General stochastic calculus

Matthis Lehmkuehler

23/05/2024

1. Introduction

Stochastic calculus is arguably one of the most important contributions to 20th century mathematics and is now one of the cornerstones of probability theory. One of the problems with the standard treatment in textbooks of this subject is that the presentation is still very close to the historical development of the subject and in particular for the case of general processes with jumps, there are now more direct constructions available.

Here, we will use a much more direct approach following [1], [2], [3] and in particular [4]. The key will be the right notion of integrability which captures the *stochastic cancellation* effect.

Let us now present a brief outline: First, we are going to describe the right notion of integrability and then define the stochastic integral. We then prove some standard results on stochastic integrals and in particular establish Itô's formula, definitely the most important result of the subject. The last step in these notes is establishing easy-to-check conditions that imply integrability and that can be used in practice easily. The Bichteler-Dellacherie theorem, which is a kind of *if and only if* statement in this context, will not be discussed, see however [5], [6] for more information on this. The reader is expected to be familiar with basic martingale results in the discrete and the continuum setting at the level of for example [7].

2. Construction of the stochastic integral

The key of the approach presented here, will be the right choice of definitions. We begin with some well-known notions. Throughout this text, $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}\right)$ will be a filtered probability space satisfying the usual conditions i.e. $\mathcal{F}_t = \mathcal{F}_{t+}$ where $\mathcal{F}_{t+} = \cap_{s>t} \mathcal{F}_s$ and $\mathcal{N} \subseteq \mathcal{F}_0$ where we let $\mathcal{N} = \{A \in \mathcal{F}_\infty : \mathbb{P}(A) \in \{0,1\}\}$ and $\mathcal{F}_\infty = \sigma(\cup_{t\geq 0} \mathcal{F}_t)$.

In these notes, a *càdlàg process* will be a random process $X : [0,\infty) \to \mathbb{R} \cup \{\partial\}$ (we use ∂ to denote a cemetery state) such that on $(0, \zeta_X)$, X is right-continuous with left limits and takes values in $\mathbb{R}$, and on $[\zeta_X, \infty)$, X is constant with value ∂ where $\zeta_X = \inf\{t \geq 0 : X_t = \partial\}$. Operations on càdlàg processes (like stopping, addition or scalar multiplication) are defined in the usual way where it is understood that evaluating any expression involving ∂ yields ∂ as the result. We also set

$$X_t^* = \begin{cases} \sup_{s\in[0,t]} |X_s| : t < \zeta_X, \\ \infty \qquad\qquad\quad : t \geq \zeta_X. \end{cases}$$

We say $X_n \to X$ u.c.p. (uniformly on compacts in probability) as $n \to \infty$ for X and (X_n) càdlàg processes if $(X - X_n)_t^* 1(t < \zeta_X) \to 0$ in probability as $n \to \infty$ for all $t \geq 0$.

Definition 2.1 (Predictable and simple processes). We call

$$\mathcal{P} = \sigma(\{0\} \times A : A \in \mathcal{F}_0) \vee \sigma((s,t] \times A : 0 \le s < t \text{ and } A \in \mathcal{F}_s)$$

the *predictable σ-algebra* on $[0,\infty) \times \Omega$. A process $\xi : [0,\infty) \times \Omega \to \mathbb{R}$ is said to be *predictable* if it is measurable w.r.t. $\mathcal{P}$ and in this case, we write (by standard abuse of notation) $\xi \in \mathcal{P}$. Moreover, we let $\mathcal{P}_b$ be the set of predictable bounded processes. A process σ is *simple* if it is of the form

$$\sigma_t = \alpha_0 1_{\{0\}}(t) + \sum_{i=1}^{n} \alpha_i 1_{(\tau_i, \tilde{\tau}_i]}(t) \quad \text{for } t \ge 0, \tag{1}$$

where $0 \le \tau_i \le \tilde{\tau}_i$ are stopping times and $\alpha_0 \in \mathcal{F}_0$ and $\alpha_i \in \mathcal{F}_{\tau_i}$ for $i \ge 1$ are random variables. We write Σ for the set of simple processes. Note that Σ is a vector space and moreover $\sigma \wedge \sigma', \sigma \vee \sigma' \in \Sigma$ whenever $\sigma, \sigma' \in \Sigma$. Finally, we define the set

$$\tilde{\Sigma} = \{A \subseteq [0,\infty) \times \Omega : 1_A \in \Sigma\}.$$

Remark 2.2. It is easy to show that $\mathcal{P}$ is generated by all left-continuous adapted processes. From the definition, it is easy to see that if ξ is simple, it is left-continuous and adapted, hence predictable. Also, by only considering simple processes with deterministic stopping times, one sees that $\tilde{\Sigma}$ is a π-system generating $\mathcal{P}$.

Definition 2.3 (Stochastic integrals with simple integrands). Let X be an adapted càdlàg process and $\sigma \in \Sigma$ be of the form (1). The *stochastic integral* $\sigma \cdot X$ is defined to be the adapted càdlàg process

$$(\sigma \cdot X)_t = \begin{cases} \sum_{i=1}^{n} \alpha_i \left(X_{\tilde{\tau}_i \wedge t} - X_{\tau_i \wedge t} \right) & : t < \zeta_X, \\ \partial & : t \ge \zeta_X. \end{cases}$$

We also write $\sigma \cdot X = \int \sigma \, dX$ and $(\sigma \cdot X)_t = \int_0^t \sigma_s \, dX_s$. It is immediate that $\sigma \cdot X$ is well-defined and linear in σ and X.

Remark 2.4. Let us remark that, just like for conditional expectations, in the case of stochastic integration, for general $\xi \in \mathcal{P}$, $\xi \cdot X$ will be *a* version of the stochastic integral and there will be no preferred (adapted càdlàg) version.

We leave the following important (but easy) lemma as an exercise.

Lemma 2.5. Let $\sigma \in \Sigma$, X be càdlàg and adapted and τ a stopping time, then we have $\left(1_{[0,\tau]}\sigma\right) \cdot X = \sigma \cdot X^\tau = (\sigma \cdot X)^\tau$ where $\left(1_{[0,\tau]}\sigma\right)_t = 1_{[0,\tau]}(t)\sigma_t$ so that $1_{[0,\tau]}\sigma \in \Sigma$.

Our goal will be to construct stochastic integrals. To do so, we will first extend the stochastic integral from Σ to $\mathcal{P}_b$ and then to the most general case of what will be called integrable processes. To avoid having to repeat the same properties over and over again, we make the following definition.

Definition 2.6 (Stochastic calculus). Fix an adapted càdlàg process X. We say that a family of predictable processes $\mathcal{G}$ admits a stochastic calculus for X if to any $\xi \in \mathcal{G}$ we can associate a càdlàg adapted process $\xi \cdot X$ such that the following properties hold:

(i) If $\sigma \in \Sigma$ is bounded then $\sigma \in \mathcal{G}$ and $\sigma \cdot X$ is given by Definition 2.3.
(ii) Stochastic dominated convergence: Let $\eta \in \mathcal{G}$, $\xi \in \mathcal{P}$ and $(\xi_n) \subseteq \mathcal{P}$ be such that $\xi_n \to \xi$ pointwise and $|\xi_n| \leq |\eta|$ for all $n \geq 1$. Then $(\xi_n) \subseteq \mathcal{G}$, $\xi \in \mathcal{G}$ and $\xi_n \cdot X \to \xi \cdot X$ u.c.p.
(iii) Linearity: If $\xi, \eta \in \mathcal{G}$ and $\lambda, \mu \in \mathbb{R}$, then $\lambda\xi + \mu\eta \in \mathcal{I}_X$ and $(\lambda\xi + \mu\eta) \cdot X = \lambda(\xi \cdot X) + \mu(\eta \cdot X)$ a.s.
(iv) Uniqueness: If $\xi * X$ for $\xi \in \mathcal{G}$ are càdlàg adapted processes satisfying the same properties (i), (ii) and (iii) as the processes $\xi \cdot X$ above, then $\xi * X = \xi \cdot X$ a.s. for all $\xi \in \mathcal{G}$.

The properties are very powerful since they allow standard measure theoretic arguments to be applied in order to extend results about stochastic integrals from simple processes to the general case: Indeed, one first verifies the desired statement on simple processes $\xi \in \Sigma$, and in particular for $\xi = 1_A$ when $A \in \tilde{\Sigma}$, then using Dynkin's lemma, one extends the result to $\xi = 1_A$ for $A \in \mathcal{P}$, and by taking linear combinations and a limit via stochastic dominated convergence, one finally deduces the desired statement for all $\xi \in \mathcal{G}$.

The last definition we need is that of integrability.

Definition 2.7 (Good integrators and integrability). For X an adapted càdlàg process. We say that X is a good integrator if

$$\sup_{\sigma\in\Sigma:\ |\sigma| \leq 1} \mathbb{P}\Big(|(\sigma \cdot X)_t| > C, t < \zeta_X\Big) \to \infty \quad \text{as } C \to \infty \text{ for all } t \geq 0.$$

If $\mathcal{P}_b$ admits a stochastic calculus for X, we say that $\xi \in \mathcal{P}$ is X-integrable if

$$\sup_{\xi'\in\mathcal{P}_b:\ |\xi'| \leq |\xi|} \mathbb{P}\Big(|(\xi' \cdot X)_t| > C, t < \zeta_X\Big) \to \infty \quad \text{as } C \to \infty \text{ for all } t \geq 0$$

and write $\mathcal{I}_X$ for the set of X-integrable processes. Both of the previous two definitions capture a form of tightness for families of stochastic integrals. Lastly for $\xi \in \mathcal{P}$, we make the following definitions which will be key in our proof strategy:

$$\|\xi\|^0_{X,t} = \sup_{\sigma'\in\Sigma:\ |\sigma'| \leq |\xi|} \mathbb{E}\Big(|(\sigma' \cdot X)_t| \wedge 1; t < \zeta_X\Big) \quad \text{and}$$

$$\|\xi\|_{X,t} = \inf\left\{\sum_n \|\sigma_n\|^0_{X,t} : |\xi| \leq \sum_n |\sigma_n| \text{ with } (\sigma_n) \subseteq \Sigma\right\} \in [0, \infty].$$

Note that $\|\xi\|_{X,t}$ is monotone and countably subadditive in $|\xi|$ and we will make use of these properties throughout without explicit mention.

The purpose of this whole section is to prove the following theorem. One should understand this as a result which improves tightness (as it appears in the previous definition) to convergence of stochastic integrals. Tightness is frequently easy to check (using moment bounds for example) which makes this theory very useful in practice.

Theorem 2.8 (Construction of stochastic integrals). If X be a good integrator, $\mathcal{P}_b$ admits a stochastic calculus for X. Moreover in this case, the set $\mathcal{I}_X$ of X-integrable processes is defined and $\mathcal{I}_X$ admits a stochastic calculus for X.

Note that we first have to construct the integral on a smaller subset of processes just in order to define the general notion of integrability. This is very analogous to first having to define (usual) integration with respect to some measure for non-negative functions in order to define when a general real-valued function is integrable.

The idea of the proof of the first part will be to show that for each $\xi \in \mathcal{P}_b$, there exists a sequence $(\sigma_n) \subseteq \Sigma$ such that $\|\xi - \sigma_n\|_{X,t} \to 0$ as $n \to \infty$. Moreover, we will establish the bound

$$\mathbb{E}\Big((\sigma_n \cdot X - \sigma_m \cdot X)_t^* \wedge 1; t < \zeta_X\Big) \leq \big(\|\sigma_n - \sigma_m\|_{X,t} \wedge 1\big)\big(1 - \log\big(\|\sigma_n - \sigma_m\|_{X,t} \wedge 1\big)\big).$$

A standard Cauchy sequence argument then allows us to construct the stochastic integral. The further extension to all processes $\mathcal{I}_X$ follows a similar strategy.

2.1. Technical lemmas

The construction of the stochastic integral builds on a few technical lemmas. The following lemma is essentially a stochastic version of the fact that an absolutely convergent series has terms going to 0.

Lemma 2.9. Let X be a good integrator and $(\sigma_n) \subseteq \Sigma$ are such that $\sum_n |\sigma_n| \leq 1$, then we have that for all $t \geq 0$, $(\sigma_n \cdot X)_t 1(t < \zeta_X) \to 0$ in probability as $n \to \infty$.

Proof. We will in fact even show $\sum_n (\sigma_n \cdot X)_t^2 < \infty$ almost surely on the event $\{t < \zeta_X\}$. Let (ε_i) be a sequence of i.i.d. random variables with distribution $\mathbb{P}(\varepsilon_i = \pm 1) = 1/2$. First, note that by the central limit theorem, for any sequence of real numbers (x_i) satisfying $\sum_i x_i^2 = \infty$, we have

$$\frac{\sum_{i=1}^n \varepsilon_i x_i}{\left(\sum_{i=1}^n x_i^2\right)^{1/2}} \xrightarrow{d} N(0,1) \quad \text{as } n \to \infty.$$

It is easy to deduce that for all $C > 0$, $\mathbb{P}\Big(|\sum_{i=1}^n \varepsilon_i x_i| \leq C\Big) \to 0$ as $n \to \infty$. By conditioning on $\Big((\sigma_i \cdot X)_t : i \geq 1\Big)$, applying the above to $x_i = (\sigma_i \cdot X)_t$ on the event $\Big\{\sum_i (\sigma_i \cdot X)_t^2 = \infty\Big\}$ and dominated convergence, we get for any $C > 0$,

$$\mathbb{P}\left(\left|\sum_{i=1}^n \varepsilon_i (\sigma_i \cdot X)_t\right| \leq C, \sum_i (\sigma_i \cdot X)_t^2 = \infty, t < \zeta_X\right) \to 0 \quad \text{as } n \to \infty.$$

However, by the definition of good integrators in Definition 2.7, the set $\Big\{\sum_{i=1}^n \varepsilon_i (\sigma_i \cdot X)_t : n \geq 1\Big\} = \Big\{\Big(\Big(\sum_{i=1}^n \varepsilon_i \sigma_i\Big) \cdot X\Big)_t : n \geq 1\Big\}$ of random variables is tight on the event $\{t < \zeta_X\}$ since

$$\left|\sum_{i=1}^n \varepsilon_i \sigma_i\right| \leq \sum_i |\sigma_i| \leq 1 \quad \text{for all } n \geq 1.$$

One readily deduces $\mathbb{P}\Big(\sum_i (\sigma_i \cdot X)_t^2 = \infty, t < \zeta_X\Big) = 0$. □

While the above result will be relevant for the construction of the stochastic integral on $\mathcal{P}_b$, the following lemma plays the same role in the construction of the stochastic integral on $\mathcal{I}_X$.

Lemma 2.10. Let X be an adapted càdlàg process and suppose that $\mathcal{P}_b$ admits a stochastic calculus for X. If $\xi \in \mathcal{I}_X$ and $(\rho_n) \subseteq \mathcal{P}_b$ are such that $\sum_n |\rho_n| \leq |\xi|$, then for all $t \geq 0$, it holds that $(\rho_n \cdot X)_t 1(t < \zeta_X) \to 0$ in probability as $n \to \infty$.

Proof. The proof is almost identical to the one of Lemma 2.9 except that now the tightness of $\left\{\sum_{i=1}^n \varepsilon_i (\rho_i \cdot X)_t : n \geq 1\right\}$ on the event $\{t < \zeta_X\}$ follows from the definition of X-integrability and the fact that for all $n \geq 1$, $\left|\sum_{i=1}^n \varepsilon_i \rho_i\right| \leq \sum_i |\rho_i| \leq |\xi|$. □

The next lemma has the same conclusion as the previous one but under a rather different monotonicity assumption. The main point is that if we had closed intervals in the definition of simple processes (1) then the lemma would be almost trivial since monotone convergence of upper-semicontinuous functions implies uniform convergence. The proof strategy is to modify the processes slightly in such a way as to preserve monotonicity and to ensure upper semicontinuity.

Lemma 2.11. Let X be a good integrator. If $(\sigma_n) \subseteq \Sigma$ is such that $|\sigma_n| \downarrow 0$ as $n \to \infty$, then for all $t \geq 0$, we have $(\sigma_n \cdot X)_t 1(t < \zeta_X) \to 0$ in probability as $n \to \infty$.

Proof. Fix $t \geq 0$, $\delta \in (0,1)$ and $\varepsilon > 0$. We aim to show $\mathbb{P}\left(|(\sigma_n \cdot X)_t| > 2\delta, t < \zeta_X\right) \leq 2\varepsilon$ for sufficiently large n. As we are considering the integral only up to time $t \geq 0$, it suffices to prove the statement when the (σ_n) are of the form

$$\sigma_n = \alpha_0^{(n)} 1_{\{0\}} + \sum_{i=1}^{N_n} \alpha_i^{(n)} 1_{(\tau_i(n), \tau_{i+1}(n)]} \quad \text{with } \alpha_0^{(n)} \in \mathcal{F}_0,\ \alpha_i^{(n)} \in \mathcal{F}_{\tau_i(n)}$$

where $0 = \tau_1(n) \leq \ldots \leq \tau_{N_n+1}(n) = t$ are dissections of $[0,t]$ consisting of stopping times which become finer as n increases. We will inductively construct stopping times $\nu_i(n) \in (\tau_i(n), \tau_{i+1}(n)]$ such that $\mathbb{E}\left(|(\sigma_n \cdot X)_t - (\sigma'_n \cdot X)_t| \wedge 1; t < \zeta_X\right) \leq \delta\varepsilon$ and $|\tilde{\sigma}_n| \downarrow 0$ where

$$\sigma'_n = \alpha_0^{(n)} 1_{\{0\}} + \sum_{i=1}^{N_n} \alpha_i^{(n)} 1_{(\nu_i(n), \tau_{i+1}(n)]}, \quad \tilde{\sigma}_n = \alpha_0^{(n)} 1_{\{0\}} + \sum_{i=1}^{N_n} \alpha_i^{(n)} 1_{[\nu_i(n), \tau_{i+1}(n)]}.$$

To see how the result follows, observe that, since $\tilde{\sigma}_n$ is upper-semicontinuous for each $n \geq 1$ and $|\tilde{\sigma}_n| \downarrow 0$, we deduce (arguing by contradiction) that $\sup_{[0,t]} |\tilde{\sigma}_n| \downarrow 0$. Hence, as clearly $|\sigma'_n| \leq |\tilde{\sigma}_n|$, we obtain $\sup_{[0,t]} |\sigma'_n| \to 0$. Therefore we get

$$\begin{aligned}
&\mathbb{P}\left(|(\sigma_n \cdot X)_t| > 2\delta, t < \zeta_X\right) \\
&\leq \mathbb{P}\left(|(\sigma'_n \cdot X)_t| > \delta, t < \zeta_X\right) + \frac{1}{\delta} \mathbb{E}\left(|(\sigma_n \cdot X)_t - (\sigma'_n \cdot X)_t| \wedge 1; t < \zeta_X\right) \\
&\leq \mathbb{P}\left(\sup_{[0,t]} |\sigma'_n| > \delta'\right) + \sup_{\sigma' \in \Sigma:\ |\sigma'| \leq 1} \mathbb{P}\left(|(\sigma' \cdot X)_t| > \frac{\delta}{\delta'}, t < \zeta_X\right) + \varepsilon.
\end{aligned}$$

Since X is a good integrator, the second term is $\leq \varepsilon/2$ for a sufficiently small $\delta' > 0$ and for fixed $\delta' > 0$, the first term is $\leq \varepsilon/2$ for n sufficiently large, thus completing the argument.

To construct the stopping times $\nu_i(n)$ we first use the fact that X is càdlàg to obtain stopping times $\tilde{\nu}_i(n) > \tau_i(n)$ (e.g. $\tilde{\nu}_i(n) = \tau_i(n) + 1/N$ for N sufficiently large and deterministic) satifying

$$\mathbb{E}\left(\sup_{u,v\in[\tau_i(n),\tilde{\nu}_i(n)]} C|X_u - X_v| \wedge 1; t < \zeta_X\right) \le \delta\varepsilon \cdot 2^{-(1+N_1+\cdots+N_n)}$$
$$\text{where} \quad C = \sup_{n\ge1}\sup_{[0,\infty)}|\sigma_n| = \sup_{[0,\infty)}|\sigma_1| \, .$$

We then set $\nu_i(1) = \tilde{\nu}_i(1) \wedge \tau_{i+1}(1)$. If $n > 1$ and $i \le N_n$, let $j \le N_{n-1}$ be the unique value such that $\big(\tau_i(n), \tau_{i+1}(n)\big] \subseteq \big(\tau_j(n-1), \tau_{j+1}(n-1)\big]$ and define

$$\nu_i(n) = \big(\tilde{\nu}_i(n) \vee \nu_j(n-1)\big) \wedge \tau_{i+1}(n).$$

Then, $|\tilde{\sigma}_n|$ is non-increasing in n and $|\tilde{\sigma}_n| \le |\sigma_n| \downarrow 0$, hence $|\tilde{\sigma}_n| \downarrow 0$. The definition of $(\sigma_n \cdot X)_t$ and $(\sigma'_n \cdot X)_t$ then yields, by inspection, $\mathbb{E}\Big(\big|(\sigma_n \cdot X)_t - (\sigma'_n \cdot X)_t\big| \wedge 1; t < \zeta_X\Big) \le \delta\varepsilon$. □

The last lemma uses the previous one to control supremum distances using the objects introduced in Definition 2.7.

Lemma 2.12. Let X be a good integrator and $\sigma \in \Sigma$. Then $\|\sigma\|^0_{X,t} = \|\sigma\|_{X,t}$ for all $t \ge 0$. Moreover, whenever $t \ge 0$,

$$\mathbb{E}\Big(\big|(\sigma\cdot X)_t\big| \wedge 1; t < \zeta_X\Big) \le \|\sigma\|_{X,t} \quad \text{and}$$
$$\mathbb{E}\Big((\sigma\cdot X)^*_t \wedge 1; t < \zeta_X\Big) \le \big(\|\sigma\|_{X,t} \wedge 1\big)\big(1 - \log\big(\|\sigma\|_{X,t} \wedge 1\big)\big).$$

Proof. By taking $\sigma_1 = \sigma$ and $\sigma_n = 0$ for $n > 1$ in Definition 2.7, it follows that $\|\sigma\|_{X,t} \le \|\sigma\|^0_{X,t}$. It therefore suffices to prove that $\|\sigma\|^0_{X,t} \le \sum_n \|\sigma_n\|^0_{X,t}$ whenever $(\sigma_n) \subseteq \Sigma$ satisfy $|\sigma| \le \sum_n |\sigma_n|$. Let

$$\sigma^{(n)} = \left(\sigma \wedge \sum_{k\le n}|\sigma_k|\right) \vee \left(-\sum_{k\le n}|\sigma_k|\right).$$

Then

$$\begin{aligned}\mathbb{E}\Big(\big|(\sigma\cdot X)_t\big| \wedge 1\Big) &\le \mathbb{E}\Big(\big|(\sigma^{(n)}\cdot X)_t\big| \wedge 1\Big) + \mathbb{E}\Big(\big|\big((\sigma - \sigma^{(n)})\cdot X\big)_t\big| \wedge 1\Big), \quad \text{so}\\ \|\sigma\|^0_{X,t} &\le \big\|\, |\sigma_1| + \cdots + |\sigma_n| \,\big\|^0_{X,t} + \mathbb{E}\Big(\big|\big((\sigma - \sigma^{(n)})\cdot X\big)_t\big| \wedge 1\Big).\end{aligned} \tag{2}$$

By applying Lemma 2.11, we see that the second term tends to 0 as $n \to \infty$ since $|\sigma - \sigma^{(n)}| \downarrow 0$. Moreover, we can decompose each $\sigma' \in \Sigma$ with $|\sigma'| \le |\sigma_1| + \cdots + |\sigma_n|$ as $\sigma' = \sigma'_1 + \cdots + \sigma'_n$ with $\sigma'_i \in \Sigma$ and $|\sigma'_i| \le |\sigma_i|$ for all $1 \le i \le n$. Hence $\big\|\, |\sigma_1| + \cdots + |\sigma_n| \,\big\|^0_{X,t} \le \|\sigma_1\|^0_{X,t} + \cdots + \|\sigma_n\|^0_{X,t}$. Letting $n \to \infty$ in (2) yields the first claim.

The first claimed display is obvious. For the second one, we define $\tau_r = \inf\big\{t \ge 0 : |(\sigma\cdot X)_t| > r\big\}$ whenever $r > 0$ (which are stopping times since $\sigma \cdot X$ is càdlàg). Note that we have

$$\begin{aligned}\mathbb{E}\Big((\sigma\cdot X)^*_t \wedge 1\Big) &= \int_0^1 \mathbb{P}\Big((\sigma\cdot X)^*_t > r\Big)\,dr \le \int_0^1 \left(1 \wedge \frac{\mathbb{E}\Big(\big|(\sigma\cdot X)_{t\wedge\tau_r}\big| \wedge 1\Big)}{r}\right)dr\\ &= \int_0^1 \left(1 \wedge \frac{\mathbb{E}\Big(\big|\big((\sigma 1_{[0,\tau_r]})\cdot X\big)_t\big| \wedge 1\Big)}{r}\right)dr \le \int_0^1 \left(1 \wedge \frac{\|\sigma\|_{X,t}}{r}\right)dr\end{aligned}$$

where we used Lemma 2.5 to obtain the second line and the claim follows readily. □

2.2. Proof of the construction theorem

We can now start putting all the different pieces together. The following two propositions are at the heart of the construction here.

Proposition 2.13. Suppose that X is a good integrator and let $\mathcal{A}_X$ be the set of $\xi \in \mathcal{P}$ with $|\xi| \leq 1$ such that there are $(\sigma_n) \subseteq \Sigma$ with $|\sigma_n| \leq 1$ and $\|\sigma_n - \xi\|_{X,t} \to 0$ as $n \to \infty$ for all $t \geq 0$. If $(\xi_n) \subseteq \mathcal{A}_X$ with $\xi_n \to \xi$ pointwise, then $\xi \in \mathcal{A}_X$ and $\|\xi - \xi_n\|_{X,t} \to 0$ as $n \to \infty$ for all $t \geq 0$.

Proof. Throughout the proof, we fix $t \geq 0$. We set $\eta = 1$ below so the proof can almost verbatim be copied to give a proof of Proposition 2.14.

Step 1: We first show that if $\xi_n \downarrow 0$ then $\|\xi_n\|_{X,t} \to 0$ as $n \to \infty$. The terms $\|\xi_n\|_{X,t}$ are non-increasing in n, so it suffices to show $\|\xi_n\|_{X,t} \to 0$ as $n \to \infty$ along a subsequence.

We first show that $\|\xi_n - \xi_m\|_{X,t} \to 0$ as $m, n \to \infty$. To this end, fix $\varepsilon > 0$ and take simple $0 \leq \sigma_n \leq |\eta|$ such that $\|\xi_n - \sigma_n\| \leq \varepsilon 2^{-n}$ for all $n \geq 1$. Now define $\tilde{\sigma}_n = \sigma_1 \wedge \cdots \wedge \sigma_n \in \Sigma$. Then of course $\xi_n - \sigma_n \leq \xi_n - \tilde{\sigma}_n$ and

$$\begin{aligned}\xi_n - \tilde{\sigma}_n &= (\xi_n - \sigma_1) \vee \cdots \vee (\xi_n - \sigma_n) \leq (\xi_1 - \sigma_1) \vee \cdots \vee (\xi_n - \sigma_n) \\ &\leq |\xi_1 - \sigma_1| + \cdots + |\xi_n - \sigma_n|\end{aligned}$$

Hence $|\xi_n - \tilde{\sigma}_n| \leq |\xi_1 - \sigma_1| + \cdots + |\xi_n - \sigma_n|$ and so by the triangle inequality, $\|\xi_n - \tilde{\sigma}_n\|_{X,t} \leq \|\xi_1 - \sigma_1\|_{X,t} + \cdots + \|\xi_n - \sigma_n\|_{X,t} \leq \varepsilon$. Thus

$$\begin{aligned}\|\xi_n - \xi_m\|_{X,t} &\leq \|\xi_n - \tilde{\sigma}_n\|_{X,t} + \|\xi_m - \tilde{\sigma}_m\|_{X,t} + \|\tilde{\sigma}_n - \tilde{\sigma}_m\|_{X,t} \\ &\leq 2\varepsilon + \|\tilde{\sigma}_n - \tilde{\sigma}_m\|_{X,t} .\end{aligned}$$

The first claim will follow once we show that $\|\tilde{\sigma}_n - \tilde{\sigma}_m\|_{X,t} \to 0$ as $n, m \to \infty$. To see this, it is enough to show that for any subsequence (n_k), we have $\|\tilde{\sigma}_{n_{k+1}} - \tilde{\sigma}_{n_k}\|^0_{X,t} = \|\tilde{\sigma}_{n_{k+1}} - \tilde{\sigma}_{n_k}\|_{X,t} \to 0$ as $k \to \infty$. By Definition 2.7, we can take simple $|\sigma'_k| \leq |\tilde{\sigma}_{n_k} - \tilde{\sigma}_{n_{k+1}}|$ such that

$$\|\tilde{\sigma}_{n_k} - \tilde{\sigma}_{n_{k+1}}\|^0_{X,t} \leq \mathbb{E}\Big(|(\sigma'_k \cdot X)_t| \wedge 1; t < \zeta_X\Big) + \frac{1}{k}.$$

Since $\sum_k |\sigma'_k| \leq \sum_k \big(\tilde{\sigma}_{n_k} - \tilde{\sigma}_{n_{k+1}}\big) \leq |\eta|$, by Lemma 2.9, $(\sigma'_k \cdot X)_t \to 0$ in probability as $k \to \infty$, so $\|\tilde{\sigma}_{n_k} - \tilde{\sigma}_{n_{k+1}}\|^0_{X,t} \to 0$ as $k \to \infty$ as required.

So we have established that $\|\xi_n - \xi_m\|_{X,t} \to 0$ as $m, n \to \infty$. We can thus choose a subsequence (m_k) such that $\|\xi_{m_{k+1}} - \xi_{m_k}\|_{X,t} \leq 2^{-k}$ for all $k \geq 1$. Since $\xi_{m_l} = \sum_{k \geq l} \big(\xi_{m_k} - \xi_{m_{k+1}}\big)$,

$$\|\xi_{m_l}\|_{X,t} \leq \sum_{k \geq l} \|\xi_{m_{k+1}} - \xi_{m_k}\|_{X,t} \leq 2 \cdot 2^{-l} \to 0 \quad \text{as } l \to \infty.$$

Step 2: Having established the first step, we can now prove the proposition. It is of course enough to show $\|\xi - \xi_n\|_{X,t} \to 0$ as $n \to \infty$ since $\xi \in \mathcal{A}_X$ follows from an easy diagonal argument. The proof strategy now uses ideas inspired by Egorov's theorem.

Fix $\varepsilon > 0$ and let

$$\eta^{(n,k)}_t = |\eta_t| \; 1\Big(\exists m \geq n : |\xi_t - (\xi_m)_t| \geq 1/k\Big).$$

Then $\eta^{(n,k)}$ is predictable, $0 \leq \eta^{(n,k)} \leq |\eta|$ and $\eta^{(n,k)} \downarrow 0$ as $n \to \infty$. Therefore by the first part, we can take (n_k) such that $\|\eta^{n_k,k}\|_{X,t} \leq \varepsilon 2^{-(k+1)}$ for all $k \geq 1$. Hence

$$\|\tilde{\eta}\|_{X,t} \leq \sum_{k\geq 1} \|\eta^{(n_k,k)}\|_{X,t} \leq \frac{\varepsilon}{2}$$

where $\tilde{\eta}_t := |\eta_t|\ 1\big(\exists k, m \geq n_k : |\xi_t - (\xi_m)_t| \geq 1/k\big) \leq \sum_{k\geq 1} \eta_t^{(n_k,k)}$. Hence for $n \geq n_k$

$$\begin{aligned}
\|\xi - \xi_n\|_{X,t} &\leq \|(\xi - \xi_n)1\big(\exists k, m \geq n_k : |\xi_t - (\xi_m)_t| \geq 1/k\big)\|_{X,t} \\
&\quad + \|(\xi - \xi_n)1\big(\forall k, m \geq n_k : |\xi_t - (\xi_m)_t| \leq 1/k\big)\|_{X,t} \\
&\leq 2\ \|\tilde{\eta}\|_{X,t} + \frac{1}{k}\ \|1\|_{X,t} \\
&\leq \varepsilon + \frac{1}{k}\ \|1\|_{X,t}^0 \leq \varepsilon + \frac{1}{k}.
\end{aligned}$$

So if $k \geq 1/\varepsilon$ then $\|\xi - \xi_n\|_{X,t} \leq 2\varepsilon$ for all $n \geq n_k$ and the proof is complete. □

Proposition 2.14. Let X be an adapted càdlàg process and suppose that $\mathcal{P}_b$ admits a stochastic calculus for X. Fix $\eta \in \mathcal{I}_X$ and let $\mathcal{A}_X^\eta$ be the set of all $\xi \in \mathcal{P}$ with $|\xi| \leq |\eta|$ such that there are $(\rho_n) \subseteq \mathcal{P}_b$ with $|\rho_n| \leq |\eta|$ and $\|\rho_n - \xi\|_{X,t} \to 0$ as $n \to \infty$ for all $t \geq 0$. If $(\xi_n) \subseteq \mathcal{A}_X^\eta$ with $\xi_n \to \xi$ pointwise, then $\xi \in \mathcal{A}_X^\eta$ and $\|\xi - \xi_n\|_{X,t} \to 0$ as $n \to \infty$ for all $t \geq 0$.

Proof. The proof is identical to the one of Proposition 2.13 except that we make use of Lemma 2.10 instead of Lemma 2.9. □

Lemma 2.15. Let X be a good integrator, ξ is predictable and $|\xi| \leq 1$. Then $\xi \in \mathcal{A}_X$.

Proof. Let us begin with the following simple additivity observation: If $\xi_1, ..., \xi_m \in \mathcal{A}_X$ and $\lambda_1, ..., \lambda_m \in \mathbb{R}$ such that $|\lambda_1\xi + \dots + \lambda_m\xi_m| \leq 1$ then $\lambda_1\xi + \dots + \lambda_m\xi_m \in \mathcal{A}_X$. Indeed, we can take simple processes $\left(\sigma_i^{(n)}\right)$ such that $|\sigma_i^{(n)}| \leq 1$ and $\|\sigma_i^{(n)} - \xi_i\|_{X,t} \to 0$ as $n \to \infty$ for all $t \geq 0$ and all $1 \leq i \leq m$. Then

$$\begin{aligned}
\left\| \sum_{i=1}^m \lambda_i \xi_i - \left(\left(\sum_{i=1}^m \lambda_i \sigma_i^{(n)} \right) \vee (-1) \right) \wedge 1 \right\|_{X,t} &\leq \left\| \sum_{i=1}^m \lambda_i \xi_i - \left(\sum_{i=1}^m \lambda_i \sigma_i^{(n)} \right) \right\|_{X,t} \\
&\leq \sum_{i=1}^m |\lambda_i| \cdot \|\xi_i - \sigma_i^{(n)}\|_{X,t} \to 0 \quad \text{as } n \to \infty
\end{aligned}$$

where the first inequality follows from

$$\left| \sum_{i=1}^m \lambda_i \xi_i - \left(\left(\sum_{i=1}^m \lambda_i \sigma_i^{(n)} \right) \vee (-1) \right) \wedge 1 \right| \leq \left| \sum_{i=1}^m \lambda_i \xi_i - \left(\sum_{i=1}^m \lambda_i \sigma_i^{(n)} \right) \right|.$$

The claim follows since $\left(\left(\sum_{i=1}^m \lambda_i \sigma_i^{(n)}\right) \vee (-1)\right) \wedge 1$ is simple and bounded by 1. We can now begin with the proof.

Step 1: Let $\mathcal{D}$ be the set of $A \in \mathcal{P}$ with $1_A \in \mathcal{A}_X$. We first show that $\mathcal{P} = \mathcal{D}$. Clearly $\tilde{\Sigma} \subseteq \mathcal{D}$ and $[0, \infty) \times \Omega \in \mathcal{D}$ so by Dynkin's lemma, it suffices to show that $\mathcal{D}$ is a Dynkin system. If $A \subseteq B$ and $A, B \in \mathcal{D}$, then $1_{B\setminus A} = 1_B - 1_A \in \mathcal{A}_X$ by the additivity observation made above. Now suppose that $(A_n) \subseteq \mathcal{D}$ with $A_n \uparrow A$, then $1_{A_n} \in \mathcal{A}_X$ and by Proposition 2.13 $1_A \in \mathcal{A}_X$ and therefore $A \in \mathcal{D}$. We thus conclude that $\mathcal{D}$ is indeed a Dynkin system.

Step 2: The process $2^{-n}\lfloor 2^n\xi\rfloor$ is a linear combination of indicators on elements of $\mathcal{P}$ and is bounded by 1, and so by the additivity observation made above, $2^{-n}\lfloor 2^n\xi\rfloor \in \mathcal{A}_X$. Sine $2^{-n}\lfloor 2^n\xi\rfloor \to \xi$ pointwise as $n \to \infty$, the claim follows from Proposition 2.13. □

We are finally able to prove the main theorem which constructs the stochastic integral. The proof will essentially just consist in applying all the results established up to now together.

Proof of Theorem 3.4. Combining Proposition 2.13 and Lemma 2.15 (together with a small scaling argument) yields the following:

- For all $\xi \in \mathcal{P}_b$ there exist simple (σ_n) such that $\|\xi - \sigma_n\|_{X,t} \to 0$ as $n \to \infty$ for all $t \geq 0$.
- If $(\xi_n) \subseteq \mathcal{P}_b$, $\xi \in \mathcal{P}_b$, $\xi_n \to \xi$ pointwise and $|\xi_n| \leq C$ for some deterministic constant $C < \infty$, then $\|\xi - \xi_n\|_{X,t} \to 0$ as $n \to \infty$ for all $t \geq 0$.

For any $\xi \in \mathcal{P}_b$ take (σ_n) as above and observe that by Lemma 2.12

$$\mathbb{E}\Big((\sigma_n \cdot X - \sigma_m \cdot X)^*_t \wedge 1; t < \zeta_X\Big) \leq \big(\|\sigma_n - \sigma_m\|_{X,t} \wedge 1\big)\big(1 - \log\big(\|\sigma_n - \sigma_m\|_{X,t} \wedge 1\big)\big) \to 0$$

as $m, n \to \infty$ for all $t \geq 0$. So $(\sigma_n \cdot X)$ is Cauchy for u.c.p. convergence and it follows that there exists a process which we call $\xi \cdot X$ such that $\sigma_n \cdot X \to \xi \cdot X$ u.c.p. as $n \to \infty$. By a similar reasoning, we see that this limit does not depend on the sequence (σ_n) chosen in the sense that the two limits will a.s. be equal. By the construction, it in particular follows that

$$\mathbb{E}\Big((\xi \cdot X)^*_t \wedge 1; t < \zeta_X\Big) \leq \big(\|\xi\|_{X,t} \wedge 1\big)\big(1 - \log\big(\|\xi\|_{X,t} \wedge 1\big)\big). \tag{3}$$

for all $\xi \in \mathcal{P}_b$ and all $t \geq 0$. Let us now check that the processes $\xi \cdot X$ for $\xi \in \mathcal{P}_b$ satisfy the properties in Definition 2.6. Indeed, (i) is clear, (ii) follows immediately from (3) and the second bullet point above, and (iii) extends directly from the simple case.

To see the uniqueness part, let $\mathcal{D} = \{A \in \mathcal{P} : 1_A \cdot X = 1_A * X \text{ a.s}\}$. Then by (i), $\tilde{\Sigma} \subseteq \mathcal{D}$ and $[0,\infty) \times \Omega \in \mathcal{D}$. Properties (ii) and (iii) easily imply that $\mathcal{D}$ is a Dynkin system and so $\tilde{\Sigma} = \mathcal{P}$. Approximating any $\xi \in \mathcal{P}_b$ by $2^{-n}\lfloor 2^n\xi\rfloor$ which are linear combinations of indicators of elements in $\mathcal{P}$ and again using (ii) combined with (iii) yields the uniqueness.

The second part of the theorem, namely that $\mathcal{I}_X$ admits a stochastic calculus for X if X is a good integrator, is analogous and just uses Proposition 2.14 instead of Proposition 2.13 and (3) instead of Lemma 2.12. □

Remark 2.16. One thing that is used in practice (and easy to check by definition) is the following: If X is a good integrator and Y is an adapted càdlàg process then $1_{[0,\zeta_Y)}Y_- \in \mathcal{I}_X$ and when we write $Y_- \cdot X$ for the process $\left(1_{[0,\zeta_Y)}Y_-\right) \cdot X$, except that we redefine it to be equal to ∂ on the interval $[\zeta_X \wedge \zeta_Y, \infty)$. By slight abuse of terminology, we also say that Y is X-integrable.

The reader can convince oneself of the fact that $Y_- \cdot X$ is a good integrator and moreover that if Z is an adapted càdlàg process then

$$Z_- \cdot (Y_- \cdot X) = (ZY)_- \cdot X \quad \text{a.s.}$$

One can of course also formulate and prove this principle for more general integrands ξ instead of Z_- (where one now needs an integrability assumption).

3. Results on stochastic integrals

The main goal of this section will be to establish Itô's formula. As a first step, we define the quadratic variation and covariation processes which measure the amount of quadratic fluctuations a process accumulates over time.

Definition 3.1. Suppose that X and Y are good integrators, then we define the càdlàg process $[X, Y]$ with $\zeta_{[X,Y]} = \zeta_X \wedge \zeta_Y$, called the *quadratic covariation process* by

$$[X, Y] = XY - X_0 Y_0 - X_- \cdot Y - Y_- \cdot X.$$

We will also write $[X] = [X, X]$ which is called the *quadratic variation process*. We have the polarization identity $[X, Y] = ([X + Y] - [X - Y])/4$ a.s.

As the lemma below shows, $[X]$ has a non-decreasing version and we will always switch to such a version. By the polarization identity $[X, Y]$ therefore always has a finite variation version and we agree to always switch to such a version in this case.

Since $[X, Y]$ is then of finite variation, it is a good integrator with the stochastic integral clearly agreeing with the Lebesgue-Stieltjes integral.

Whenever X is càdlàg process, we may define $\Delta X_t = X_t - X_{t-}$. Then ΔX_t is non-zero for only countably many $t < \zeta_X$. For many technical purposes, it is best to think of the process ΔX as being identified with the point process

$$\sum_{t : \Delta X_t \neq 0} \delta_{(t, \Delta X_t)}.$$

By first verifying the following for simple processes and extending the result using u.c.p. convergence, it is a straightforward to check, that $\Delta(\xi \cdot X)_t = \xi_t \Delta X_t$ for all $t < \zeta_X$ a.s. whenever X is a good integrator and $\xi \in \mathcal{I}_X$. In particular, if X is continuous a.s. then so is $\xi \cdot X$. Moreover, by definition $\Delta[X, Y]_t = \Delta X_t \Delta Y_t$ for all $t < \zeta_X \wedge \zeta_Y$ a.s.

Lemma 3.2. Suppose that X and Y are good integrators, and for each n let $(\tau_m^n)_m$ be a non-decreasing sequence of stopping times such that $\tau_0^n = 0$, $\tau_m^n \to \infty$ as $m \to \infty$. Also assume that

$$\max_{m \geq 0} |(\tau_{m+1}^n \wedge t) - (\tau_m^n \wedge t)| \to 0 \quad \text{as } n \to \infty \text{ for all } t \geq 0.$$

Let A be an adapted càdlàg process, then

$$A_X^{(n)} := \sum_{m \geq 0} A_{\tau_m^n} \left(X_{\cdot \wedge \tau_{m+1}^n} - X_{\cdot \wedge \tau_m^n} \right) \to A_- \cdot X \quad \text{u.c.p. as } n \to \infty.$$

Moreover, $[X]$ has a version which is non-decreasing process on $[0, \zeta_X)$ and $[X, Y]$ has a finite variation version on $[0, \zeta_X)$ (to which we switch). Finally,

$$A_{X,Y}^{(n)} := \sum_{m \geq 0} A_{\tau_m^n} \left(X_{\cdot \wedge \tau_{m+1}^n} - X_{\cdot \wedge \tau_m^n} \right) \left(Y_{\cdot \wedge \tau_{m+1}^n} - Y_{\cdot \wedge \tau_m^n} \right)$$
$$\to A_- \cdot [X, Y] \quad \text{u.c.p. as } n \to \infty.$$

Proof. To get the first convergence statement, note that $A_X^{(n)}$ is still adapted and has the càdlàg property since on any compact time interval, only finitely terms contribute to the sum. Now observe that $A_X^{(n)} = \xi^{(n)} \cdot X$ on $\left[0, \zeta_{A_- \cdot X}\right) = [0, \zeta_A \wedge \zeta_X)$ a.s. where

$$\xi^{(n)} = \sum_{m \geq 0} A_{\tau_m^n} 1\left(A_{\tau_m^n} \neq \partial\right) 1_{(\tau_m^n, \tau_{m+1}^n]}.$$

We have $|\xi_n| \leq X_-^* 1_{[0,\zeta_X)}$ for all $n \geq 1$ and the right hand side is X-integrable since X^* is adapted and càdlàg. Since $\xi_n \to A_- 1_{[0,\zeta_A)}$ pointwise as $n \to \infty$, the first convergence claim follows from stochastic dominated convergence.

To see that $[X]$ is non-decreasing a.s. we observe that

$$\begin{aligned}\sum_{m \geq 0} \left(X_{\cdot \wedge \tau_{m+1}^n} - X_{\cdot \wedge \tau_m^n}\right)^2 &= X^2 - X_0^2 - 2 \sum_{m \geq 0} X_{\tau_m^n} \left(X_{\cdot \wedge \tau_{m+1}^n} - X_{\cdot \wedge \tau_m^n}\right) \\ &\to X^2 - X_0^2 - 2 X_- \cdot X\end{aligned}$$

u.c.p. as $n \to \infty$ by the first part, and the claim follows from the fact that the right hand side equals $[X]$ a.s. and it is easy to see that the limit of the terms on the left is non-decreasing a.s.

Now to see the last claim, we observe that

$$\begin{aligned}A_{X,Y}^{(n)} = &\sum_{m \geq 0} A_{\tau_m^n} \left((XY)_{\cdot \wedge \tau_{m+1}^n} - (XY)_{\cdot \wedge \tau_m^n}\right) - \sum_{m \geq 0} (AX)_{\tau_m^n} \left(Y_{\cdot \wedge \tau_{m+1}^n} - Y_{\cdot \wedge \tau_m^n}\right) \\ &- \sum_{m \geq 0} (AY)_{\tau_m^n} \left(X_{\cdot \wedge \tau_{m+1}^n} - X_{\cdot \wedge \tau_m^n}\right) \\ \to\ &A_- \cdot (XY) - (AX)_- \cdot Y - (AY)_- \cdot X\end{aligned}$$

u.c.p. as $n \to \infty$ by applying the first part of the lemma three times. Note that XY is a good integrator since $XY = X_0 Y_0 + X_- \cdot Y + Y_- \cdot X$ a.s., and in fact using this equation at the end of the display above yields the claim by Remark 2.16. ☐

Remark 3.3. Note that if X is a good integrator, then

$$\sum_{s \leq t} (\Delta X_s)^2 = \sum_{s \leq t} \Delta [X]_s \leq [X]_t < \infty \quad \text{for all } t < \zeta_X$$

since $[X]$ is non-decreasing.

We are now ready to prove Itô's formula, essentially the fundamental theorem of stochastic calculus. It is worth mentioning that this is not the most general version possible, for instance, one can generalize the result by replacing f by a random twice continuously differentiable function $f : \mathbb{R} \times D \to \mathbb{R}$ with the property that $f(t, \cdot)$ is measurable with respect to $\mathcal{F}_t$ for all $t \geq 0$ (i.e. an adaptedness property holds for this function) and consider $f(t, X_t)$ instead. The variant just mentioned is for instance relevant in the context of Schramm-Loewner evolutions.

Theorem 3.4 (Itô's formula). Suppose that $X^1, ..., X^d$ are good integrators such that $X = (X^1, ..., X^d)$ is such that $X_t, X_{t-} \in D \subseteq \mathbb{R}^d$ for all $t < \zeta_X = \zeta_{X^1} \wedge ... \wedge \zeta_{X^d}$, D some open set and $f : D \to \mathbb{R}$ is twice continuously differentiable, then

$$f(X_T) = f(X_0) + \sum_i \int_0^T \partial_i f(X_{t-}) dX_t + \frac{1}{2} \sum_{i,j} \int_0^T \partial_{ij} f(X_{t-}) d[X^i, X^j]_t$$
$$+ \sum_{t \leq T} \left(\Delta f(X_t) - \sum_i \partial_i f(X_{t-}) \Delta X_t^i - \frac{1}{2} \sum_{i,j} \partial_{ij} f(X_{t-}) \Delta X_t^i \Delta X_t^j \right)$$

for all $T \geq 0$ almost surely (by the conventions we made, both sides equal ∂ if $T \geq \zeta_X$). Note that the sum over times $t \leq T$ above converges absolutely and defines an adapted càdlàg process since $\sum_{t \leq T} \left(\Delta X_t^i\right)^2 < \infty$ for all $i \leq d$ and $T < \zeta_X$.

Proof. Since both sides of the formula are càdlàg, it suffices that almost surely the claim holds for a fixed $T \geq 0$. Let (τ_m^n) be any sequence satisfying the requirements of Lemma 3.2 and fix $T < \zeta_X$. For $0 \leq s < t \leq T$ let $R_{s,t}$ be defined via

$$f(X_t) = f(X_s) + \sum_i \partial_i f(X_s)(X_t^i - X_s^i) + \frac{1}{2} \sum_{i,j} \partial_{ij} f(X_s)(X_t^i - X_s^i)\left(X_t^j - X_s^j\right) + R_{s,t}.$$

Since $X([0,T]) \cup X_-([0,T])$ is compact and f is twice continuously differentiable, we get

$$C_\varepsilon := \sup_{0 \leq s < t \leq T : 0 < \|X_s - X_t\| \leq \varepsilon} \frac{|R_{s,t}|}{\|X_t - X_s\|^2} \to 0 \quad \text{as } \varepsilon \to 0.$$

By using the definition of $R_{s,t}$ and a telescoping series, we obtain

$$\begin{aligned}
f(X_T) - f(X_0) &= \sum_{m \geq 0} \left(f\left(X_{T \wedge \tau_{m+1}^n}\right) - f\left(X_{T \wedge \tau_m^n}\right) \right) \\
&= \sum_i \sum_{m \geq 0} \partial_i f\left(X_{\tau_m^n}\right) \left(X_{T \wedge \tau_{m+1}^n}^i - X_{T \wedge \tau_m^n}^i\right) \\
&\quad + \frac{1}{2} \sum_{i,j} \sum_{m \geq 0} \partial_{ij} f\left(X_{\tau_m^n}\right) \left(X_{T \wedge \tau_{m+1}^n}^i - X_{T \wedge \tau_m^n}^i\right) \left(X_{T \wedge \tau_{m+1}^n}^j - X_{T \wedge \tau_m^n}^j\right) \\
&\quad + \sum_{m \geq 0} R_{T \wedge \tau_m^n, T \wedge \tau_{m+1}^n}.
\end{aligned}$$

By Lemma 3.2 it therefore suffices to show that

$$\sum_{m \geq 0} R_{T \wedge \tau_m^n, T \wedge \tau_{m+1}^n} \to \sum_{t \leq T} J_t \quad \text{in probability as } n \to \infty \text{ on the event } \{T < \zeta_X\},$$
$$\text{where} \quad J_t = \Delta f(X_t) - \sum_i \partial_i f(X_{t-}) \Delta X_t^i - \frac{1}{2} \sum_{i,j} \partial_{ij} f(X_{t-}) \Delta X_t^i \Delta X_t^j.$$

To see this, fix $\varepsilon > 0$ and define the finite set $\mathcal{T}_\varepsilon = \{t \leq T : \|\Delta X_t\| \geq \varepsilon\}$. Moreover, we let $I_\varepsilon^n = \{m \geq 0 : \mathcal{T}_\varepsilon \cap [T \wedge \tau_m^n, T \wedge \tau_{m+1}^n) \neq \emptyset\}$. Then we have the deterministic statement

$$\sum_{m \in I_\varepsilon^n} R_{T \wedge \tau_m^n, T \wedge \tau_{m+1}^n} \to \sum_{t \leq T:\, \|\Delta X_t\| \geq \varepsilon} J_t \quad \text{as } n \to \infty \text{ on } \{T < \zeta_X\}.$$

For n sufficiently large, we have $\|X_{T \wedge \tau_{m+1}^n} - X_{T \wedge \tau_{m+1}^n}\| \leq 2\varepsilon$ for all $m \notin I_\varepsilon^n$ and hence

$$\sum_{m\in I_\varepsilon^n} |R_{T\wedge\tau_m^n, T\wedge\tau_{m+1}^n}| \le C_{2\varepsilon} \sum_{m\ge 0} \|X_{T\wedge\tau_{m+1}^n} - X_{T\wedge\tau_{m+1}^n}\|^2 \to C_{2\varepsilon} \sum_i [X^i]_T$$

in probability as $n \to \infty$ on the event $\{T < \zeta_X\}$.

The claim now follows directly from $C_{2\varepsilon} \to 0$ as $\varepsilon \to 0$ and the absolute convergence of the series $\sum_{t\le T} J_t$ when $T < \zeta_X$. □

To conclude this section, we mention the following result for time changes.

Theorem 3.5 (Time changes). Suppose that X is a good integrator and that τ_t are stopping times for all $t \ge 0$ such that τ is a non-decreasing and right-continuous map, and define the random time $\zeta_{X\circ\tau} = \inf\{t \ge 0 : \tau_t \ge \zeta_X\}$. Then the process $X \circ \tau$ is a good integrator with respect to the filtration $\left(\mathcal{F}_{\tau_t}\right)_{t\ge 0}$ and for any $\xi \in \mathcal{I}_X$ we have $\xi \circ \tau \in \mathcal{I}_{X\circ\tau}$ with

$$(\xi \circ \tau) \cdot (X \circ \tau) = (\xi \cdot X) \circ \tau \quad \text{a.s.}$$

Proof. This is immediate whenever ξ (and therefore $\xi \circ \tau$) is a simple process and extends easily to the general case. □

4. Results on integrability

In this section, we will mention the setting in which the result from the remainder of the note are supposed to be applied. The main point is that martingales turn out to be good integrators precisely because their unbiased fluctuations induce the stochastic cancellations needed to make the definition of being a good integrator true.

We say that an adapted càdlàg process M is a local martingale (resp. locally square integrable martingale) if there are stopping times $\tau_n < \zeta_M$ such that $\tau_n \to \zeta_M$ a.s. as $n \to \infty$ and such that M^{τ_n} is a martingale (resp. square integrable martingale) for all $n \ge 1$. An adapted càdlàg process C with values in $[0, \infty] \cup \{\partial\}$ is called locally integrable if there exist stopping times $\tau_n < \zeta_C$ satisfying $\tau_n \to \zeta_C$ a.s. as $n \to \infty$ and C^{τ_n} is an integrable process for all $n \ge 1$. Moreover, we say that A is a finite variation process if A has finite variation on $[0, T]$ for all $T < \zeta_A$.

Theorem 4.1. Let $X = M + A$ where M is a locally square integrable càdlàg martingale and A is a càdlàg adapted finite variation process. Then X is a good integrator. Suppose that $\xi \in \mathcal{P}$, that $\xi^2 \cdot [M]$ is a locally integrable process and

$$\int_0^t |\xi_s|\, |dA_s| < \infty \quad \text{for all } t < \zeta_X \text{ a.s.}$$

Then $\xi \in \mathcal{I}_X$. Also $\xi \cdot X = \xi \cdot M + \xi \cdot A$ a.s. and $\xi \cdot M$ is a locally square integrable martingale while $\xi \cdot A$ is a finite variation process a.s.

Proof. Without loss of generality, we may assume that $\zeta_X = \zeta_M = \zeta_A$. From Lebesgue-Stieltjes integration one deduces that A is a good integrator and $\xi \in \mathcal{I}_X$ by the assumption on the integral of ξ with respect to the total variation process of A. So it remains to consider M as an integrator.

Before we begin, let us make the observation that if (N^n) are càdlàg martingales (with $\zeta_{N^n} = \infty$ for all $n \ge 1$) such that

$$\sup_{n\ge 1} \mathbb{E}\left((N_t^n)^2\right) \quad \text{for all } t \ge 0$$

and N is an adapted càdlàg process (with $\zeta_N = \infty$) such that $N_n \to N$ u.c.p., then N is a square integrable martingale. The martingale property follows from the fact that convergence in probability together with L^2 boundedness (in particular uniform integrability) implies L^1 convergence, and the square integrability from Fatou's lemma.

Let (τ_n) be a sequence of stopping times such that $\tau_n < \zeta_M$, $\tau_n \to \zeta_M$ a.s. as $n \to \infty$, M^{τ_n} is a square integrable martingale for all $n \geq 1$, $|M_-| \leq n$ on $[0, \tau_n]$ and

$$\mathbb{E}\Big(\big(\xi^2 \cdot M^{\tau_n}\big)_t\Big) < \infty \quad \text{for all } t \geq 0 \text{ and } n \geq 1.$$

Then for a simple bounded process σ we have

$$\begin{aligned}\mathbb{P}\Big(\big|(\sigma \cdot M)_t\big| > C, t < \zeta_M\Big) &\leq \mathbb{P}\Big(\big|(\sigma \cdot M^{\tau_n})_t\big| > C\Big) + \mathbb{P}(\tau_n < t < \zeta_M) \\ &\leq \frac{1}{C^2}\,\mathbb{E}\Big((\sigma \cdot M^{\tau_n})_t^2\Big) + \mathbb{P}(\tau_n < t < \zeta_M).\end{aligned} \tag{4}$$

Let σ be of the form (1) and without loss of generality assume that $\tilde{\tau}_i \leq \tau_{i+1}$ for all $i < n$ The key is then that by the martingale property and the definition of simple integrands,

$$\begin{aligned}\mathbb{E}\Big((\sigma \cdot M^{\tau_n})_t^2\Big) &= \sum_{i=1}^{n} \mathbb{E}\Big(\alpha_i^2 \big(M^{\tau_n}_{\tilde{\tau}_i \wedge t} - M^{\tau_n}_{\tau_i \wedge t}\big)^2\Big) \\ &\quad + \sum_{1 \leq i < j \leq n} \mathbb{E}\Big(\alpha_i \alpha_j \big(M^{\tau_n}_{\tilde{\tau}_i \wedge t} - M^{\tau_n}_{\tau_i \wedge t}\big)\Big(M^{\tau_n}_{\tilde{\tau}_j \wedge t} - M^{\tau_n}_{\tau_j \wedge t}\Big)\Big) \\ &= \sum_{i=1}^{n} \mathbb{E}\Big(\alpha_i^2 \big(M^{\tau_n}_{\tilde{\tau}_i \wedge t} - M^{\tau_n}_{\tau_i \wedge t}\big)^2\Big) \\ &\quad + \sum_{1 \leq i < j \leq n} \mathbb{E}\Big(\alpha_i \alpha_j \big(M^{\tau_n}_{\tilde{\tau}_i \wedge t} - M^{\tau_n}_{\tau_i \wedge t}\big)\mathbb{E}\Big(M^{\tau_n}_{\tilde{\tau}_j \wedge t} - M^{\tau_n}_{\tau_j \wedge t} \mid \mathcal{F}_{\tau_j \wedge t}\Big)\Big) \\ &= \sum_{i=1}^{n} \mathbb{E}\Big(\alpha_i^2 \big(M^{\tau_n}_{\tilde{\tau}_i \wedge t} - M^{\tau_n}_{\tau_i \wedge t}\big)^2\Big).\end{aligned}$$

This is really the key insight of Itô and known as the Itô isometry. If $|\sigma| \leq 1$ then

$$\begin{aligned}\mathbb{E}\Big((\sigma \cdot M^{\tau_n})_t^2\Big) &\leq \sum_{i=1}^{n} \mathbb{E}\Big(\big(M^{\tau_n}_{\tilde{\tau}_i \wedge t} - M^{\tau_n}_{\tau_i \wedge t}\big)^2\Big) = \sum_{i=1}^{n} \mathbb{E}\Big(\big(M^{\tau_n}_{\tilde{\tau}_i \wedge t}\big)^2 - \big(M^{\tau_n}_{\tau_i \wedge t}\big)^2\Big) \\ &\leq \mathbb{E}\Big(\big(M^{\tau_n}_t\big)^2\Big) < \infty.\end{aligned}$$

One readily deduces that M is a good integrator, and one easily checks that $\sigma \cdot M^{\tau_n}$ is a square integrable martingale and so $\sigma \cdot M$ is a locally square integrable martingale. Thus the stochastic integral extends to bounded predictable processes and $\rho \cdot M$ is a locally square integrable martingale for all $\rho \in \mathcal{P}_b$ by the observation made at the beginning of the proof.

In particular, $(M_- \cdot M)^{\tau_n} = (M^{\tau_n})_- \cdot M^{\tau_n}$ a.s. is a square integrable martingale and from the definition of $[M]$ and the square martingale property of M^{τ_n} it follows that

$$\mathbb{E}\Big((\sigma \cdot M^{\tau_n})_t^2\Big) = \sum_{i=1}^{n} \mathbb{E}\Big(\alpha_i^2 \Big([M^{\tau_n}]_{\tilde{\tau}_i \wedge t} - [M^{\tau_n}]_{\tau_i \wedge t}\Big)\Big) = \mathbb{E}\Big(\big(\sigma^2 \cdot [M^{\tau_n}]\big)_t\Big).$$

By approximating a bounded predictable process pointwise by by simple processes all bounded by some common constant (it is a standard argument in measure theory that this is possible) and using Fatou's lemma on the left and dominated convergence on the right, we get

$$\mathbb{E}\Big((\rho \cdot M^{\tau_n})_t^2\Big) \le \mathbb{E}\Big((\rho^2 \cdot [M^{\tau_n}])_t\Big)$$

whenever $\rho \in \mathcal{P}_b$. Showing that $\xi \in \mathcal{I}_M$ is then straightforward by using a variant of (4) combined with the previous display. The local square martingale property of $\xi \cdot M$ follows again by the observation made at the beginning of the proof after taking a sequence of bounded predictable processes converging pointwise to ξ and all bounded by $|\xi|$. ☐

Remark 4.2. The above can be generalized. Indeed, one can show that all local martingales are good integrators and that it is enough to assume that $(\xi^2 \cdot [M])^{1/2}$ is locally integrable to ensure that $\xi \in \mathcal{I}_X$ (to see this one needs to invoke the BDG inequality). However, in this generalized setting, it is no longer true that $\xi \cdot M$ is a local martingale which makes this generalization less useful for applications; a counterexample is *Emery's example* as mentioned in [1].

Let us conclude these notes by mentioning perhaps the most important case: Let us fix a diffusivity parameter $\sigma \ge 0$, drift $\mu \in \mathbb{R}$ and an intensity measure μ on $\mathbb{R}$ satisfying

$$\int_{[-1,1]} x^2 \mu(dx) < \infty \quad \text{and} \quad \mu(\mathbb{R} \setminus [-1,1]) < \infty.$$

As always, we define the Lévy process $X = M + A$ in terms of càdlàg processes M and A given via

$$M_t = \sigma B_t + \int_{[0,t]\times[-1,1]} x(P - \lambda \otimes \mu)(ds, dx)\,, \quad A_t = \mu t + \int_{[0,t]\times(\mathbb{R}\setminus[-1,1])} xP(ds, dx)$$

with P being a Poisson point process with intensity measure $\lambda \otimes \mu$ where λ denotes the Lebesgue measure. Then M is a square integrable martingale and A is a finite variation process. Hence X is a good integrator and one can show using Lemma 3.2 that

$$[M]_t = \sigma^2 t + \sum_{s\le t} (\Delta M_s)^2 = \sigma^2 t + \sum_{s\le t:\ |\Delta X_s|\le 1} (\Delta X_s)^2 \quad \text{for all } t \ge 0 \text{ a.s.}$$

From this it follows that $\xi \in \mathcal{I}_X$ provided that the process

$$\left(\sigma^2 \int_0^t \xi_s^2 ds + \sum_{s\le t:\ |\Delta X_s|\le 1} \xi_s^2 (\Delta X_s)^2\right)_{t\ge 0}$$

is locally integrable and

$$|\mu| \int_0^t |\xi_s|\ ds < \infty \quad \text{for all } t \ge 0 \text{ a.s.}$$

Note that one always has

$$\sum_{s\le t:\ |\Delta X_s|>1} |\xi_s| \cdot |\Delta X_s| < \infty \quad \text{for all } t \ge 0 \text{ a.s.}$$

simply because there are only finitely many jumps of magnitude > 1 on each interval $[0, t]$ when $t \ge 0$. In fact, in this particular case, the whole theory can also be derived by combining classical Itô calculus just for Brownian motion with Palm's formula for Poisson point processes.

References

[1] P. Protter, *Stochastic integration and differential equations*, Second Edition. Springer, 2005. doi: 10.1007/978-3-662-10061-5

[2] K. Bichteler, “Stochastic integration and L^p-theory of semimartingales”, *Ann. Prob.*, vol. 9, no. 1, pp. 49–89, 1981, doi: 10.1214/aop/1176994509

[3] K. Bichteler, *Stochastic integration with jumps*, vol. 89. Cambridge University Press, 2002. doi: 10.1017/CBO9780511549878

[4] G. Lowther, “Stochastic calculus notes”, 2010. Available: https://almostsure.wordpress.com/category/stochastic-calculus-notes/

[5] K. Bichteler, “Stochastic integrators”, *Bull. Amer. Math. Soc.*, vol. 1, no. 5, pp. 761–765, 1979.

[6] M. Beiglböck and P. Siorpaes, “Riemann-integration and a new proof of the Bichteler-Dellacherie theorem”, *Stochastic Process. Appl.*, vol. 124, no. 3, pp. 1226–1235, 2014, doi: 10.1016/j.spa.2013.10.001. Available: https://arxiv.org/abs/1201.1996

[7] O. Kallenberg, *Foundations of modern probability*, Second edition. Springer, 2021. doi: 10.1007/978-3-030-61871-1