

Itô Integrals

Vladimir Feinberg

September 4, 2021

1 Construction of the Itô integral

We recall our original objective, which is to define some stochastic notion of differential equations parameterized by b, σ to identify some process X_t satisfying

$$X'(t) = b(t, X_t) + \sigma(t, X_t) \cdot \text{“noise”} .$$

It's sensible to consider the forward Euler method, for some shrinking split sequence t_i , which would set

$$\Delta_k X = b(t_k, X_{t_k}) \Delta_k t + \sigma(t_k, X_{t_k}) \Delta_k V ,$$

where $\Delta_k f = f(t_{k+1}) - f(t_k)$ and V_t is some stochastic process.

Based on natural assumptions about $\Delta_k V$ such as stationarity, independent increments, and 0-mean, we might be tempted to choose Brownian motion $V_t = B_t$. In the limit as $\Delta_k t \rightarrow 0$, we would be interested in finding a definition of an integral which is consistent with the limit of the discrete sums implied by the Euler method:

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

where $\int_0^t \sigma(s, X_s) dB_s$ is not yet sensibly defined or shown to exist uniquely as $\lim_n \sum_{k=1}^n \sigma(t_k, X_{t_k}) \Delta_k B$.

Definition 1.1 (Itô Integral). Consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ where $\mathcal{F} = \mathcal{F}_\infty$ for a filtration $\{\mathcal{F}_t\}$. Let $\mathcal{V} = \mathcal{V}(S, T)$ be the class of functions $f : \mathbb{R}_+ \times \Omega \rightarrow \mathbb{R}$ satisfying the following properties:

1. f is $\mathcal{B}(\mathbb{R}_+) \times \mathcal{F}$ measurable.
2. $f(t, \cdot)$ is \mathcal{F}_t -adapted.
3. $L^2(\mathbb{P} \times [S, T])(f) = \mathbb{E}^{1/2} \left[\int_S^T f(t, \omega)^2 dt \right] < \infty$

Then under some shrinking split sequence (see previous chapter notes), the Itô Integral is defined as the limit

$$\int_S^T f(t, \omega) dB_t(\omega) = \lim_{n \rightarrow \infty} \sum_{k=1}^n \phi_n(t_k, B_{t_k}(\omega)) \Delta_k B(\omega) ,$$

where ϕ_n are simple functions chosen by diagonalizing multiple other unique limit sequences. These ϕ_n will be recovered from multiple other sequences as we grow the complexity of f to the full extent of \mathcal{V} . To start, if f is a simple function, then of course $\phi_n = f$ suffices. Afterwards, we show the convergence under $L^2(\mathbb{P} \times [S, T])$ of limit sequences for growing subsets of \mathcal{V} :

1. Bounded and everywhere continuous processes. Since simple functions approximate our continuous process everywhere on Ω , by bounded convergence they approximate in $L^2(\mathbb{P} \times [S, T])$.

2. Bounded processes $g \in \mathcal{V}$. As $n \rightarrow \infty$, consider a bump function ψ_n with shrinking support immediately left of the origin. Consider the convolution $g_n = \psi_n * g$, also in \mathcal{V} by virtue of the negative support of ψ_n . Note that $g_n \rightarrow g$ pointwise as ψ_n progressively approximates the Dirac measure. Then again by bounded convergence we have $g_n \rightarrow g$ in $L^2(\mathbb{P} \times [S, T])$, where g_n are bounded and everywhere continuous by construction.
3. Any process in $g \in \mathcal{V}$. Let g_n be g clamped to $[-n, n]$. By the previous step, we have convergence to each such g_n via limit sequence in \mathcal{V} . Then since $g \in L^2(\mathbb{P} \times [S, T])$ itself, by dominated convergence we have convergence in $g \in L^2(\mathbb{P} \times [S, T])$.

For any $f \in \mathcal{V}$ we may diagonalize the above sequences for some *elementary* functions ϕ_n to define the Itô integral. Why does this definition give rise to a *unique* integral, independent of the choice of ϕ_n (and thus, also, independent of split sequence choice)? To see this, we need a critical tool.

Lemma 1 (Itô Isometry). For any $f \in \mathcal{V}$,

$$\mathbb{E} \left[\int_S^T f(t, \omega)^2 dt \right] = \mathbb{E} \left[\left(\int_S^T f(t, \omega) dt \right)^2 \right],$$

or, in operator form,

$$L^2(\mathbb{P} \times [S, T])(f) = L^2(\mathbb{P}) \circ \int_S^T dt.$$

For elementary functions, this can be proven by computing both sides explicitly and relying on stationarity of increments. In turn, this implies that for any $f \in \mathcal{V}$, for any $L^2(\mathbb{P} \times [S, T])$ sequence ϕ_n of elementary functions, which is necessary Cauchy in that metric, must give rise to a sequence of integrals $I_n = \int_S^T \phi_n dB_t$ which are Cauchy in $L^2(\mathbb{P})$ and thus converge as random variables in this way to a fixed limit $I = I(f, \phi)$ in $L^2(\mathbb{P})$. But for any other similar elementary function sequence ϕ'_n converging to f in $L^2(\mathbb{P} \times [S, T])$, we construct an analogous sequence of integrals I'_n , also Cauchy in $L^2(\mathbb{P})$, thus also yielding an Itô integral $I' = I(f, \phi')$. Through our elementary isometry,

$$\begin{aligned} \|I_n - I'\|_{L^2(\mathbb{P})} &\leq \|I_n - I'_n\|_{L^2(\mathbb{P})} + \|I'_n - I'\|_{L^2(\mathbb{P})} \\ &= \|\phi_n - \phi'_n\|_{L^2(\mathbb{P}) \circ \int_S^T dt} + \|I'_n - I'\|_{L^2(\mathbb{P})} \\ &= \|\phi_n - \phi'_n\|_{L^2(\mathbb{P} \times [S, T])} + \|I'_n - I'\|_{L^2(\mathbb{P})} \\ &\rightarrow 0, \end{aligned}$$

but since I_n converges to both I, I' in $L^2(\mathbb{P})$, by completeness of the corresponding metric space, $I = I'$ almost surely, showing both invariance of the Itô integral to the specific limit sequence ϕ_n chosen and Itô isometry on all of \mathcal{V} .

2 Some properties of the Itô integral

By taking limits from the definition, the following properties of the Itô integral hold almost always:

1. $\int_S^T f dB_t = \int_S^U f dB_t + \int_U^T f dB_t$,
2. $\int_S^T (cf + g) dB_t = c \int_S^T f dB_t + \int_S^T g dB_t$ for constant c ,
3. $\mathbb{E} \int_S^T f dB_t = 0$,
4. $\int_S^T f dB_t \in \mathcal{F}_T$,

5. and $M_s = \int_0^s f dB_t$ is a martingale.

Theorem 1. Let $f \in \mathcal{V}(0, T)$. Then there exists a t -continuous version of $I_t(\omega) = \int_0^t f(s, \omega) dB_s(\omega)$.

Proof. Let ϕ_n be elementary functions tending to f in $L^2(\mathbb{P} \times [0, T])$ from Itô's construction, with corresponding Itô integrals $I_t^{(n)}$. For any n, m , $I_t^{(n)} - I_t^{(m)}$ is a martingale since its components are. [Doob's Martingale Inequality](#), which is basically Markov for martingales, relates

$$\mathbb{P} \left\{ \sup_{0 \leq t \leq T} |I_t^{(n)} - I_t^{(m)}| \geq \epsilon \right\} \leq \frac{1}{\epsilon^2} \mathbb{E} |I_t^{(n)} - I_t^{(m)}|^2 = \frac{1}{\epsilon^2} \|\phi_n - \phi_m\|_{L^2(\mathbb{P} \times [0, T])}^2.$$

Since the RHS vanishes as n, m grow large, a subsequence $I_t^{(n_k)}$ may be chosen ensuring with geometrically decreasing ϵ between adjacent $I_t^{(n_k)}, I_t^{(n_{k+1})}$. This allows Borel-Cantelli to be applied over the set of events $\left\{ \sup_{0 \leq t \leq T} |I_t^{(n_k)} - I_t^{(n_{k+1})}| \geq 2^{-k} \right\}$ for varying $k \in \mathbb{N}$, so in fact with probability one $\sup_{0 \leq t \leq T} |I_t^{(n_k)} - I_t^{(n_{k+1})}| \leq 2^{-k}$ so long as $k \geq K(\omega)$. This provides uniform convergence for almost all ω , and thus continuity in the limit. \square

Note that we *cannot* assume that the elementary functions ϕ_n , known to exist, which define the Itô integral $\int_S^T f dB_t$ for $f \in \mathcal{V}(S, T)$, are of the form $\phi_n(t, \omega) \stackrel{?}{=} \sum_j f(t_j, \omega) 1\{t \in [t_j, t_{j+1})\}$. **Observation.** For general $f \in \mathcal{V}(S, T)$, we do not know a-priori the form of the elementary functions which define its Itô integral.

However, from the original construction of the Itô integral, we do have some basic lemmas which help.

Lemma 2. If $f, f_n \in \mathcal{V}(S, T)$ and $f_n \xrightarrow{L^2(\mathbb{P} \times [S, T])} f$, then $\int_S^T f_n dB_t \xrightarrow{L^2(\mathbb{P})} \int_S^T f dB_t$. Proof immediately follows from Itô's lemma.

Lemma 3. If $f \in \mathcal{V}(S, T)$ almost always has continuous sample paths, then $\phi_n \xrightarrow{L^2(\mathbb{P} \times [S, T])} f$ where $\phi_n(t, \omega) = \sum_j f(t_j, \omega) 1\{t \in [t_j, t_{j+1})\}$.

Proof. A stronger version of this Lemma is proven as an exercise. We need only rely on a weaker notion of continuity; it suffices to show f is CMS; see Exercise 4.13.3 below.

To show an almost always continuous f is CMS, interchange limits by DCT and appeal to continuity.

$$\lim_{s \rightarrow t} \mathbb{E} (f(s, \cdot) - f(t, \cdot))^2 = \mathbb{E} \lim_{s \rightarrow t} (f(s, \cdot) - f(t, \cdot))^2 = 0$$

\square

3 Extensions of the Itô integral

Without much effort, the filtration \mathcal{F} can be expanded to any other filtration \mathcal{H} such that $\mathcal{F}_t \subset \mathcal{H}_t$.

Relaxing the integrator to be an arbitrary semimartingale instead of Brownian motion is possible. Revisiting our construction steps, the only direct properties used were bounded quadratic variation and adaptedness. These can be further relaxed to just [semimartingale](#) properties.

The second moment condition in Itô's construction can be relaxed to a mere condition requiring that the second moment of our integrand is almost surely finite, $\mathbb{P} \left\{ \int_S^T f(s, \omega)^2 ds < \infty \right\} = 1$, however, this of course will end up creating non-integrable Itô limits, and thus drop the martingale property. However, continuity can be recovered, along with a local martingale property.

There also exists another interpretation of stochastic integrals. The Stratonovich is built from a trapezoidal rule (unlike Itô's left-sided integration rule) as the building block:

$$\int_S^T f(t, \omega) \circ dB_t(\omega) = \lim_{n \rightarrow \infty} \sum_{k=1}^n \phi_n \left(\frac{t_k + t_{k+1}}{2}, B_{t_k}(\omega) \right) \Delta_i B(\omega).$$

One reason the Stratonovich interpretation might be more natural is because of the following motivation. If we consider t -continuously differentiable approximations $B_t^{(n)}$ to our noise process B_t which on bounded intervals converge uniformly almost always, then the solution $X_t^{(n)}$ to the differential equation (for constant ω)

$$\frac{dX_t}{dt} = b(t, X_t) + \sigma(t, X_t) \frac{dB_t^{(n)}}{dt} ,$$

converges to the Stratonovich solution $X_t = X_0 + \int_0^t b(t, X_s) ds + \int_0^t \sigma(t, X_s) \circ dB_s$, whereas the solution expressed in the Itô interpretation would be

$$X_t = X_0 + \int_0^t b(t, X_s) ds + \frac{1}{2} \int_0^t \sigma'(s, X_s) \sigma(s, X_s) ds + \int_0^t \sigma(t, X_s) dB_s$$

4 Exercises

4.1

We show

$$\int_0^t s dB_s = tB_t - \int_0^t B_s ds$$

directly from the definition of Itô integrals. We consider the progressively finer elementary functions defined by the discretization of the identity integrand as $n \rightarrow \infty$ for some shrinking split sequence t_j , which tend to the identity in $L^2(\mathbb{P} \times [0, t])$.

$$\sum_i t_i (B_{t_{i+1}} - B_{t_i}) = t_n B_t - \sum_i (t_{i+1} - t_i) B_{t_i} ,$$

so that the limit converges to the desired solution, because almost-always continuity of B_t permits ω -pointwise Riemann integration.

Note the above relies on Lemma 2 and Lemma 3 to derive the Itô integral from the definition with the obvious elementary function approximation.

4.2

We show

$$\int_0^t B_s^2 dB_s = \frac{1}{3} B_t^3 - \int_0^t B_s ds$$

directly from the definition of Itô integrals. A more general technique for solving such integrals from the definition follows in three steps:

1. Rewrite the expected naive Riemann integral solution (here, $\frac{1}{3} B_t^3$) as a telescoping sum. $B_t^3 = \sum_j \Delta_j B^3$.
2. Express each term in an Itô-friendly way, i.e., as functions of increments or the left sided-timestep, i.e., $B_{t_j}, \Delta_j B$.

$$\begin{aligned} \Delta_j B^3 &= \Delta_j B \left(B_{t_{j+1}}^2 + B_{t_{j+1}} B_{t_j} + B_{t_j}^2 \right) \\ &= \Delta_j B \left(B_{t_{j+1}}^2 - 2B_{t_{j+1}} B_{t_j} + B_{t_j}^2 \right) + 3B_{t_{j+1}} B_{t_j} \Delta_j B \\ &= \Delta_j^3 B + 3(\Delta_j B + B_{t_j}) B_{t_j} \Delta_j B \\ &= \Delta_j^3 B + 3B_{t_j} \Delta_j^2 B + 3B_{t_j}^2 \Delta_j B \end{aligned}$$

3. Summing the above formula over j and rearranging terms from the first telescope we have

$$\sum_j B_{t_j}^2 \Delta_j B = \frac{1}{3} B_t^3 - \frac{1}{3} \sum_j \Delta_j^3 B - \sum_j B_{t_j} \Delta_j^2 B .$$

In the limit as $n \rightarrow \infty$, the LHS converges to the desired Itô integral $\int_0^t B_s^2 dB_s$ in $L^2(\mathbb{P})$. The middle term vanishes since $\left| \sum_j \Delta_j^3 B \right| \leq \sum_j |\Delta_j^3 B| \rightarrow [B]_t^{(3)} = 0$. To finish, it remains to show that the last term simplifies to $\int_0^t B_s ds$.

Notice it suffices to show that $\| \sum_j B_{t_j} \Delta_j^2 B - \sum_j B_{t_j} \Delta_j t \|_{L^2(\mathbb{P})} \rightarrow 0$ as the L^2 convergence implies an almost surely convergent subsequence, and by uniqueness of the Itô integral definition, this shows almost sure equality on the final integrals regardless of the elementary function sequence we look at. Indeed,

$$\mathbb{E} \left(\sum_j B_{t_j} (\Delta_j^2 B - \Delta_j t) \right)^2 = \mathbb{E} \sum_j B_{t_j}^2 (\Delta_j^2 B - \Delta_j t)^2 + 2 \mathbb{E} \sum_j \sum_{k>j} B_{t_j} B_{t_k} (\Delta_j^2 B - \Delta_j t) (\Delta_k^2 B - \Delta_k t) .$$

In the first term, independence simplifies calculation, yielding $\mathbb{E} B_{t_j}^2 \mathbb{E} (\Delta_j^2 B - \Delta_j t)^2 = O(tn^{-2})$ term-wise, leveraging the variance of a standard χ^2 distribution, which still vanishes in an n -sized sum. The second term can be decomposed

$$\mathbb{E} [B_{t_j} B_{t_k} (\Delta_j^2 B - \Delta_j t) (\Delta_k^2 B - \Delta_k t)] = \mathbb{E} [B_{t_j} B_{t_k} (\Delta_j^2 B - \Delta_j t)] \mathbb{E} [(\Delta_k^2 B - \Delta_k t)] ,$$

where the last term vanishes. Of course, $\Delta_j t = n^{-1}$ for all j above.

Note the above relies on Lemma 2 and Lemma 3 to derive the Itô integral from the definition with the obvious elementary function approximation.

4.3

Let X_t be a stochastic process.

4.3.1

If X_t is a martingale with respect to any filtration \mathcal{G}_t , it must be a martingale with respect to its own natural filtration \mathcal{F}_t . Measurability and integrability conditions hold immediately, so it suffice to check the conditional martingale property. In addition, by virtue of $X_t \in \mathcal{G}_t$ we must have $\mathcal{F}_t \subset \mathcal{G}_t$ through an inductive argument. Taking $s \leq t$, by tower, $\mathbb{E}[X_t | \mathcal{F}_s] = \mathbb{E}[\mathbb{E}[X_t | \mathcal{G}_s] | \mathcal{F}_s] = \mathbb{E}[X_s | \mathcal{F}_s] = X_s$.

4.3.2

By the above, for any martingale X_t , we have for all $t \geq 0$,

$$\mathbb{E} X_t = \mathbb{E} \mathbb{E}[X_t | \mathcal{F}_0] = \mathbb{E} X_0 .$$

4.3.3

The above may be satisfied for non-martingales; consider $X_t = t\epsilon$ where ϵ is Rademacher-distributed.

4.4

4.4.1

$X_t = B_t + 4t$ is not a martingale because its expectation is strictly increasing with time (see previous problem).

4.4.2

$X_t = B_t^2$ is not a martingale because its expectation is strictly increasing with time.

4.4.3

$X_t = t^2 B_t - 2 \int_0^t s B_s ds$ is a martingale. Integrability and measurability hold immediately. Then notice

$$\mathbb{E}[X_t | \mathcal{F}_s] = t^2 B_s - 2 \int_0^s B_r dr - 2 \int_s^t r \mathbb{E}[B_r | \mathcal{F}_s] dr = t^2 B_s - 2 \int_0^s B_r dr - 2 B_s \int_s^t r dr = X_s .$$

4.4.4

$X_t = B_t^{(1)} B_t^{(2)}$ is a martingale, where $B_t^{(i)}$ are independent brownian motions, so the conditional expectation operator will distribute.

4.5

$M_t = B_t^2 - t$ is a martingale. Measurability and integrability are apparent, and $\mathbb{E}[M_t | \mathcal{F}_s] = B_s^2 + \mathbb{E}[2B_s(B_t - B_s) + (B_t - B_s)^2 | \mathcal{F}_s] - t = M_s$ by splitting up $B_t = B_s + (B_t - B_s)$.

4.6

$N_t = B_t^3 - t B_t$ is a martingale. Measurability and integrability are apparent. Performing the same expansion $B_t = B_s + \Delta$ as before, we find

$$\mathbb{E}[N_t | \mathcal{F}_s] = \mathbb{E}[\Delta^3 + 3\Delta^2 B_s + 3\Delta B_s^2 - 3t B_t + B_s^3 | \mathcal{F}_s] = 3(t - s) B_s - 3t B_s + B_s^3 = N_s$$

4.7

Consider the iterated Itô integral

$$I_n = n! \int_0^t \int_0^{u_n} \cdots \int_0^{u_2} dB_{u_1} dB_{u_2} \cdots dB_{u_n} .$$

4.7.1

For every tuple u_2, \dots, u_n , necessarily in sorted order already, the innermost integrand 1 is measurable with respect to $\mathcal{B}(\mathbb{R}_+) \times \mathcal{F}$. It is also adapted to \mathcal{F}_{u_1} trivially. Finally, it's square integrable, as the entire integral $\int_0^{u_2} 1^2 dB_1 \leq u_2 \leq t$. The resulting integral $I_1^{(n)}(u_2, \dots, u_n) = \int_0^{u_2} 1 dB_1$ is by properties of Itô integration itself measurable and adapted to \mathcal{F}_{u_2} , as well as dominated by t over $u_2 \in [0, u_3]$.

Then, inductively, assume for a tuple u_i, \dots, u_n , the corresponding integrand $\int_0^{u_i} \cdots dB_{u_{i-1}}$ is measurable, adapted, and square integrable due to being dominated by $t^{i-1}/(i-1)!$. Then adding another integral again results in a measurable and adapted random process dominated in $L^2(\mathbb{P} \times [S, T])$ by $t^i/i!$.

4.7.2

Manually, we can verify that

$$I_1 = B_t ,$$

and since $\int_0^{u_2} dB_{u_1} = B_{u_2}$ that

$$I_2 = B_t^2 - t ,$$

and finally,

$$I_3 = 3 \int_0^t B_{u_2}^2 - u_2 \, dB_{u_2} = B_t^3 - 3tB_t ,$$

by combining Exercises 4.1 and 4.2.

We notice that all of the above are of the form $I_n = t^{n/2} h_n \left(\frac{B_t}{\sqrt{t}} \right)$, where h_n is the Hermite polynomial of degree n , where

$$h_n(x) = (-1)^n \exp \left(\frac{x^2}{2} \right) \partial_x^n \exp \left(-\frac{x^2}{2} \right)$$

4.7.3

By the above, N_t from Exercise 4.6 must be a martingale as it is an Itô integral.

4.8

4.8.1

Let Y be an integrable rv and let $M_t = \mathbb{E}[Y | \mathcal{F}_t]$ M_t is a martingale, following immediately from the tower property.

4.8.2

Observation. The below doesn't seem to hold without the additional continuity assumption, which disagrees with the primary text.

Conversely, suppose M_t is a real-valued \mathcal{F}_t -martingale which satisfies for some p that $\sup_t \mathbb{E} |M_t|^p < \infty$. We can show there exists some integrable Y such that $M_t = \mathbb{E}[Y | \mathcal{F}_t]$.

By Corollary C.7 in the appendix (also, for discrete settings, by Theorem 4.7 from Çinlar's Probability and Stochastics, p.201, Theorem 4.7, the characterization of uniformly integrable martingales), *with the additional assumption that M_t is continuous, an assumption absent from Øksendal*, we have that an integrable Y exists such that $M_t \xrightarrow{L^1} Y$.

Then, notice that $\mathbb{E}[Y - X_t | \mathcal{F}_t] = 0$ if $\mathbb{E}[1_H(Y - X_t)] = 0$ is for all events $H \in \mathcal{F}_t$. But $1_H(X_s - X_t) \xrightarrow{L^1} 1_H(Y - X_t)$ as $s \rightarrow \infty$ by analyzing the cases of $1_H \in \{0, 1\}$. So $\mathbb{E}[1_H(Y - X_t)] = \lim_{s \rightarrow \infty} \mathbb{E}[1_H(X_s - X_t)]$ where for $s \geq t$

$$\mathbb{E}[1_H(X_s - X_t)] = \mathbb{E}[1_H \mathbb{E}[X_s - X_t | \mathcal{F}_t]] = 0 .$$

The continuity assumption seems load-bearing, but perhaps uninterestingly so. With ϵ a Rademacher random variable, $M_t = \epsilon 1\{t \geq 1\}$, and an unnaturally large constant filtration $\mathcal{H}_t = \sigma\{\epsilon\}$ (note the natural filtration has $\mathcal{F}_t = \emptyset$ for $t < 1$), then $M_t = 0 \neq \epsilon = \mathbb{E}[\epsilon | \mathcal{H}_t]$ for any $t < 1$.

4.9

Recall the Stratonovich integral is defined as

$$\int_0^T f(t, \omega) \circ dB_t(\omega) = \lim_n \sum_j f(t_j^*, \omega) \Delta_j B ,$$

where $t_j^* = \frac{t_j + t_{j+1}}{2}$. We suppose that for $f \in \mathcal{V}(0, T)$ and continuous (such that Lemma 2 and Lemma 3 apply) the above Stratonovich integral exists in $L^2(\mathbb{P})$ and is unique regardless of choice of shrinking split partition (without proof).

We compute $\int_0^T B_t \circ dB_t$.

Consider the term of the sum on the RHS, $f(t_j^*, \omega) \Delta_j B$, for $f(t, \omega) = B_t(\omega)$.

$$\begin{aligned}
B_{t_j^*} \Delta_j B &= B_{t_j^*} \left((B_{t_{j+1}} - B_{t_j^*}) + (B_{t_j^*} - B_{t_j}) \right) \\
&= B_{t_j^*} (B_{t_{j+1}} - B_{t_j^*}) + B_{t_j} (B_{t_j^*} - B_{t_j}) - B_{t_j} (B_{t_j^*} - B_{t_j}) + B_{t_j^*} (B_{t_j^*} - B_{t_j}) \\
&= \sum_{i=2j-1}^{2j} B_{t_i^*} \Delta_i^\dagger B + (B_{t_j^*} - B_{t_j})^2
\end{aligned}$$

Where $\{t_i^\dagger\}_i$ iterates over twice as many points as $\{t_j\}_{j=1}^n$ and includes the midpoints of the latter sequence. Similarly, $\Delta_i^\dagger g = g(t_{i+1}^\dagger) - g(t_i^\dagger)$.

Introduce the notation $a_n \xrightarrow{L^2(\mathbb{P})} b_n$ when $\lim_n \|a_n - b_n\|_{L^2(\mathbb{P})} = 0$ and one of the limits converges as $a_n \xrightarrow{L^2(\mathbb{P})} a$ to the same a for any subsequence of a_n . This implies that $a_n, b_n \xrightarrow{\text{as}} a$ and $b_n \xrightarrow{L^2(\mathbb{P})} a$ as well.

Then, starting from assumption,

$$\begin{aligned}
\int_0^T B_t \circ dB_t &\stackrel{L^2(\mathbb{P})}{=} \sum_j B_{t_j^*} \Delta_j B \\
&\stackrel{L^2(\mathbb{P})}{=} \sum_i B_{t_i^*} \Delta_i^\dagger B + \sum_j (B_{t_j^*} - B_{t_j})^2 \\
&\stackrel{L^2(\mathbb{P})}{=} \int_0^T B_t dB_t + \sum_j (B_{t_j^*} - B_{t_j})^2.
\end{aligned}$$

Next, notice that each term of the sum $\sum_j (B_{t_j^*} - B_{t_j})^2$ is jointly independent and distributed exactly as each term of the sum $\sum_{i=1}^n (\Delta_i^\dagger B)^2$, which is just the half interval, respectively. Following the proof from Chapter 2 Exercise 17, we can re-derive what is essentially the quadratic variation $[B]_t^{(2)}$ over the interval $[0, T/2]$, so that $\sum_j (B_{t_j^*} - B_{t_j})^2 = \frac{T}{2}$. Combining this with the known Itô integral $\int_0^T B_t dB_t$, we have that

$$\int_0^T B_t \circ dB_t = \frac{1}{2} B_T^2.$$

4.10

Consider the setup in Exercise 4.9 for any deterministic choice $t_j^* \in [t_j, t_{j+1}]$, not necessarily the midpoint. Recall $f \in \mathcal{V}(0, T)$ and let $f_t = f(t, \cdot)$ be shorthand for the corresponding random variable. Further, assume that

$$\mathbb{E} |f_s - f_t|^2 \leq K |s - t|^{1+\epsilon}$$

for any fixed $\epsilon > 0$.

Define $I_n^*(\omega) = \sum_j f(t_j^*, \omega) \Delta_j B$. We show that $I_n^* \xrightarrow{L^1(\mathbb{P})} \int_0^T f dB_t$. Let $I_n(\omega) = \sum_j f(t_j, \omega) \Delta_j B$ be the standard elementary sum approximating our Itô integral.

Before we begin, notice that through Jensen concavity, we can derive a weaker L^1 continuity:

$$\mathbb{E}^2 |f_s - f_t| \leq \left(\mathbb{E}^{1/2} |f_s - f_t|^2 \right)^2 = \mathbb{E} |f_s - f_t|^2 \leq K |s - t|^{1+\epsilon}.$$

By DCT,

$$\lim_n \mathbb{E} |I_n - I_n^*| = \lim_n \mathbb{E} \left| \int f dB_t - \lim_n I_n \right|,$$

so we need only show the LHS vanishes.

$$\begin{aligned}
\mathbb{E} |I_n - I_n^*| &\leq \sum_j \mathbb{E} \left| (f_{t_j} - f_{t_j^*}) \Delta_j B \right| \\
&\leq \mathbb{E} |\Delta_1 B| \sum_j \mathbb{E} |f_{t_j} - f_{t_j^*}| \\
&\leq \sqrt{\frac{2}{\pi n}} \sum_j K^{1/2} |t_j^* - t_j|^{0.5+\epsilon/2} = O(n^{-\epsilon/2}) ,
\end{aligned}$$

Since $|t_j^* - t_j| \leq \frac{1}{n}$ by choice of split sequence.

Observation. One of my earlier proof attempts made it seem like one could use the much weaker condition,

$$\mathbb{E} |f_s - f_t|^2 \leq K |s - t|^\delta ,$$

holding for any fixed $\delta > 0$. With this, I thought I could prove an even stronger result, that $I_n^* \xrightarrow{L^2(\mathbb{P})} \int_0^T f dB_t$, but was eventually able to find the bug.

First, we define the elementary functions

$$\phi_n^*(t, \omega) = \sum_j f(t_j^*, \omega) 1_{\{t \in [t_j, t_{j+1})\}} .$$

Now $I_n^* = \int_0^T \phi_n^* dB_t$. Let ϕ_n be an elementary function sequence such that $\phi_n \rightarrow f$ in $L^2(\mathbb{P} \times [0, T])$ (and therefore $I_n = \int_0^T \phi_n dB_t$ satisfies $I_n \xrightarrow{L^2(\mathbb{P})} \int_0^T f dB_t$). As before, by DCT suffices to show that the following vanishes:

$$\begin{aligned}
\|I_n - I_n^*\|_{L^2(\mathbb{P})} &= \|\phi_n - \phi_n^*\|_{L^2(\mathbb{P} \times [0, T])} \\
&= \mathbb{E} \sum_j \left(f_{t_j} - f_{t_j^*} \right)^2 \Delta_j t \\
&\leq \frac{K}{n} \sum_j |t_j^* - t_j|^{1+\epsilon} = O(n^{-1-\epsilon}) ,
\end{aligned}$$

where the critical first step supposedly follows from Itô's lemma.

However, upon closer inspection we notice $I_n^* = \int_0^T \phi_n^* dB_t$ is not well-defined, because $\phi_n^*(t, \omega)$, though elementary, is not in $\mathcal{V}(0, T)$ as it is not \mathcal{F}_t -measurable. This demonstrates the importance of the filtration measurability condition.

4.11

Consider a stochastic process W_t such that all of the following attributes hold:

1. W_t is stationary; namely the joint distribution of $\{W_{t_1+t}, W_{t_2+t}, \dots, W_{t_k+t}\}$ does not depend on t for any finite collection $\{t_i\}_{i=1}^k \subset \mathbb{R}_+$,
2. $W_t \perp W_s$ where $t \neq s$,
3. $\mathbb{E} W_t = 0$ for all t ,
4. and W_t is continuous.

In this case, W_t must be trivial.

Define for fixed $N > 0$ $W_t^{(N)} = -N \vee (N \wedge W_t)$ as the original process clamped to $[-N, N]$. Then by the bounded convergence theorem,

$$\lim_{s \rightarrow t} \mathbb{E} \left(W_t^{(N)} - W_s^{(N)} \right)^2 = \mathbb{E} \lim_{s \rightarrow t} \left(W_t^{(N)} - W_s^{(N)} \right)^2 = 0 .$$

And yet, by expanding the square and applying independence, we have that for $s \neq t$ that

$$\lim_{s \rightarrow t} \mathbb{E} \left(W_t^{(N)} - W_s^{(N)} \right)^2 = \lim_{s \rightarrow t} \mathbb{E} \left(W_t^{(N)} \right)^2 + \mathbb{E} \left(W_s^{(N)} \right)^2 = 2 \mathbb{E} \left(W_t^{(N)} \right)^2 ,$$

where we must actually apply BCT again.

But all this implies that $\mathbb{E} \left(W_t^{(N)} \right)^2 = 0$. But this implies $W_t^{(N)} = 0$ almost surely for any N . Taking limits we have our result.

4.12

For brevity, given some endpoints S, T in context (typically, $S = 0$, the notation

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

is taken to be an almost sure equality

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

so that corresponding assumptions are placed such as $\sigma \in \mathcal{V}(S, T)$ and b is Lebesgue integrable.

In this case, per the chapter text, if the Stratonovich process

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

exists (which we have assumed under the implicit assumptions we've equipped this notation with), then the corresponding Itô below must be almost surely equal to it:

$$dX_t = \left(b(X_t, t) - \frac{1}{2} \sigma(X_t, t) \partial_1 \sigma(X_t, t) \right) dt + \sigma(X_t, t) \circ dB_t .$$

Observation. This results in a straightforward computation for going in the Stratonovich to Itô direction of converting stochastic differential equations. But this also shows us that Itô process which happen to also be Stratonovich (the other direction) are rarer; they require the additional regularity $\partial_1 \sigma \in \mathcal{V}$ on the noise term. This makes sense; an Itô process is necessarily a martingale, but a Stratonovich one is not necessarily one. However, as described in [these notes](#) comparing the two integrals, this difference is overblown most of the time. The below exercises, which convert from one to the other, demonstrate this.

4.12.1

$$\begin{aligned} dX_t &= \gamma X_t dt + \alpha X_t \circ dB_t \\ &= \left(\gamma - \frac{1}{2} \alpha^2 \right) X_t dt + \alpha X_t dB_t \end{aligned}$$

4.12.2

$$\begin{aligned} dX_t &= \sin X_t \cos X_t dt + (t^2 + \cos X_t) \circ dB_t \\ &= \frac{\sin X_t}{2} (\cos X_t - t^2) X_t dt + (t^2 + \cos X_t) dB_t \end{aligned}$$

4.12.3

$$\begin{aligned}
dX_t &= rX_t dt + \alpha X_t dB_t \\
&= \left(r - \frac{\alpha^2}{2}\right) X_t dt + \frac{1}{2} (\alpha X_t) \alpha dt + \alpha X_t dB_t \\
&= \left(r - \frac{\alpha^2}{2}\right) X_t dt + \alpha X_t \circ dB_t
\end{aligned}$$

4.12.4

$$\begin{aligned}
dX_t &= 2 \exp(-X_t) X_t dt + X_t^2 dB_t \\
&= (2 \exp(-X_t) - X_t^3) X_t dt + \frac{1}{2} X_t^2 (2X_t) dt + X_t^2 dB_t \\
&= (2 \exp(-X_t) - X_t^3) X_t dt + X_t^2 \circ dB_t
\end{aligned}$$

4.13

A stochastic process X_t is considered to be continuous in mean square (CMS) if for all t it satisfies $EX_t^2 < \infty$ and $\lim_{s \rightarrow t} \mathbb{E}(X_t - X_s)^2 = 0$

4.13.1

B_t is CMS because $\mathbb{E}(B_s - B_t)^2 = |s - t| \rightarrow 0$.

4.13.2

If f is C -Lipschitz then $f(B_t)$ is CMS because $\mathbb{E}(f(B_s) - f(B_t))^2 \leq C \mathbb{E}|B_s - B_t|^2 = C|s - t| \rightarrow 0$.

4.13.3

Let $X_t \in \mathcal{V}(S, T)$ be CMS. Then we show

$$\int_0^T X_t dB_t = \lim_n \int_S \phi_n(t, \omega) dB_t(\omega),$$

where $\phi_n(t, \omega) = \sum_j X_{t_j^{(n)}}(\omega) 1\{t \in [t_j, t_{j+1})\}$.

Per Lemma 2, it suffices to show convergence in $L^2(\mathbb{P} \times [S, T])$.

$$\begin{aligned}
\mathbb{E} \int_S^T (X_t - \phi_n(t, \cdot))^2 dt &= \sum_j \mathbb{E} \int_{t_j^{(n)}}^{t_{j+1}^{(n)}} (X_t - X_{t_j^{(n)}})^2 dt \\
&= \sum_j \Delta_j t \mathbb{E} \sup_{s \in [t_j^{(n)}, t_{j+1}^{(n)}]} (X_{t_j} - X_s)^2 \\
&\leq \frac{1}{n} \sum_j \epsilon_n^2 = \epsilon_n^2 \rightarrow 0,
\end{aligned}$$

where we apply a uniform bound $\epsilon_n \geq \mathbb{E}^{1/2}(X_t - X_s)^2$ for any t, s within n^{-1} distance of one another holding on $[S, T]$, with $\epsilon_n \rightarrow 0$.

The aforementioned ϵ_n sequence exists by virtue of uniform continuity of X as a function from indices in $[S, T]$ to random variables in $\Omega \rightarrow \mathbb{R}$.

CMS is by definition continuity of $X : [S, T] \rightarrow \mathbb{R}^\Omega$ viewing $[S, T]$ as a standard Euclidean compact metric space and \mathbb{R}^Ω as a standard topological space induced by the metric $L^2(\mathbb{P})$. By compactness X must be uniformly continuous.

4.14

A function $h(\omega)$ is \mathcal{F}_t measurable if and only if it can be written as

$$h(\omega) = \sum_{i=1}^{\infty} \prod_{j=1}^{k_i} g_j^{(i)} \left(B_{t_j^{(i)}} \right) ,$$

where \mathcal{F}_t is the natural filtration over Brownian motion up to time t , and the infinite sum converges as a limit to h almost surely.

Considering clamped versions of h and diagonalizing, without loss of generality h may be bounded.

Next, consider any shrinking split sequence $t_j^{(n)}$ and let $\mathcal{H}_n = \sigma \left\{ B_{t_j^{(n)}} \right\}_j$. By applying Øksendal Appendix C.9, $\mathbb{E}[h | \mathcal{H}_n] \xrightarrow{\text{as}} \mathbb{E}[h | \mathcal{H}_\infty]$.

First, we show that, almost surely,

$$\mathbb{E}[h | \mathcal{H}_\infty] = \mathbb{E}[h | \mathcal{F}_t] = h ,$$

which follows from the near-left-continuity of the filtration $\mathcal{F}_s \subset \bigcap_{\epsilon < 0} \mathcal{F}_{s+\epsilon} \subset \bigcup_n \mathcal{H}_n$ for each s (since we can take intersection over rational ϵ).

Then by Doob-Dynkin, $\mathbb{E}[h | \mathcal{H}_n] = G_n(B_{t_1}, B_{t_2}, \dots, B_{t_k})$ for some Borel G_n ; which per the text can be approximated by a sequence of continuous functions F_n (we diagonalize the iteration again).

Stone-Weierstrass then tells us that each such continuous function can be approximated by a polynomial basis; which yields the desired result after expansion and a further diagonalization of iteration.

4.15

Suppose $f, g \in \mathcal{V}$ and that

$$C + \int_S^T f dB_t = D + \int_S^T g dB_t$$

holds almost surely.

Then, since the equation holds in expectation, it must be that $C = D$. This implies the integrals are almost surely equal (*).

Let $\phi_n \rightarrow f$ and $\gamma_n \rightarrow g$ in $L^2(\mathbb{P} \times [S, T])$, elementary functions which must exist by Itô's construction. Then,

$$\begin{aligned} 0 &= \left\| \int_S^T f dB_t - \int_S^T g dB_t \right\|_{L^2(\mathbb{P})} & (*) \\ &= \left\| \lim_n \int_S^T \phi_n dB_t - \int_S^T \gamma_n dB_t \right\|_{L^2(\mathbb{P})} \\ &= \lim_n \left\| \int_S^T \phi_n dB_t - \int_S^T \gamma_n dB_t \right\|_{L^2(\mathbb{P})} & \text{DCT } L^2(\mathbb{P}) \\ &= \lim_n \|\phi_n - \gamma_n\|_{L^2(\mathbb{P} \times [S, T])} & \text{Itô's Lemma} \\ &= \left\| \lim_n \phi_n - \gamma_n \right\|_{L^2(\mathbb{P} \times [S, T])} & \text{DCT } L^2(\mathbb{P} \times [S, T]) \\ &= \|f - g\|_{L^2(\mathbb{P} \times [S, T])} , \end{aligned}$$

which implies that f, g are indeed $([S, T] \times \Omega)$ -almost equal.

4.16

By Jensen's inequality, for a square integrable X (it must be square integrable to define the conditional expectation), for any σ -algebra \mathcal{F} :

$$\mathbb{E} \mathbb{E}[X | \mathcal{F}]^2 \leq \mathbb{E} \mathbb{E}[X^2 | \mathcal{F}] = \mathbb{E} X^2 ,$$

by applying the tower property.

4.17

Let X be an integrable random variable and \mathcal{G} be a finite σ -algebra generated by a partition $\{G_i\}$ of Ω , with nonnull parts. We then show that, surely,

$$\mathbb{E}[X | \mathcal{G}] = \sum_i \frac{1_{G_i}}{\mathbb{P} G_i} \int_{G_i} X \, d\mathbb{P} .$$

Recall by the projection property that $\mathbb{E}[X | \mathcal{G}]$ is the unique random variable satisfying

$$\mathbb{E}[\mathbb{E}[X | \mathcal{G}] 1_G] = \mathbb{E}[X 1_G]$$

for all $G \in \mathcal{G}$. Taking in particular $G = G_i$; as shown in Exercise 2.7, as $\mathbb{E}[X | \mathcal{G}]$ must be constant on G_i . Let this constant be v_i . To match expectations, we must have

$$v_i \mathbb{P} G_i + 0 = \mathbb{E}[X 1_{G_i}] ,$$

by directly evaluating the expectation on the LHS and solving for v_i concludes the proof.