5. Itô formula

We will now move to the fundamental tool associated with the stochastic integral: the celebrated Itô formula. We begin by an extended version of a definition from 275D:

Definition 5.1 *Given a filtration $\{\mathcal{F}_t\}_{t\geqslant 0}$, a* semimartingale *is a stochastic process of the form $\{M_t + A_t \colon t \geqslant 0\}$, where $M = \{M_t \colon t \geqslant 0\}$ is a local martingale and $A = \{A_t \colon t \geqslant 0\}$ is an adapted process with $A_0 = 0$ and finite first variation on any compact interval of times. The semimartingale is continuous if both M and A have continuous paths.*

As a consequence of Lemma 2.2, we have:

Lemma 5.2 *The processes M and A in the decomposition of a continuous semimartingale are unique up to indistinguishability.*

Proof. If $\{M_t + A_t \colon t \geqslant 0\}$ and $\{\widetilde{M}_t + \widetilde{A}_t \colon t \geqslant 0\}$ are two representations of the same process, then $M_t - \widetilde{M}_t = \widetilde{A}_t - A_t$ holds for all $t \geqslant 0$. Since $A - \widetilde{A}$ is of bounded variation, Lemma 2.2 implies that $P(\forall t \geqslant 0 \colon M_t - \widetilde{M}_t = M_t - \widetilde{M}_t) = 1$ a.s. $A_0 = \widetilde{A}_0 = 0$ implies $M_0 = \widetilde{M}_0$ and so we get that M and $\widetilde{M}$, and then also A and $\widetilde{A}$, are indistinguishable. $\square$

Definition 5.3 *Let X be a continuous semimartingale with decomposition $X = M + A$ for M a continuous local martingale and A a continuous, adapted process of locally bounded variation. We then set:*

$$\langle X \rangle_t := \langle M \rangle_t \tag{5.1}$$

and let

$$\int_0^t Y_s \mathrm{d}X_s := \int_0^t Y_s \mathrm{d}M_s + \int_0^t Y_s \mathrm{d}A_s \tag{5.2}$$

for any $t \geqslant 0$ whenever $Y \in \mathcal{V}_M^{\mathrm{loc}}$ is also Lebesgue-Stieltjes integrable with respect to A a.s. (The latter is the meaning of the integral on the right.)

A few remarks are in order. First, the formula (5.1) is consistent with the interpretation of $X \mapsto \langle X \rangle_t$ as quadratic variation (restricted to limit in probability of second variation for deterministic partitions whose mesh tends to infinity). Indeed, the fact that A, being continuous and of bounded first variation, will not contribute to the limiting second variation. The reader should bear in mind though that, in this interpretation, $\{X_t^2 - \langle X \rangle_t \colon t \geqslant 0\}$ is *not* necessarily a local martingale.

Our second remark concerns the existence of an adapted continuous version of the process in (5.2). Since the Lebesgue-Stieltjes integral of an adapted process Y with respect to a continuous and adapted A is necessarily continuous and adapted, this coincides with the existence of the continuous version for the Itô integral, for which $\mathcal{F}_0$ containing all P-null sets is sufficient.

With the above notions in place, we can now state:

Theorem 5.4 (Itô formula for semimartingales) *Let X be a continuous semimartingale and $f \in C^2(\mathbb{R})$. Then for all $t \geqslant 0$,*

$$f(X_t) = f(X_0) + \int_0^t f'(X_s)\mathrm{d}X_s + \frac{1}{2}\int_0^t f''(X_s)\mathrm{d}\langle X\rangle_s \quad \text{a.s.} \tag{5.3}$$

where the integrals exist in their respective meaning.

The proof of this theorem conveniently uses the following facts about Itô integral:

Lemma 5.5 *Let $M \in \mathscr{M}_{\text{loc}}^{\text{cont}}$ and suppose that $Y \in \mathcal{V}_M^{\text{loc}}$ and $\{Y^{(n)}\}_{n\in\mathbb{N}} \in (\mathcal{V}_M^{\text{loc}})^{\mathbb{N}}$ are such that, for some $t \geqslant 0$,*

$$\int_0^t \bigl(Y_s - Y_s^{(n)}\bigr)^2 \mathrm{d}\langle M\rangle_s \underset{n\to\infty}{\overset{P}{\longrightarrow}} 0. \tag{5.4}$$

Then also

$$\int_0^t Y_s^{(n)}\mathrm{d}M_s \underset{n\to\infty}{\overset{P}{\longrightarrow}} \int_0^t Y_s\mathrm{d}M_s. \tag{5.5}$$

Proof. This was left as a homework exercise in 275D so we give a proof. By linearity, we may assume that $Y = 0$. Define

$$\epsilon_n := \inf\left\{\epsilon \geqslant 0\colon P\left(\int_0^t (Y_s^{(n)})^2\mathrm{d}\langle M\rangle_s > 2\epsilon\right) > \epsilon\right\}. \tag{5.6}$$

The assumed convergence in probability then gives $\epsilon_n \to 0$. Next, for each $n \in \mathbb{N}$ consider the stopping time

$$\tau_n := \inf\left\{t \geqslant 0\colon \langle M\rangle_t \geqslant n \,\vee\, \int_0^t \bigl(Y_s^{(n)}\bigr)^2\mathrm{d}\langle M\rangle_s > \epsilon_n\right\}. \tag{5.7}$$

Since the probability that $\int_0^t (Y_s^{(n)})^2\mathrm{d}\langle M\rangle_s > \epsilon_n$ is no larger than ϵ_n, we have $\tau_n \to \infty$ in probability. But Itô isometry then shows

$$\begin{aligned} E\left(1_{\{\tau_n > t\}}\left|\int_0^t Y_s^{(n)}\mathrm{d}M_s\right|^2\right) &\leqslant E\left(\left|\int_0^t Y_s^{(n)} 1_{\{\tau_n > s\}}\mathrm{d}M_{s\wedge\tau_n}\right|^2\right) \\ &= E\left(\int_0^{t\wedge\tau_n} \bigl(Y_s^{(n)}\bigr)^2\mathrm{d}\langle M\rangle_s\right) \leqslant \epsilon_n. \end{aligned} \tag{5.8}$$

Chebyshev's inequality now gives $\int_0^t Y_s^{(n)}\mathrm{d}M_s \to 0$ in probability as desired. □

We will use this lemma through:

Lemma 5.6 *Let $M \in \mathscr{M}_{\text{loc}}^{\text{cont}}$ and suppose that $Y \in \mathcal{V}_M^{\text{loc}}$ has left-continuous locally bounded sample paths. Then for any $t \geqslant 0$ and any sequence $\{\Pi_n\}_{n\in\mathbb{N}}$ of partitions of $[0,t]$ such that $\Pi_n = \{0 = t_0^n < \cdots < t_{m_n}^n = t\}$ we have*

$$\|\Pi_n\| \to 0 \;\;\Rightarrow\;\; \sum_{i=1}^{m_n} Y_{t_{i-1}^n}(M_{t_i} - M_{t_{i-1}}) \underset{n\to\infty}{\overset{P}{\longrightarrow}} \int_0^t Y_s\mathrm{d}M_s. \tag{5.9}$$

In short, under left-continuity the left-endpoint Riemann-Stieltjes sums converge to the Itô integral in probability as the mesh of the partition tends to zero.

Proof. Setting

$$Y_s^{(n)} := \sum_{i=1}^{n} Y_{t_{i-1}^n} 1_{(t_{i-1}^n, t_i^n]}(s) \tag{5.10}$$

we have

$$\sum_{i=1}^{m_n} Y_{t_{i-1}^n}(M_{t_i} - M_{t_{i-1}}) = \int_0^t Y_s^{(n)} \mathrm{d}M_s. \tag{5.11}$$

The assumed left-continuity and local boundedness with the help of the Bounded Convergence Theorem validate (5.4) with convergence in pointwise-everywhere (no probability needed) sense. The claim then follows from Lemma 5.5. □

Proof of Theorem 5.4. As the proof is quite similar to that for integrals with respect to standard Brownian motion, we spell out only the main steps. Fix $t > 0$ and assume first that f, f' and f'' are bounded and that, for some deterministic $K > 0$, the objects in the semimartingale representation $X = M + A$ obey

$$\sup_{s \leqslant t} |M_s| \leqslant K \;\wedge\; \langle M \rangle_t \leqslant K \;\wedge\; V_t^{(1)}(A) \leqslant K. \tag{5.12}$$

Given $n \geqslant 0$, let $\Pi_n = \{0 = t_0 < \cdots < t_n = t\}$ be a partition such that $t_i - t_{i-1} = 1/n$ for each $i = 1, \dots, n$. Then

$$\begin{aligned} f(X_t) &= f(X_0) + \sum_{i=1}^{n} \big[f(X_{t_i}) - f(X_{t_{i-1}})\big] \\ &= f(X_0) + \sum_{i=1}^{n} f'(X_{t_{i-1}})(X_{t_i} - X_{t_{i-1}}) + \frac{1}{2}\sum_{i=1}^{n} f''(X_{t_{i-1}})\big(\langle X \rangle_{t_i} - \langle X \rangle_{t_{i-1}}\big) \\ &\quad + \frac{1}{2}\sum_{i=1}^{n} f''(X_{t_{i-1}})\Big[(X_{t_i} - X_{t_{i-1}})^2 - \big(\langle X \rangle_{t_i} - \langle X \rangle_{t_{i-1}}\big)\Big] \\ &\quad + \sum_{i=1}^{n} \int_0^1 \Big[f''\big(uX_{t_{i-1}} + (1-u)X_{t_i}\big) - f''(X_{t_{i-1}})\Big](X_{t_i} - X_{t_{i-1}})^2 u \mathrm{d}u, \end{aligned} \tag{5.13}$$

where we first invoked Taylor's theorem with a remainder and then employed some convenient rearrangements of terms in the sums. Lemma 5.6 along with the fact that Stietjels integral $\int f \mathrm{d}g$ exists whenever f is continuous and g of bounded variation imply

$$\sum_{i=1}^{n} f'(X_{t_{i-1}})(X_{t_i} - X_{t_{i-1}}) \underset{n\to\infty}{\overset{P}{\longrightarrow}} \int_0^t f'(X_s)\mathrm{d}X_s \tag{5.14}$$

and

$$\sum_{i=1}^{n} f''(X_{t_{i-1}})\big(\langle X \rangle_{t_i} - \langle X \rangle_{t_{i-1}}\big) \underset{n\to\infty}{\longrightarrow} \int_0^t f''(X_s)\mathrm{d}\langle X \rangle_s \tag{5.15}$$

We thus have to show that, as $n \to \infty$, the last two terms in (5.13) tend to zero in probability. Let us call these terms I_n and J_n, respectively.

As to the term I_n, here we write it as

$$I_n = \frac{1}{2}\big[I_n^{(1)} + 2I_n^{(2)} + I_n^{(3)}\big], \tag{5.16}$$

where

$$\begin{aligned} I_n^{(1)} &:= \sum_{i=1}^n f''(X_{t_{i-1}})(A_{t_i} - A_{t_{i-1}})^2 \\ I_n^{(2)} &:= \sum_{i=1}^n f''(X_{t_{i-1}})(A_{t_i} - A_{t_{i-1}})(M_{t_i} - M_{t_{i-1}}) \\ I_n^{(3)} &:= \sum_{i=1}^n f''(X_{t_{i-1}})\Big[(M_{t_i} - M_{t_{i-1}})^2 - \big(\langle M\rangle_{t_i} - \langle M\rangle_{t_{i-1}}\big)\Big] \end{aligned} \tag{5.17}$$

For the first of these we invoke the bound $|I_n^{(1)}| \leqslant \|f''\| V_t^{(2)}(A, \Pi_n)$ and then observe that

$$V_t^{(2)}(A, \Pi) \leqslant \operatorname{osc}_A\big([0,t], \|\Pi\|\big) V_t^{(1)}(A) \tag{5.18}$$

The total variation is less than K by (5.12) while the oscillation vanishes as $\|\Pi\| \to 0$ by uniform continuity of A. Hence $I_n^{(1)} \to 0$ pointwise. For the second term a Cauchy-Schwarz bound similarly gives

$$|I_n^{(2)}| \leqslant \|f''\|\, V_t^{(2)}(A, \Pi_n)^{1/2}\, V_t^{(2)}(M, \Pi_n)^{1/2} \tag{5.19}$$

which tends to zero in probability by the fact $V_t^{(2)}(M, \Pi_n)$ is convergent in probability and thus forms a tight sequence.

For the last term we have to work a bit harder. Noting that $I_n^{(3)}$ is the sum of uniformly bounded martingale increments shows

$$\begin{aligned} E\big[(I_n^{(3)})^2\big] &\leqslant \|f''\|^2 \sum_{i=1}^n E\bigg(\Big[(M_{t_i} - M_{t_{i-1}})^2 - \big(\langle M\rangle_{t_i} - \langle M\rangle_{t_{i-1}}\big)\Big]^2\bigg) \\ &\leqslant 2\|f''\|^2 \Big[E\big(V_t^{(4)}(M, \Pi_n)\big) + E\big(V_t^{(2)}(\langle M\rangle, \Pi_n)\big)\Big] \end{aligned} \tag{5.20}$$

where we invoked the inequality $(a+b)^2 \leqslant 2a^2 + 2b^2$ and interpreted the resulting sums using the fourth and second variation. Note that, for any $p > p' > 1$ and any $h\colon [0,t] \to \mathbb{R}$, we have $V_t^{(p)}(h, \Pi) \leqslant \operatorname{osc}_h([0,t], \|\Pi\|)^{p'} V_t^{(p-p')}(f, \Pi)$. In light of $V_t^{(1)}(M) \leqslant K$ assumed in (5.12), this implies

$$E\big(V_t^{(2)}(\langle M\rangle, \Pi_n)\big) \leqslant K E\big(\operatorname{osc}_{\langle M\rangle}([0,t], \|\Pi_n\|)\big) \tag{5.21}$$

which tends to zero by the Bounded Convergence Theorem and the fact that the oscillation is bounded by $2K$ thanks to the second inequality in (5.12) and it tends to zero pointwise by continuity of $\langle M\rangle$. Using the above with $p := 4$ and $p' := 2$ and employing Cauchy-Schwarz similarly yields

$$E\big(V_t^{(4)}(M, \Pi_n)\big) \leqslant E\big(\operatorname{osc}_M([0,t], \|\Pi_n\|)^2\big) \tag{5.22}$$

which again tends to zero as $n \to \infty$ by the Bounded Convergence Theorem and the fact that the oscillation is bounded by $2K$ by the first inequality in (5.12). This implies $I_n^{(3)} \to 0$ in L^2 and, combining with the above, $I_n \to 0$ in probability.

Concerning J_n, here we use the bound

$$|J_n| \leqslant \operatorname{osc}_{f''\circ X}\big([0,t], \|\Pi_n\|\big)\, V_t^{(2)}(X,\Pi_n). \tag{5.23}$$

The assumed continuity of f'' implies continuity of $f'' \circ X$ which means that the oscillation tends to zero as $n \to \infty$ pointwise. Since

$$V_t^{(2)}(M+A,\Pi) \leqslant 2V_t^{(2)}(M,\Pi) + 2V_t^{(2)}(A,\Pi) \tag{5.24}$$

the sequence $V_t^{(2)}(X,\Pi_n)$ is tight. Hence, $J_n \to 0$ in probability as well.

The above proves the Itô formula under the assumptions made at the beginning of the proof; namely, that f, f' and f'' are bounded and (5.13) holds. Let

$$\tau_K := \inf\Big\{t \geqslant 0\colon \sup_{s\leqslant t}|X_s| \geqslant K \,\vee\, \text{(5.12) fails}\Big\}. \tag{5.25}$$

Then the above proof gives

$$f(X_{t\wedge\tau_K}) = f(X_0) + \int_0^t f'(X_s)1_{\{\tau_K>s\}}\mathrm{d}X_s + \frac{1}{2}\int_0^t f''(X_s)1_{\{\tau_K>s\}}\mathrm{d}\langle X\rangle_s. \tag{5.26}$$

(Technically speaking, we also need to formally argue that f can be changed outside the interval $[-K,K]$ so that f, f' and f'' are bounded everywhere. This is done by extending f by the second-order Taylor polynomials centered at $\pm K$.) As $\tau_K \to \infty$ a.s. as $K \to \infty$, we now pass to $K \to \infty$ limit inside these integrals using ordinary Bounded Convergence Theorem in the Lebesgue-Stieltjes integrals supplied in addition with Lemma 5.5 for the stochastic integral. This leads to the desired limit formula (5.3). $\square$

The formula (5.3) is actually the simplest of several Itô formulas. Indeed, one can consider a function $f\colon \mathbb{R}^d \to \mathbb{R}$ of $\mathbb{R}^d$-valued X, in which case we get

$$f(X_t) = f(X_0) + \int_0^t \nabla f(X_s)\cdot \mathrm{d}X_s + \frac{1}{2}\int_0^t \nabla\nabla f(X_s)\cdot \mathrm{d}\langle X,X\rangle_s \tag{5.27}$$

where we think of $\nabla\nabla f$ as a $d\times d$ matrix and contact the indices against those of the $d\times d$-matrix $\langle X,X\rangle$. (In particular, for X being a standard Brownian motion $d\langle X,X\rangle_t$ is the identity matrix times $\mathrm{d}t$ and $\nabla\nabla f$ then contracts into the Laplacian Δf.)

Another generalization is in the direction of functions of X_t and also of the time t and the quadratic variation process $\langle X\rangle_t$. An ordinary integral of the corresponding first derivative then pops up for these additional components on the right-hand side. Yet another generalization would be that of several processes entering the arguments of f. All of these are handled by the infinitesimal rules of stochastic calculus

$$\mathrm{d}M_t\mathrm{d}\widetilde{M}_t = \mathrm{d}\langle M,\widetilde{M}\rangle_t \;\wedge\; \mathrm{d}t\mathrm{d}M_t = 0 \;\wedge\; (\mathrm{d}t)^2 = 0 \tag{5.28}$$

that allow treating each case of interest depending on context.

As our final remark we note that requiring $f \in C^2(\mathbb{R})$ is not necessary for Itô formula to hold. Indeed, straightforward approximation arguments (transfered to the stochastic integral with the help of Lemma 5.5) show that continuity of f'' can be weakened to f' being absolutely continuous. We will discuss this extension when we push the Itô formula even beyond that case using the concept of the Brownian local time.

Further reading: Section 3.3A of Karatzas-Shreve