega} H_t^2(\omega) d\mathbb{P}(\omega) dt.$$

This is certainly not an obvious identity: when going from left to right we have to move the square under the second integral sign, to get it on $f_t(\omega)$, and this is false in general. To be able to do this, we needed H_t to be adapted, and we also used the basic properties of Brownian motion, in particular independence of the

future and the past-up-to-present. A closer analysis shows that the martingale property of Brownian motion, which amounts to the statement that

$$\mathbb{E}(W_t | \mathcal{F}_s) = W_s,$$

is in fact all that is needed, and that one can generalize the stochastic integral by replacing $(W_t)_t$ by any square integrable martingale. We will precisely define and discuss martingales in a later chapter.

The identity (11.15) is fundamental for extending the Itô integral to more general f_t s than just the simple ones.

11.5 The Itô integral in full generality

The Itô integral will now be extended from simple step functions to more general ones by a limit argument. The basic idea is simply to approximate a general adapted integrand $H = (H_t)_t$ by a sequence of simple adapted integrands, $H_n = (H_{n,t})_t$,

$$H_{n,t}(\omega) \rightarrow H_t(\omega), \quad n \rightarrow \infty, \quad (11.16)$$

and to set

$$I_T(H) = \lim_{n \rightarrow \infty} I_T(H_n) = \lim_{n \rightarrow \infty} \int_0^T H_{n,t}(\omega) dW_t(\omega), \quad (11.17)$$

hoping that this limit exists. Now the whole mathematical subtlety of the Itô integral lies in the way in which we have to interpret these two limits (11.16) and (11.17). Note that these are not simple limits of numbers, with which you should be familiar from elementary calculus, but limits of *functions*, namely functions on the sample space, Ω . The subject of limits in function spaces generated intense mathematical research during the first half of the 20th century, and has given rise to the field of functional analysis. There exist several notions of convergence of sequences of functions. The Itô integral is best understood in the context of Hilbert spaces; indeed, (11.15) can be understood as asserting that the Itô integral on simple adapted functions is a length-preserving map, or isometry, between two suitably defined spaces of square integrable functions (L^2 -spaces). We will therefore slightly change viewpoint and re-interpret everything in terms of functions on Ω and $[0, \infty) \times \Omega$, respectively. First observe that we can look upon a process $(H_t)_{t \leq T}$ as simply being a function

$$H : [0, T] \times \Omega \rightarrow \mathbb{R},$$

sending

$$H : (t, \omega) \rightarrow H_t(\omega).$$

Thus we will also write $H_t(\omega)$ as $H(t, \omega)$. We now introduce the space of functions:

$$\mathcal{H}_T^2 = \left\{ H : [0, T] \times \Omega \rightarrow \mathbb{R} \text{ adapted, } \int_0^T \int_{\Omega} H(t, \omega)^2 d\mathbb{P}(\omega) dt < \infty \right\}, \quad (11.18)$$

where we will tacitly assume that H is measurable with respect to the product σ -algebra $\mathcal{B}([0, T]) \times \mathcal{F}$, in order to be able to give sense to the double integral; ignore this point if it confuses you.

If we introduce the inner product

$$(H, K)_{\mathcal{H}} = \int_0^T \int_{\Omega} H(t, \omega) K(t, \omega) dt d\mathbb{P}(\omega),$$

then \mathcal{H}_T^2 can be shown to be a Hilbert space.³ The corresponding ‘infinite-dimensional Euclidean length’ is given by

$$\|H\|_{\mathcal{H}}^2 = \int_0^T \int_{\Omega} H(t, \omega)^2 d\mathbb{P}(\omega) dt,$$

which we can also write as

$$\|H\|_{\mathcal{H}}^2 = \int_0^T \mathbb{E}(H_t)^2 dt.$$

\mathcal{H}_T^2 is the L^2 -space of adapted processes. Recall from Chapter 8 that there is also the (simpler) Hilbert space of random variables:

$$L^2(\Omega) = \left\{ X : \Omega \rightarrow \mathbb{R} : \mathbb{E}(X^2) = \int_{\Omega} X(\omega)^2 d\mathbb{P}(\omega) < \infty \right\},$$

with ‘Euclidean length’

$$\|X\|_{L^2}^2 = \int_{\Omega} X(\omega)^2 d\mathbb{P}(\omega) = \mathbb{E}(X^2).$$

We note in passing that using a more precise notation, $L^2(\Omega)$ should be designated as $L^2(\Omega, \mathcal{F}, \mathbb{P})$, since it depends on both the σ -algebra and on the measure. A similar remark applies to \mathcal{H}_T^2 , but we will keep to the simplified notation, so as not to make the following totally unreadable. Thus both \mathcal{H}_T^2 and $L^2(\Omega)$ are spaces of the same nature, namely functions whose square can be integrated, or are *square-integrable*, with respect to a suitable measure.

We can now show the following important fact concerning \mathcal{H}_T^2 .

Lemma 11.5. *Simple adapted processes, regarded as functions on $[0, T] \times \Omega$, are called dense in \mathcal{H}_T^2 : for each $H \in \mathcal{H}_T^2$ there exists a sequence of simple functions H_n such that*

$$\|H - H_n\|_{\mathcal{H}} \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

Equivalently, interpreting everything as processes, if $H = (H_t)_{0 \leq t \leq T}$ is a stochastic process adapted to the filtration $\mathcal{F}_t, t \leq T$, then we can always find a sequence $H_n = (H_{n,t})_{0 \leq t \leq T}$ of simple adapted processes, such that

$$\|H - H_n\|_{\mathcal{H}} = \int_0^T \mathbb{E}((H_t - H_{n,t})^2) dt \rightarrow 0. \quad (11.19)$$

The lemma is non-trivial, and requires a proof, for which we refer to the literature (see, for example, the book by Oksendal [18]). Below we will indicate how to construct such an approximating sequence if H has continuous paths.

We now show how to use these properties to define the Itô integral $I_T(H) = \int_0^T H_t dW_t$ of an arbitrary process $H = (H_t)_{t \geq 0}$ in \mathcal{H}_T^2 .

³In fact, \mathcal{H}_T^2 is a closed subspace of the L^2 -space $L^2([0, T] \times \Omega, \mathcal{B}([0, T]) \times \mathcal{F}, dt d\mathbb{P})$.

Step 1. By the lemma, there exists a sequence of simple adapted processes, $H_n = (H_{n,t})_t$, such that

$$\|H - H_n\|_{\mathcal{H}} \rightarrow 0.$$

By the triangle inequality,

$$\|H_n - H_m\|_{\mathcal{H}} \leq \|H_n - H\|_{\mathcal{H}} + \|H - H_m\|_{\mathcal{H}} \rightarrow 0,$$

as $n, m \rightarrow \infty$ simultaneously.

Step 2. Being a simple adapted process, the Itô integral of H_n is already defined. We set

$$Y_n := I_T(H_n) = \int_0^T H_{n,t} dW_t.$$

Then $I_n - I_m = I_T(H_n - H_m)$, and (11.15) implies that

$$\begin{aligned} \|Y_n - Y_m\|_{L^2(\Omega)}^2 &= \mathbb{E}((Y_n - Y_m)^2) \\ &= \int_0^T \mathbb{E}((H_n - H_m)^2) dt \\ &= \|H_n - H_m\|_{\mathcal{H}}^2 \rightarrow 0, \end{aligned} \quad (11.20)$$

as $n, m \rightarrow \infty$ simultaneously. Hence:

$$(Y_n)_n \text{ is a Cauchy sequence in } L^2(\Omega)!$$

Step 3. $L^2(\Omega)$, being a Hilbert space, is complete, meaning that every Cauchy sequence converges. Hence there exists an element $X \in L^2(\Omega)$ such that

$$\|X - Y_n\|_{L^2(\Omega)}^2 \rightarrow 0,$$

that is,

$$\mathbb{E}((X - I_T(H_n))^2) \rightarrow 0,$$

and we define the stochastic integral of our original process H simply as being this limit:

$$I_T(H) = \int_0^T H_t dW_t := X = \lim_{n \rightarrow \infty} I_T(H_n).$$

Step 4. We finally have to show that this is a good definition, in the sense that the $I_T(H) := X$ we found does not depend on the approximating sequence $(H_n)_n$ of simple adapted processes (in general, there is more than one such sequence). But this is relatively easy: if $(\widehat{H}_n)_n$ is another such sequence, then

$$\|H_n - \widehat{H}_n\|_{\mathcal{H}} \rightarrow 0,$$

and therefore, using (11.15),

$$\|I_T(H_n) - I_T(\widehat{H}_n)\| = \|H_n - \widehat{H}_n\|_{\mathcal{H}} \rightarrow 0.$$

Hence, necessarily

$$\lim_{n \rightarrow \infty} I_T(H_n) = \lim_{n \rightarrow \infty} I_T(\widehat{H}_n).$$

This completes the construction of the Itô integral.

It can be shown that I_T inherits the properties (11.14) and (11.15) established the simple Itô integrals: the approximating $I_T(g_n)$ have these properties, and they persist in the limit (we will not give a formal proof). We record this formally as follows.

Theorem 11.6. *Let $H = (H_t)_{0 \leq t \leq T} \in \mathcal{H}_T^2$, and let*

$$I_T(H) = \int_0^T H_t dW_t,$$

its Itô integral, which we just defined. Then

$$\mathbb{E}(I_T(H)) = 0, \tag{11.21}$$

and

$$\mathbb{E}(I_T(H)^2) = \int_0^T \mathbb{E}(H_t^2) dt. \tag{11.22}$$

More generally, if $K = (K_t)_t$ is also in \mathcal{H}_T^2 , then

$$\mathbb{E}(I_T(H) I_T(K)) = \int_0^T \mathbb{E}(H_t K_t) dt. \tag{11.23}$$

The only point that perhaps needs comment is (11.23). This can either be established along the same lines as (11.22), by first verifying it for simple functions, and then passing to the limit (for which you will need Cauchy's inequality if you want to prove it rigorously). Alternatively, it can be derived from (11.22) by using a small trick called *polarization*: see the exercises at the end of this chapter.

11.6 Itô's integral for integrands with continuous sample paths

The above procedure is admittedly a bit abstract. However, if the integrand H_t has continuous sample paths $t \rightarrow H_t(\omega)$, and is bounded (which accounts for the vast majority of stochastic integrals used in finance), we can give a less abstract and more concrete description of $I_T(H)$, which corresponds closely to the usual picture of integration in ordinary calculus. The point is that for such H_t we can choose a very simple type of approximating sequence of simple functions. For convenience, we will sometimes write

$$H(t, \omega) \text{ for } H_t(\omega).$$

We then set

$$H_{n,t}(\omega) = \sum_{j=0}^{n-1} H\left(\frac{jT}{n}, \omega\right) I_{(jT/n, (j+1)T/n]}(t). \tag{11.24}$$

These are clearly simple adapted functions, since H_t is adapted. If the sample paths of H_t are continuous then, for each $\omega \in \Omega$ and $0 \leq t \leq T$,

$$H_{n,t}(\omega) \rightarrow H_t(\omega), \quad n \rightarrow \infty.$$

Under the additional condition that H is bounded, meaning that there exists a constant $C > 0$ such that

$$|H_t(\omega)| \leq C, \quad \text{for all } \omega \in \Omega, \quad 0 \leq t \leq T,$$

we show⁴ that H_n tends to H in the space \mathcal{H}_T^2 : $\|H - H_n\|_{\mathcal{H}} \rightarrow 0$

Next, if we compute the Itô integral of $H_{n,t}$, then we obtain $I_T(H_{n,t}) = S_n(H, T)$, where

$$S_n(H, T)(\omega) := \sum_{j=0}^{n-1} H\left(\frac{jT}{n}, \omega\right) (W_{(j+1)T/n}(\omega) - W_{jT/n}(\omega)). \quad (11.25)$$

Observe that this is just like a Riemann sum, with the integrand H_t always evaluated at the *left* end-point of the interval $[jT/n, (j+1)T/n]$. We then have the following.

Theorem 11.7. *If H_t is adapted and bounded, then*

$$\begin{aligned} I_T(H) &= \lim_{n \rightarrow \infty} S_n(H, T) \\ &= \lim_{n \rightarrow \infty} \sum_{j=0}^{n-1} H\left(\frac{jT}{n}, \omega\right) (W_{(j+1)T/n}(\omega) - W_{jT/n}(\omega)), \end{aligned} \quad (11.26)$$

in the sense that

$$\mathbb{E}((I_T(H) - S_n(H, T))^2) \rightarrow 0, \quad \text{as } n \rightarrow \infty. \quad (11.27)$$

The convergence in (11.27) is called (for obvious reasons) *convergence in mean square sense*.

Observe that (11.26) immediately suggests how to compute numerically different realizations of $I_T(H)$, given sample paths of H_t and of W_t . As for ordinary numerical integration, this simple algorithm can be improved: see the book by Kloeden and Platen [12].

11.7 Itô processes

Let $H = (H_t)_{t \geq 0}$ be an adapted processes, which is in \mathcal{H}_t^2 for each time $t \geq 0$:

$$\int_0^t \mathbb{E}(H_s^2) ds < \infty, \quad \text{for all } t > 0$$

for each t . For such H the Itô integrals

$$\int_0^t H_s(\omega) dW_s(\omega)$$

are well-defined and, as a function of the upper limit of integration t , define a new process. More generally, we can add an integral of the type

$$\int_0^t A_s(\omega) ds,$$

⁴For instance, by applying Lebesgue's dominated convergence theorem.

which are unproblematic if, for example, the process A_t has continuous paths: this is just an ordinary integral with respect to ds . Set

$$X_t = X_0 + \int_0^t A_s ds + \int_0^t H_s dW_s, \tag{11.28}$$

where X_0 is a constant; X_t is called an *Itô process* and symbolically written as

$$dX_t = A_t dt + H_t dW_t. \tag{11.29}$$

Remark on notation. To stress the fact that our integrands H_s in integrals such as $\int_0^t H_s dW_s$ are allowed to be stochastic, we have systematically designated them by capital letters in this chapter. However, in the end it becomes a bit tiresome to use capital letters for random variables, and we will often revert to lower-case letters h_t, a_t for stochastic processes, especially when they occur as integrands of Itô processes, in accordance with general notational practice in stochastics and finance. Itô processes are thus written as

$$dX_t = a_t dt + h_t dW_t, \quad \text{etc.},$$

and it should be clear from the context whether or not a_t and h_t are stochastic.

In a rigorous mathematical development of Itô calculus, stochastic differentials such as (11.29) are just a symbolic shorthand for the corresponding stochastic integral (11.28), and are not in themselves considered *bona fide* mathematical objects. However, the basic intuition obtained from manipulating stochastic differentials according to the Itô rules we gave in Chapter 3 is correct, and the results established using these rules can be rigorously proved after translating them into integrals. As an example we take another look at Itô's lemma.

11.8 ★ Itô's lemma revisited

For simplicity we will limit ourselves to the simplest and most basic form of Itô's lemma, formula (4.10): if $f = f(w)$ is 3-times continuously differentiable, with bounded derivatives, then

$$df(W_t) = f'(W_t) dW_t + \frac{1}{2} f''(W_t) dt. \tag{11.30}$$

In integral form, this becomes

$$f(W_t) - f(0) = f(W_t) - f(W_0) = \int_0^t f'(W_s) dW_s + \int_0^t \frac{1}{2} f''(W_s) ds. \tag{11.31}$$

★ *Proof.* We sketch a proof based on the construction of the Itô integral in the previous sections. The idea is to cut things up in small intervals again, and write

$$f(W_t) - f(W_0) = \sum_{j=1}^{n-1} f(W_{(j+1)t/n}) - f(W_{jt/n}),$$

where both sides of the equation are of course functions of ω , which we suppress for legibility. We then use the Taylor expansion of f to analyse each of the

terms in the sum on the right, as follows:

$$\begin{aligned} f(W_{(j+1)t/n}) - f(W_{jt/n}) &= f'(W_{jt/n})(W_{(j+1)t/n} - W_{jt/n}) \\ &\quad + \frac{1}{2}f''(W_{jt/n})(W_{(j+1)t/n} - W_{jt/n})^2 \\ &\quad + \text{remainder } R_j. \end{aligned}$$

Since we assume that f is 3-times differentiable, with continuous and bounded derivatives, it follows from one of the standard calculus formulas for Taylor with remainder, that the remainder term can be estimated by

$$|R_j(w)| \leq C|W_{(j+1)t/n} - W_{jt/n}|^3, \quad (11.32)$$

where C is some sufficiently big constant which dominates the third derivative of f . Taking the sum over all j from 0 to $n-1$, proving (11.31) then amounts to showing that the following limits hold, in mean square sense:

$$\sum_{j=0}^{n-1} f'(W_{jt/n})(W_{(j+1)t/n} - W_{jt/n}) \rightarrow \int_0^t f'(W_s) dW_s, \quad (11.33)$$

$$\sum_{j=0}^{n-1} f''(W_{jt/n})(W_{(j+1)t/n} - W_{jt/n})^2 \rightarrow \int_0^t f''(W_s) ds, \quad (11.34)$$

and

$$\sum_{j=0}^{n-1} R_j \rightarrow 0. \quad (11.35)$$

To simplify the notation, we set

$$t_j = \frac{jt}{n}.$$

The first limit, (11.33), follows from Theorem 11.7. To get some insight into the second, observe that the expectation of the left-hand side of (11.34),

$$\begin{aligned} \mathbb{E} \left(\sum_{j=0}^{n-1} f''(W_{t_j})(W_{t_{j+1}} - W_{t_j})^2 \right) &= \sum_{j=0}^{n-1} f''(W_{t_j})(t_{j+1} - t_j) \\ &=: J_n \rightarrow \int_0^t f''(W_s) ds, \end{aligned}$$

the last line by the definition of the ordinary (Riemann) integral from calculus. This is encouraging, but not yet enough, since we have to prove that the expectation of the square of the difference of the left-hand side with the right-hand side goes to 0. This is of course the same as saying that its square root goes to 0, which is the L^2 -norm. But since $\|X\|_{L^2} = \sqrt{\mathbb{E}(X^2)}$ satisfies the triangle inequality, it suffices to show that

$$\left\| \sum_{j=0}^{n-1} f''(W_{t_j})(W_{t_{j+1}} - W_{t_j})^2 - J_n \right\|_{L^2} \rightarrow 0, \quad (11.36)$$

for then

$$\begin{aligned} & \left\| \sum_{j=0}^{n-1} f''(W_{t_j})(W_{t_{j+1}} - W_{t_j})^2 - \int_0^t f''(W_s) ds \right\|_{L^2} \\ & \leq \left\| \sum_{j=0}^{n-1} f''(W_{t_j})(W_{t_{j+1}} - W_{t_j})^2 - J_n \right\|_{L^2} \\ & \quad + \left\| J_n - \int_0^t f''(W_s) ds \right\|_{L^2}. \end{aligned}$$

The first term on the right tends to zero, by (11.36), and the second tends to zero by definition of the ordinary integral.⁵

So we are left with establishing (11.36). Expand, *i.e.*,

$$\begin{aligned} & \left(\left[\sum_{j=0}^{n-1} f''(W_{t_j})(W_{t_{j+1}} - W_{t_j})^2 \right] - J_n \right)^2 \\ & = \left(\sum_{j=0}^{n-1} f''(W_{t_j})(W_{t_{j+1}} - W_{t_j})^2 - (t_{j+1} - t_j) \right)^2 \\ & = \sum_{j=0}^{n-1} \sum_{k=0}^{n-1} f''(W_{t_j}) f''(W_{t_k}) ((W_{t_{j+1}} - W_{t_j})^2 - (t_{j+1} - t_j)) \\ & \quad \cdot ((W_{t_{k+1}} - W_{t_k})^2 - (t_{k+1} - t_k)). \end{aligned}$$

We now take the expectation of all this, and examine the diagonal ($j = k$) and off-diagonal ($j \neq k$) separately. The last ones are easy since if, for example, $j < k$, then by independence of future and past-to-present, their expectation equals

$$\mathbb{E}(\dots) \mathbb{E}((W_{t_{k+1}} - W_{t_k})^2 - (t_{k+1} - t_k)) = 0,$$

since $W_{t_{k+1}} - W_{t_k}$ has variance $t_{k+1} - t_k$.

As regards the diagonal terms, we (again) expand, *i.e.*,

$$\begin{aligned} ((W_{t_{j+1}} - W_{t_j})^2 - (t_{j+1} - t_j))^2 & = (W_{t_{j+1}} - W_{t_j})^4 - 2(t_{j+1} - t_j)(W_{t_{j+1}} - W_{t_j})^2 \\ & \quad + (t_{j+1} - t_j)^2. \end{aligned}$$

Inserting this, and using again the independence of future increments, together with

$$\mathbb{E}((W_{t_{j+1}} - W_{t_j})^4) = 3(t_{j+1} - t_j)^3 = \frac{3t^2}{n^2},$$

and

$$\mathbb{E}((W_{t_{j+1}} - W_{t_j})^2) = t_{j+1} - t_j = \frac{t}{n},$$

⁵To be slightly more precise, by ordinary calculus, for each ω , $J_n(\omega) \rightarrow \int_0^t f''(W_s(\omega)) ds$, for any $\omega \in \Omega$, and since everything is bounded, one can use Lebesgue's dominated convergence theorem to conclude that the L^2 -norm of the difference also tends to 0.

we finally obtain

$$\begin{aligned} \left| \mathbb{E} \left(\sum_{j=0}^{n-1} f''(W_{t_j})^2 ((W_{t_{j+1}} - W_{t_j})^2 - (t_{j+1} - t_j))^2 \right) \right| &= \left| \sum_{j=1}^{n-1} \frac{2t^2}{n^2} \mathbb{E}(f''(W_{t_j})^2) \right| \\ &\leq 2C' \cdot \frac{t^2}{n^2} \cdot n = O\left(\frac{1}{n}\right), \end{aligned}$$

which tends to zero as $n \rightarrow \infty$, since $|f''(t)| \leq C$, for all t , implies $\mathbb{E}[f''(W_{t_j})^2] \leq C!$ This proves

Finally, similar arguments can be used to prove (11.35). This is in fact slightly easier since, by Cauchy–Schwarz and the bound (11.32),

$$\begin{aligned} \mathbb{E} \left(\left(\sum_j R_j \right)^2 \right) &\leq n \mathbb{E} \left(\sum_j R_j^2 \right)^2 \\ &\leq Cn \sum_{j=1}^{n-1} \mathbb{E}(W_{t_{j+1}} - W_{t_j})^6 \\ &= (Cn) \cdot \left(\frac{c_6 t^3}{n^3} \right) \cdot n \simeq \frac{1}{n} \rightarrow 0, \end{aligned}$$

where c_6 is the 6th moment of the standard normal. This proves (11.35), and thereby the Itô formula in integral form (11.31). \square

By using Taylor’s formula with integral remainder, we can weaken the hypothesis on f to twice continuously differentiable. We can also do away with the hypothesis that f'' be bounded; we refer to the literature [cite something!].

More general forms of Itô’s lemma, as for $dF(t, W_t)$, can be proved along the same lines (now we also have to expand to first order with respect to the t -variable).