20

Stochastic Integral and the Ito Formula

20.1 Quadratic Variation of Square-Integrable Martingales

In this section we shall apply the Doob-Meyer Decomposition to submartingales of the form X_t^2, where X_t is a square-integrable martingale with continuous sample paths. This decomposition will be essential in the construction of the stochastic integral in the next section.

We shall call two random processes equivalent if they are indistinguishable. We shall often use the same notation for a process and the equivalence class it represents.

Definition 20.1. *Let $\mathcal{F}_t$ be a filtration on a probability space $(\Omega, \mathcal{F}, \mathrm{P})$. Let $\mathcal{M}_2^c$ denote the space of all equivalence classes of square-integrable martingales which start at zero, and whose sample paths are continuous almost surely. That is, $X_t \in \mathcal{M}_2^c$ if $(X_t, \mathcal{F}_t)$ is a square-integrable martingale, $X_0 = 0$ almost surely, and X_t is continuous almost surely.*

We shall always assume that the filtration $\mathcal{F}_t$ satisfies the usual conditions (as is the case, for example, if $\mathcal{F}_t$ is the augmented filtration for a Brownian motion).

Let us consider the process X_t^2. Since it is equal to a convex function (namely x^2) applied to the martingale X_t, the process X_t^2 is a submartingale. Let S_a be the set of all stopping times bounded by a. If $\tau \in S_a$, by the Optional Sampling Theorem

$$\int_{\{X_\tau^2 > \lambda\}} X_\tau^2 d\mathrm{P} \leq \int_{\{X_\tau^2 > \lambda\}} X_a^2 d\mathrm{P}.$$

By the Chebyshev Inequality,

$$\mathrm{P}(X_\tau^2 > \lambda) \leq \frac{\mathrm{E}X_\tau^2}{\lambda} \leq \frac{\mathrm{E}X_a^2}{\lambda} \to 0 \quad \text{as} \quad \lambda \to \infty.$$

L. Koralov and Y.G. Sinai, *Theory of Probability and Random Processes*, Universitext, DOI 10.1007/978-3-540-68829-7_20,

Since the integral is an absolutely continuous function of sets,

$$\lim_{\lambda\to\infty} \sup_{\tau\in S_a} \int_{\{X_\tau^2>\lambda\}} X_\tau^2 d\mathrm{P} = 0,$$

that is, the set of random variables $\{X_\tau\}_{\tau\in S_a}$ is uniformly integrable.

Therefore, we can apply the Doob-Meyer Decomposition (Theorem 13.26) to conclude that there are unique (up to indistinguishability) processes M_t and A_t, whose paths are continuous almost surely, such that $X_t^2 = M_t + A_t$, where $(M_t, \mathcal{F}_t)$ is a martingale, and A_t is an adapted non-decreasing process, and $M_0 = A_0 = 0$ almost surely.

Definition 20.2. *The process A_t in the above decomposition $X_t^2 = M_t + A_t$ of the square of the martingale $X_t \in \mathcal{M}_2^c$ is called the quadratic variation of X_t and is denoted by $\langle X\rangle_t$.*

Example. Let us prove that $\langle W\rangle_t = t$. Indeed for $s \le t$,

$$\mathrm{E}(W_t^2|\mathcal{F}_s) = \mathrm{E}((W_t - W_s)^2|\mathcal{F}_s) + 2\mathrm{E}(W_tW_s|\mathcal{F}_s) - \mathrm{E}(W_s^2|\mathcal{F}_s) = W_s^2 + t - s.$$

Therefore, $W_t^2 - t$ is a martingale, and $\langle W\rangle_t = t$ due to the uniqueness of the Doob-Meyer Decomposition.

Example. Let $X_t \in \mathcal{M}_2^c$ and τ be a stopping time of the filtration $\mathcal{F}_t$ (here τ is allowed to take the value ∞ with positive probability). Thus, the process $Y_t = X_{t\wedge\tau}$ also belongs to $\mathcal{M}_2^c$. Indeed, it is a continuous martingale by Lemma 13.29. It is square-integrable since $Y_t\chi_{\{t<\tau\}} = X_t\chi_{\{t<\tau\}}$, while

$$Y_t\chi_{\{\tau\le t\}} = X_\tau\chi_{\{\tau\le t\}} = \mathrm{E}(X_t\chi_{\{\tau\le t\}}|\mathcal{F}_\tau) \in L^2(\Omega, \mathcal{F}, \mathrm{P}).$$

Since $X_t^2 - \langle X\rangle_t$ is a continuous martingale, the process $X_{t\wedge\tau}^2 - \langle X\rangle_{t\wedge\tau}$ is also a martingale by Lemma 13.29. Since $\langle X\rangle_{t\wedge\tau}$ is an adapted non-decreasing process, we conclude from the uniqueness of the Doob-Meyer Decomposition that $\langle Y\rangle_t = \langle X\rangle_{t\wedge\tau}$.

Lemma 20.3. *Let $X_t \in \mathcal{M}_2^c$. Let τ be a stopping time such that $\langle X\rangle_\tau = 0$ almost surely. Then $X_t = 0$ for all $0 \le t \le \tau$ almost surely.*

Proof. Since $\langle X\rangle_t$ is non-decreasing, $\langle X\rangle_{t\wedge\tau} = 0$ almost surely for each t. By Lemma 13.29, the process $X_{t\wedge\tau}^2 - \langle X\rangle_{t\wedge\tau}$ is a martingale. Therefore, since the expectation of a martingale is a constant,

$$\mathrm{E}X_{t\wedge\tau}^2 = \mathrm{E}(X_{t\wedge\tau}^2 - \langle X\rangle_{t\wedge\tau}) = 0$$

for each $t \ge 0$, that is $X_{t\wedge\tau} = 0$ almost surely. Since X_t is continuous, $X_t = 0$ for all $0 \le t \le \tau$ almost surely. □

Clearly, the linear combinations of elements of $\mathcal{M}_2^c$ are also elements of $\mathcal{M}_2^c$.

Definition 20.4. *Let two processes X_t and Y_t belong to $\mathcal{M}_2^c$. We define their cross-variation as*

$$\langle X, Y\rangle_t = \frac{1}{4}(\langle X+Y\rangle_t - \langle X-Y\rangle_t). \tag{20.1}$$

Clearly, $X_tY_t - \langle X, Y\rangle_t$ is a continuous martingale, the cross-variation is bi-linear and symmetric in X and Y, and $|\langle X, Y\rangle_t|^2 \leq \langle X\rangle_t\langle Y\rangle_t$.

Let us introduce a metric which will turn $\mathcal{M}_2^c$ into a complete metric space.

Definition 20.5. *For $X, Y \in \mathcal{M}_2^c$ and $0 \leq t < \infty$, we define*

$$||X||_t = \sqrt{\mathrm{E}X_t^2}, \quad \text{and} \quad d_{\mathcal{M}}(X, Y) = \sum_{n=1}^{\infty} \frac{1}{2^n} \min(||X-Y||_n, 1).$$

In order to prove that $d_{\mathcal{M}}$ is a metric, we need to show, in particular, that $d_{\mathcal{M}}(X, Y) = 0$ implies that $X_t - Y_t$ is indistinguishable from zero. If $d_{\mathcal{M}}(X, Y) = 0$, then $X_n - Y_n = 0$ almost surely for every positive integer n. Since $X_t - Y_t$ is a martingale, $X_t - Y_t = \mathrm{E}(X_n - Y_n|\mathcal{F}_t) = 0$ almost surely for every $0 \leq t \leq n$. Therefore,

$$\mathrm{P}(\{\omega : X_t(\omega) - Y_t(\omega) = 0 \ \text{ for all rational } t\}) = 1.$$

This implies that $X_t - Y_t$ is indistinguishable from zero, since it is continuous almost surely. It is clear that $d_{\mathcal{M}}$ has all the other properties required of a metric. Let us show that the space $\mathcal{M}_2^c$ is complete, which will be essential in the construction of the stochastic integral.

Lemma 20.6. *The space $\mathcal{M}_2^c$ with the metric $d_{\mathcal{M}}$ is complete.*

Proof. Let X_t^m be a Cauchy sequence in $\mathcal{M}_2^c$. Then X_n^m is a Cauchy sequence in $L^2(\Omega, \mathcal{F}_n, \mathrm{P})$ for each n. If $t \leq n$, then $\mathrm{E}|X_t^{m_1} - X_t^{m_2}|^2 \leq \mathrm{E}|X_n^{m_1} - X_n^{m_2}|^2$ for all m_1 and m_2, since $|X_t^{m_1} - X_t^{m_2}|^2$ is a submartingale. This proves that X_t^m is a Cauchy sequence in $L^2(\Omega, \mathcal{F}_t, \mathrm{P})$ for each t. Let X_t be defined for each t as the limit of X_t^m in $L^2(\Omega, \mathcal{F}_t, \mathrm{P})$. Let $0 \leq s \leq t$, and $A \in \mathcal{F}_s$. Then,

$$\int_A X_t d\mathrm{P} = \lim_{m\to\infty} \int_A X_t^m d\mathrm{P} = \lim_{m\to\infty} \int_A X_s^m d\mathrm{P} = \int_A X_s d\mathrm{P},$$

where the middle equality follows from X_t^m being a martingale, and the other two are due to the L^2 convergence. This shows that $(X_t, \mathcal{F}_t)$ is a martingale. By Lemma 13.25, we can choose a right-continuous modification of X_t. We can therefore apply the Doob Inequality (Theorem 13.30) to the submartingale $|X_t^m - X_t|^2$ to obtain

$$\mathrm{P}(\sup_{0\leq s\leq t} |X_s^m - X_s| \geq \lambda) \leq \frac{1}{\lambda^2}\mathrm{E}|X_t^m - X_t|^2 \to 0 \ \text{ as } m \to \infty$$

for any t. We can, therefore, extract a subsequence m_k such that

$$\mathrm{P}(\sup_{0 \leq s \leq t} |X_s^{m_k} - X_s| \geq \frac{1}{k}) \leq \frac{1}{2^k} \quad \text{for } k \geq 1.$$

The First Borel-Cantelli Lemma implies that $X_t^{m_k}$ converges to X_t uniformly on $[0, t]$ for almost all ω. Since t was arbitrary, this implies that X_t is continuous almost surely, and thus $\mathcal{M}_2^c$ is complete. □

Next we state a lemma which explains the relation between the quadratic variation of a martingale (as in Definition 20.1) and the second variation of the martingale over a partition (as in Sect. 3.2).

More precisely, let f be a function defined on an interval $[a, b]$ of the real line. Let $\sigma = \{t_0, t_1, \ldots, t_n\}$, $a = t_0 \leq t_1 \leq \ldots \leq t_n = b$, be a partition of the interval $[a, b]$ into n subintervals. We denote the length of the largest interval by $\delta(\sigma) = \max_{1 \leq i \leq n}(t_i - t_{i-1})$. Let $V_{[a,b]}^2(f, \sigma) = \sum_{i=1}^n |f(t_i) - f(t_{i-1})|^2$ be the second variation of the function f over the partition σ.

Lemma 20.7. *Let $X_t \in \mathcal{M}_2^c$ and $t \geq 0$ be fixed. Then, for any $\varepsilon > 0$*

$$\lim_{\delta(\sigma) \to 0} \mathrm{P}(|V_{[0,t]}^2(X_s, \sigma) - \langle X \rangle_t| > \varepsilon) = 0.$$

We omit the proof of this lemma, instead referring the reader to "Brownian Motion and Stochastic Calculus" by I. Karatzas and S. Shreve. Note, however, that Lemma 18.24 contains a stronger statement (convergence in L^2 instead of convergence in probability) when the martingale X_t is a Brownian motion.

Corollary 20.8. *Assume that $V_{[0,t]}^1(X_s(\omega)) < \infty$ for almost all $\omega \in \Omega$, where $X_t \in \mathcal{M}_2^c$ and $t \geq 0$ is fixed. Then $X_s(\omega) = 0$, $s \in [0, t]$, for almost all $\omega \in \Omega$.*

Proof. Let us assume the contrary. Then, by Lemma 20.3, there is a positive constant c_1 and an event $A' \subseteq \Omega$ with $\mathrm{P}(A') > 0$ such that $\langle X \rangle_t(\omega) \geq c_1$ for almost all $\omega \in A'$. Since $V_{[0,t]}^1(X_s(\omega)) < \infty$ for almost all $\omega \in A'$, we can find a constant c_2 and a subset $A'' \subseteq A'$ with $\mathrm{P}(A'') > 0$ such that $V_{[0,t]}^1(X_s(\omega)) \leq c_2$ for almost all $\omega \in A''$.

Let σ_n be a sequence of partitions of $[0, t]$ into 2^n intervals of equal length. By Lemma 20.7, we can assume, without loss of generality, that $V_{[0,t]}^2(X_s(\omega), \sigma_n) \neq 0$ for large enough n almost surely on A''. Since a continuous function is also uniformly continuous,

$$\lim_{n \to \infty} \frac{V_{[0,t]}^1(X_s(\omega), \sigma_n)}{V_{[0,t]}^2(X_s(\omega), \sigma_n)} = \infty \quad \text{almost surely on } A''.$$

This, however, contradicts $V_{[0,t]}^2(X_s(\omega), \sigma_n) \to \langle X \rangle_t(\omega) \geq c_1$ (in probability), while $\lim_{n \to \infty} V_{[0,t]}^1(X_s(\omega), \sigma_n) = V_{[0,t]}^1(X_s(\omega)) \leq c_2$ for almost all $\omega \in A''$. □

Lemma 20.9. *Let $X_t, Y_t \in \mathcal{M}_2^c$. There is a unique (up to indistinguishability) adapted continuous process of bounded variation A_t such that $A_0 = 0$ almost surely and $X_t Y_t - A_t$ is a martingale. In fact, $A_t = \langle X, Y \rangle_t$.*

Proof. The existence part was demonstrated above. Suppose there are two processes A_t^1 and A_t^2 with the desired properties. Then $M_t = A_t^1 - A_t^2$ is a continuous martingale with bounded variation. Define the sequence of stopping times $\tau_n = \inf\{t \geq 0 : |M_t| = n\}$, where the infimum of an empty set is equal to $+\infty$. This is a non-decreasing sequence, which tends to infinity almost surely. Note that $M_t^{(n)} = M_{t \wedge \tau_n}$ is a square-integrable martingale for each n (by Lemma 13.29), and that $M_t^{(n)}$ is also a process of bounded variation. By Corollary 20.8, $M_t^{(n)} = 0$ for all t almost surely. Since $\tau_n \to \infty$, A_t^1 and A_t^2 are indistinguishable. □

An immediate consequence of this result is the following lemma.

Lemma 20.10. *Let $X_t, Y_t \in \mathcal{M}_2^c$ with the filtration $\mathcal{F}_t$, and let τ be a stopping time for $\mathcal{F}_t$. Then $\langle X, Y \rangle_{t \wedge \tau}$ is the cross-variation of the processes $X_{t \wedge \tau}$ and $Y_{t \wedge \tau}$.*

20.2 The Space of Integrands for the Stochastic Integral

Let $(M_t, \mathcal{F}_t)_{t \in \mathbb{R}^+}$ be a continuous square-integrable martingale on a probability space $(\Omega, \mathcal{F}, \mathrm{P})$, and let X_t be an adapted process. In this chapter we shall define the stochastic integral $\int_0^t X_s dM_s$, also denoted by $I_t(X)$.

We shall carefully state additional assumptions on X_t in order to make sense of the integral. Note that the above expression cannot be understood as the Lebesgue-Stieltjes integral defined for each ω, unless $\langle M \rangle_t(\omega) = 0$. Indeed, the function $M_s(\omega)$ has unbounded first variation on the interval $[0, t]$ if $\langle M \rangle_t(\omega) \neq 0$, as discussed in the previous section.

While the stochastic integral could be defined for a general square integrable martingale M_t (by imposing certain restrictions on the process X_t), we shall stick to the assumption that $M_t \in \mathcal{M}_2^c$. Our prime example is $M_t = W_t$.

Let us now discuss the assumptions on the integrand X_t. We introduce a family of measures μ_t, $0 \leq t < \infty$, associated to the process M_t, on the product space $\Omega \times [0, t]$ with the σ-algebra $\mathcal{F} \times \mathcal{B}([0, t])$.

Namely, let $\mathcal{K}$ be the collection of sets of the form $A = B \times [a, b]$, where $B \in \mathcal{F}$ and $[a, b] \subseteq [0, t]$. Let $\mathcal{G}$ be the collection of measurable sets $A \in \mathcal{F} \times \mathcal{B}([0, t])$ for which $\int_0^t \chi_A(\omega, s) d\langle M \rangle_s(\omega)$ exists for almost all ω and is a measurable function of ω. Note that $\mathcal{K} \subseteq \mathcal{G}$, that $\mathcal{K}$ is a π-system, and that $\mathcal{G}$ is closed under unions of non-intersecting sets and complements in $\Omega \times [0, t]$. Therefore, $\mathcal{F} \times \mathcal{B}([0, t]) = \sigma(\mathcal{K}) = \mathcal{G}$, where the second equality is due to Lemma 4.12.

We can now define

$$\mu_t(A) = \mathrm{E} \int_0^t \chi_A(\omega, s) d\langle M\rangle_s(\omega),$$

where $A \in \mathcal{F} \times \mathcal{B}([0,t])$. The expectation exists since the integral is a measurable function of ω bounded from above by $\langle M\rangle_t$. The fact that μ_t is σ-additive (that is a measure) follows from the Levi Convergence Theorem. If f is defined on $\Omega \times [0,t]$ and is measurable with respect to the σ-algebra $\mathcal{F} \times \mathcal{B}([0,t])$, then

$$\int_{\Omega\times[0,t]} f d\mu_t = \mathrm{E} \int_0^t f(\omega, s) d\langle M\rangle_s(\omega).$$

(If the function f is non-negative, and the expression on one side of the equality is defined, then the expression on the other side is also defined. If the function f is not necessarily non-negative, and the expression on the left-hand side is defined, then the expression on the right-hand side is also defined). Indeed, this formula is true for indicator functions of measurable sets, and therefore, for simple functions with a finite number of values. It also holds for non-negative functions since they can be approximated by monotonic sequences of simple functions with a finite number of values. Furthermore, any function can be represented as a difference of two non-negative functions, and thus, if the expression on the left-hand side is defined, so is the one on the right-hand side.

We can also consider the σ-finite measure μ on the product space $\Omega \times \mathbb{R}^+$ with the σ-algebra $\mathcal{F} \times \mathcal{B}(\mathbb{R}^+)$ whose restriction to $\Omega \times [0,t]$ coincides with μ_t for each t. For example, if $M_t = W_t$, then $d\langle M\rangle_t(\omega)$ is the Lebesgue measure for each ω, and μ is equal to the product of the measure P and the Lebesgue measure on the half-line.

Let $\mathcal{H}_t = L^2(\Omega \times [0,t], \mathcal{F} \times \mathcal{B}([0,t]), \mu_t)$, and $||\cdot||_{\mathcal{H}_t}$ be the L^2 norm on this space. We define $\mathcal{H}$ as the space of classes of functions on $\Omega \times \mathbb{R}^+$ whose restrictions to $\Omega \times [0,t]$ belong to $\mathcal{H}_t$ for every $t \geq 0$. Two functions f and g belong to the same class, and thus correspond to the same element of $\mathcal{H}$, if $f = g$ almost surely with respect to the measure μ. We can define the metric on $\mathcal{H}$ by

$$d_{\mathcal{H}}(f,g) = \sum_{n=1}^{\infty} \frac{1}{2^n} \min(||f-g||_{\mathcal{H}_n}, 1).$$

It is easy to check that this turns $\mathcal{H}$ into a complete metric space.

We shall define the stochastic integral $I_t(X)$ for all progressively measurable processes X_t such that $X_t \in \mathcal{H}$. We shall see that $I_t(X)$ is indistinguishable from $I_t(Y)$ if X_t and Y_t coincide as elements of $\mathcal{H}$. The set of elements of $\mathcal{H}$ which have a progressively measurable representative will be denoted by $\mathcal{L}^*$ or $\mathcal{L}^*(M)$, whenever it is necessary to stress the dependence on the martingale M_t. It can be also viewed as a metric space with the metric $d_{\mathcal{H}}$, and it can be shown that this space is also complete (although we will not use this fact).

Lemma 20.11. *Let X_t be a progressively measurable process and A_t a continuous adapted processes such that $A_t(\omega)$ almost surely has bounded variation on any finite interval and*

$$Y_t(\omega) = \int_0^t X_s(\omega)dA_s(\omega) < \infty \quad \textit{almost surely.}$$

Then Y_t is progressively measurable.

Proof. As before, X_t can be approximated by simple functions from below, proving $\mathcal{F}_t$-measurability of Y_t for fixed t. The process Y_t is progressively measurable since it is continuous. □

20.3 Simple Processes

In this section we again assume that we have a probability space $(\Omega, \mathcal{F}, \mathrm{P})$ and a continuous square-integrable martingale $M_t \in \mathcal{M}_2^c$.

Definition 20.12. *A process X_t is called simple if there are a strictly increasing sequence of real numbers t_n, $n \geq 0$, such that $t_0 = 0$, $\lim_{n\to\infty} t_n = \infty$, and a sequence of bounded random variables ξ_n, $n \geq 0$, such that ξ_n is $\mathcal{F}_{t_n}$-measurable for every n and*

$$X_t(\omega) = \xi_0(\omega)\chi_{\{0\}}(t) + \sum_{n=0}^{\infty} \xi_n(\omega)\chi_{(t_n,t_{n+1}]}(t) \quad \text{for } \omega \in \Omega, t \geq 0. \qquad (20.2)$$

The class of all simple processes will be denoted by $\mathcal{L}_0$.

It is clear that $\mathcal{L}_0 \subseteq \mathcal{L}^*$. We shall first define the stochastic integral for simple processes. Then we shall extend the definition to all the integrands from $\mathcal{L}^*$ with the help of the following lemma.

Lemma 20.13. *The space $\mathcal{L}_0$ is dense in $\mathcal{L}^*$ in the metric $d_{\mathcal{H}}$ of the space $\mathcal{H}$.*

The lemma states that, given a process $X_t \in \mathcal{L}^*$, we can find a sequence of simple processes X_t^n such that $\lim_{n\to\infty} d_{\mathcal{H}}(X_t^n, X_t) = 0$. We shall only prove this for X_t continuous for almost all ω, the general case being somewhat more complicated.

Proof. It is sufficient to show that for each integer m there is a sequence of simple processes X_t^n such that

$$\lim_{n\to\infty} ||X_t^n - X_t||_{\mathcal{H}_m} = 0. \qquad (20.3)$$

Indeed, if this is the case, then for each m we can find a simple process $X_t^{(m)}$ such that $||X_t^{(m)} - X_t||_{\mathcal{H}_m} \leq 1/m$. Then $\lim_{m\to\infty} d_{\mathcal{H}}(X_t^{(m)}, X_t) = 0$ as required. Let m be fixed, and

$$X_t^n(\omega) = X_0(\omega)\chi_{\{0\}}(t) + \sum_{k=0}^{n-1} X_{km/n}(\omega)\chi_{(km/n,(k+1)m/n]}(t).$$

This sequence converges to X_t almost surely uniformly in $t \in [0, m]$, since X_t is continuous almost surely. If X_t is bounded on the interval $[0, m]$ (that is, $|X_t(\omega)| \le c$ for all $\omega \in \Omega$, $t \in [0, m]$), then, by the Lebesgue Dominated Convergence Theorem, $\lim_{n\to\infty} ||X_t^n - X_t||_{\mathcal{H}_m} = 0$. If X_t is not necessarily bounded, it can be first approximated by bounded processes as follows. Let

$$Y_t^n(\omega) = \begin{cases} -n & \text{if} \quad X_t(\omega) < -n, \\ X_t(\omega) & \text{if} \quad -n \le X_t(\omega) \le n, \\ n & \text{if} \quad X_t(\omega) > n. \end{cases}$$

Note that Y_t^n are continuous progressively measurable processes, which are bounded on $[0, m]$. Moreover, $\lim_{n\to\infty} ||Y_t^n - X_t||_{\mathcal{H}_m} = 0$. Each of the processes Y_t^n can, in turn, be approximated by a sequence of simple processes. Therefore, (20.3) holds for some sequence of simple processes. Thus, we have shown that for an almost surely continuous progressively measurable process X_t, there is a sequence of simple processes X_t^n such that $\lim_{n\to\infty} d_{\mathcal{H}}(X_t^n, X_t) = 0$. □

20.4 Definition and Basic Properties of the Stochastic Integral

We first define the stochastic (Ito) integral for a simple process,

$$X_t(\omega) = \xi_0(\omega)\chi_{\{0\}}(t) + \sum_{n=0}^{\infty} \xi_n(\omega)\chi_{(t_n,t_{n+1}]}(t) \quad \text{for } \omega \in \Omega, t \ge 0. \tag{20.4}$$

Definition 20.14. *The stochastic integral $I_t(X)$ of the process X_t is defined as*

$$I_t(X) = \sum_{n=0}^{m(t)-1} \xi_n(M_{t_{n+1}} - M_{t_n}) + \xi_{m(t)}(M_t - M_{t_{m(t)}}),$$

where $m(t)$ is the unique integer such that $t_{m(t)} \le t < t_{m(t)+1}$.

When it is important to stress the dependence of the integral on the martingale, we shall denote it by $I_t^M(X)$. While the same process can be represented in the form (20.4) with different ξ_n and t_n, the definition of the integral does not depend on the particular representation.

Let us study some properties stochastic integral. First, note that $I_0(X) = 0$ almost surely. It is clear that the integral is linear in the integrand, that is,

$$I_t(aX + bY) = aI_t(X) + bI_t(Y) \tag{20.5}$$

for any $X, Y \in \mathcal{L}_0$ and $a, b \in \mathbb{R}$. Also, $I_t(X)$ is continuous almost surely since M_t is continuous. Let us show that $I_t(X)$ is a martingale. If $0 \le s < t$, then

$$\mathrm{E}((I_t(X) - I_s(X))|\mathcal{F}_s)$$

$$= \mathrm{E}(\xi_{m(s)-1}(M_{t_{m(s)}} - M_s) + \sum_{n=m(s)}^{m(t)-1} \xi_n(M_{t_{n+1}} - M_{t_n}) + \xi_n(M_t - M_{t_{m(t)}})|\mathcal{F}_s).$$

Since ξ_n is $\mathcal{F}_{t_n}$-measurable and M_t is a martingale, the conditional expectation with respect to $\mathcal{F}_s$ of each of the terms on the right-hand side is equal to zero. Therefore, $\mathrm{E}(I_t(X) - I_s(X)|\mathcal{F}_s) = 0$, which proves that I_t is a martingale.

The process $I_t(X)$ is square-integrable since M_t is square-integrable and the random variables ξ_n are bounded. Let us find its quadratic variation. Let $0 \le s < t$. Assume that $t_{m(t)} > s$ (the case when $t_{m(t)} \le s$ can be treated similarly). Then,

$$\begin{aligned}
&\mathrm{E}(I_t^2(X) - I_s^2(X)|\mathcal{F}_s) \\
&\quad = \mathrm{E}((I_t(X) - I_s(X))^2|\mathcal{F}_s) \\
&\quad = \mathrm{E}((\xi_{m(s)}(M_{t_{m(s)+1}} - M_s) + \sum_{n=m(s)+1}^{m(t)-1} \xi_n(M_{t_{n+1}} - M_{t_n}) \\
&\qquad + \xi_{m(t)}(M_t - M_{t_{m(t)}}))^2|\mathcal{F}_s) \\
&\quad = \mathrm{E}(\xi_{m(s)}^2(M_{t_{m(s)+1}} - M_s)^2 + \sum_{n=m(s)+1}^{m(t)-1} \xi_n^2(M_{t_{n+1}} - M_{t_n})^2 \\
&\qquad + \xi_{m(t)}^2(M_t - M_{t_{m(t)}})^2|\mathcal{F}_s) \\
&\quad = \mathrm{E}(\xi_{m(s)}^2(\langle M\rangle_{t_{m(s)+1}} - \langle M\rangle_s) + \sum_{n=m(s)+1}^{m(t)-1} \xi_n^2(\langle M\rangle_{t_{n+1}} - \langle M\rangle_{t_n}) \\
&\qquad + \xi_{m(t)}^2(\langle M\rangle_t - \langle M\rangle_{t_{m(t)}})|\mathcal{F}_s) = \mathrm{E}(\int_s^t X_u^2 d\langle M\rangle_u|\mathcal{F}_s).
\end{aligned}$$

This implies that the process $I_t^2(X) - \int_0^t X_u^2 d\langle M\rangle_u$ is a martingale. Since the process $\int_0^t X_u^2 d\langle M\rangle_u$ is $\mathcal{F}_t$-adapted (as follows from the definition of a simple process), we conclude from the uniqueness of the Doob-Meyer Decomposition that $\langle I(X)\rangle_t = \int_0^t X_u^2 d\langle M\rangle_u$. Also, by setting $s = 0$ in the calculation above and taking expectation on both sides,

$$\mathrm{E}I_t^2(X) = \mathrm{E}\int_0^t X_u^2 d\langle M\rangle_u. \tag{20.6}$$

Recall that we have the metric $d_{\mathcal{M}}$ given by the family of norms $||\cdot||_n$ on the space $\mathcal{M}_2^c$ of martingales, and the metric $d_{\mathcal{H}}$ given by the family of norms

$||\cdot||_{\mathcal{H}_n}$ on the space $\mathcal{L}^*$ of integrands. So far, we have defined the stochastic integral as a mapping from the subspace $\mathcal{L}_0$ into $\mathcal{M}_2^c$,

$$\mathcal{I} : \mathcal{L}_0 \to \mathcal{M}_2^c.$$

Equation (20.6) implies that $\mathcal{I}$ is an isometry between $\mathcal{L}_0$ and its image $\mathcal{I}(\mathcal{L}_0) \subseteq \mathcal{M}_2^c$, with the norms $||\cdot||_{\mathcal{H}_n}$ and $||\cdot||_n$ respectively. Therefore, it is an isometry with respect to the metrics $d_{\mathcal{H}}$ and $d_{\mathcal{M}}$, that is

$$d_{\mathcal{M}}(I_t(X), I_t(Y)) = d_{\mathcal{H}}(X, Y)$$

for any $X, Y \in \mathcal{L}_0$. Since $\mathcal{L}_0$ is dense in $\mathcal{L}^*$ in the metric $d_{\mathcal{H}}$ (Lemma 20.13), and the space $\mathcal{M}_2^c$ is complete (Lemma 20.6), we can now extend the mapping $\mathcal{I}$ to an isometry between $\mathcal{L}^*$ (with the metric $d_{\mathcal{H}}$) and a subset of $\mathcal{M}_2^c$ (with the metric $d_{\mathcal{M}}$),

$$\mathcal{I} : \mathcal{L}^* \to \mathcal{M}_2^c.$$

Definition 20.15. *The stochastic integral of a process $X_t \in \mathcal{L}^*$ is the unique (up to indistinguishability) martingale $I_t(X) \in \mathcal{M}_2^c$ such that*

$$\lim_{Y \to X, Y \in \mathcal{L}_0} d_{\mathcal{M}}(I_t(X), I_t(Y)) = 0.$$

Given a pair of processes $X_t, Y_t \in \mathcal{L}^*$, we can find two sequences $X_t^n, Y_t^n \in \mathcal{L}_0$ such that $X_t^n \to X_t$ and $Y_t^n \to Y_t$ in $\mathcal{L}^*$. Then $aX_t^n + bY_t^n \to aX_t + bY_t$ in $\mathcal{L}^*$, which justifies (20.5) for any $X, Y \in \mathcal{L}^*$.

For $X_t \in \mathcal{L}_0$, we proved that

$$\mathrm{E}(I_t^2(X) - I_s^2(X)|\mathcal{F}_s) = \mathrm{E}(\int_s^t X_u^2 d\langle M\rangle_u|\mathcal{F}_s). \tag{20.7}$$

If $X_t \in \mathcal{L}^*$, we can find a sequence X_t^n such that $X_t^n \to X_t$ in $\mathcal{L}^*$. For any $A \in \mathcal{F}_s$,

$$\int_A (I_t^2(X) - I_s^2(X))d\mathrm{P} = \lim_{n\to\infty} \int_A (I_t^2(X^n) - I_s^2(X^n))d\mathrm{P} \tag{20.8}$$
$$= \lim_{n\to\infty} \int_A \int_s^t (X_u^n)^2 d\langle M\rangle_u = \int_A \int_s^t X_u^2 d\langle M\rangle_u.$$

This proves that (20.7) holds for all $X_t \in \mathcal{L}^*$. By Lemma 20.11, the process $\int_0^t X_u^2 d\langle M\rangle_u$ is $\mathcal{F}_t$-adapted. Thus, due to the uniqueness in the Doob-Meyer Decomposition, for all $X \in \mathcal{L}^*$,

$$\langle I(X)\rangle_t = \int_0^t X_u^2 d\langle M\rangle_u. \tag{20.9}$$

Remark 20.16. We shall also deal with stochastic integrals over a segment $[s, t]$, where $0 \le s \le t$. Namely, let a process X_u be defined for $u \in [s, t]$. We can consider the process $\widetilde{X}_u$ which is equal to X_u for $s \le u \le t$ and to zero for $u < s$ and $u > t$. If $\widetilde{X}_u \in \mathcal{L}^*$, we can define

$$\int_s^t X_u dM_u = \int_0^t \widetilde{X}_u dM_u.$$

Clearly, for $X_u \in \mathcal{L}^*$, $\int_s^t X_u dM_u = I_t(X) - I_s(X)$.

20.5 Further Properties of the Stochastic Integral

We start this section with a formula similar to (20.9), but which applies to the cross-variation of two stochastic integrals.

Lemma 20.17. *Let $M_t^1, M_t^2 \in \mathcal{M}_2^c$, $X_t^1 \in \mathcal{L}^*(M^1)$, and $X_t^2 \in \mathcal{L}^*(M^2)$. Then*

$$\langle I^{M^1}(X^1), I^{M^2}(X^2)\rangle_t = \int_0^t X_s^1 X_s^2 d\langle M^1, M^2\rangle_s, \quad t \ge 0, \quad \textit{almost surely.} \tag{20.10}$$

We only sketch the proof of this lemma, referring the reader to "Brownian Motion and Stochastic Calculus" by I. Karatzas and S. Shreve for a more detailed exposition. We need the Kunita-Watanabe Inequality, which states that under the assumptions of Lemma 20.17,

$$\int_0^t |X_s^1 X_s^2| dV_{[0,s]}^1(\langle M^1, M^2\rangle)$$

$$\le (\int_0^t (X_s^1)^2 d\langle M^1\rangle_s)^{1/2} (\int_0^t (X_s^2)^2 d\langle M^2\rangle_s)^{1/2}, \quad t \ge 0, \quad \text{almost surely,}$$

where $V_{[0,s]}^1(\langle M^1, M^2\rangle)$ is the first total variation of the process $\langle M^1, M^2\rangle_t$ over the interval $[0, s]$. In particular, the Kunita-Watanabe Inequality justifies the existence of the integral on the right-hand side of (20.10).

As we did with (20.7), we can show that for $0 \le s \le t < \infty$,

$$\mathrm{E}((I_t^{M^1}(X^1) - I_s^{M^1}(X^1))(I_t^{M^2}(X^2) - I_s^{M^2}(X^2))|\mathcal{F}_s)$$

$$= \mathrm{E}(\int_s^t X_u^1 X_u^2 d\langle M^1 M^2\rangle_u |\mathcal{F}_s)$$

for simple processes $X_t^1, X_t^2 \in \mathcal{L}_0$. This implies that (20.10) holds for simple processes X_t^1 and X_t^2. If $X_t^1 \in \mathcal{L}^*(M^1)$, $X_t^2 \in \mathcal{L}^*(M^2)$, then they can be approximated by simple processes as in the proof of (20.9). The transition from the statement for simple processes to (20.10) can be justified using the Kunita-Watanabe Inequality.

The following lemma will be used in the next section to define the stochastic integral with respect to a local martingale.

Lemma 20.18. *Let $M_t^1, M_t^2 \in \mathcal{M}_2^c$ (with the same filtration), $X_t^1 \in \mathcal{L}^*(M^1)$, and $X_t^2 \in \mathcal{L}^*(M^2)$. Let τ be a stopping time such that*

$$M_{t\wedge\tau}^1 = M_{t\wedge\tau}^2, \quad X_{t\wedge\tau}^1 = X_{t\wedge\tau}^2 \quad \text{for } 0 \le t < \infty \quad \text{almost surely.}$$

Then $I_{t\wedge\tau}^{M^1}(X^1) = I_{t\wedge\tau}^{M^2}(X^2)$ for $0 \le t < \infty$ almost surely.

Proof. Let $Y_t = X_{t\wedge\tau}^1 = X_{t\wedge\tau}^2$ and $N_t = M_{t\wedge\tau}^1 = M_{t\wedge\tau}^2$. Take an arbitrary $t \ge 0$. By the formula for cross-variation of two integrals,

$$\langle I^{M^i}(X^i), I^{M^j}(X^j)\rangle_{t\wedge\tau} = \int_0^{t\wedge\tau} X_s^i X_s^j d\langle M^i, M^j\rangle_s = \int_0^t Y_s^2 d\langle N\rangle_s,$$

where $1 \le i, j \le 2$. Therefore,

$$\begin{aligned} &\langle I^{M^1}(X^1) - I^{M^2}(X^2)\rangle_{t\wedge\tau} \\ &= \langle I^{M^1}(X^1)\rangle_{t\wedge\tau} + \langle I^{M^2}(X^2)\rangle_{t\wedge\tau} - 2\langle I^{M^1}(X^1), I^{M^2}(X^2)\rangle_{t\wedge\tau} = 0. \end{aligned}$$

Lemma 20.3 now implies that $I_s^{M^1}(X^1) = I_s^{M^2}(X^2)$ for all $0 \le s \le t \wedge \tau$ almost surely. Since t was arbitrary, $I_s^{M^1}(X^1) = I_s^{M^2}(X^2)$ for $0 \le s < \tau$ almost surely, which is equivalent to the desired result. □

The next lemma will be useful when applying the Ito formula (to be defined later in this chapter) to stochastic integrals.

Lemma 20.19. *Let $M_t \in \mathcal{M}_2^c$, $Y_t \in \mathcal{L}^*(M)$, and $X_t \in \mathcal{L}^*(I^M(Y))$. Then $X_tY_t \in \mathcal{L}^*(M)$ and*

$$\int_0^t X_s d\left(\int_0^s Y_u dM_u\right) = \int_0^t X_s Y_s dM_s. \tag{20.11}$$

Proof. Since $\langle I^M(Y)\rangle_t = \int_0^t Y_s^2 d\langle M\rangle_s$, we have

$$\mathrm{E}\int_0^t X_s^2 Y_s^2 d\langle M\rangle_s = \mathrm{E}\int_0^t X_s^2 d\langle I^M(Y)\rangle_s < \infty,$$

which shows that $X_tY_t \in \mathcal{L}^*(M)$. Let us examine the quadratic variation of the difference between the two sides of (20.11). By the formula for cross-variation of two integrals,

$$\begin{aligned} &\langle I^{I^M(Y)}(X) - I^M(XY)\rangle_t \\ &= \langle I^{I^M(Y)}(X)\rangle_t + \langle I^M(XY)\rangle_t - 2\langle I^{I^M(Y)}(X), I^M(XY)\rangle_t \\ &= \int_0^t X_s^2 d\langle I^M(Y)\rangle_s + \int_0^t X_s^2 Y_s^2 d\langle M\rangle_s - 2\int_0^t X_s^2 Y_s d\langle I^M(Y), M\rangle_s \\ &= \int_0^t X_s^2 Y_s^2 d\langle M\rangle_s + \int_0^t X_s^2 Y_s^2 d\langle M\rangle_s - 2\int_0^t X_s^2 Y_s^2 d\langle M\rangle_s = 0. \end{aligned}$$

By Lemma 20.3, (20.11) holds. □

20.6 Local Martingales

In this section we define the stochastic integral with respect to continuous local martingales.

Definition 20.20. *Let X_t, $t \in \mathbb{R}^+$, be a process adapted to a filtration $\mathcal{F}_t$. Then $(X_t, \mathcal{F}_t)$ is called a local martingale if there is a non-decreasing sequence of stopping times $\tau_n : \Omega \to [0, \infty]$ such that $\lim_{n\to\infty} \tau_n = \infty$ almost surely, and the process $(X_{t\wedge\tau_n}, \mathcal{F}_t)$ is a martingale for each n.*

This method of introducing a non-decreasing sequence of stopping times, which convert a local martingale into a martingale, is called localization.

The space of equivalence classes of local martingales whose sample paths are continuous almost surely and which satisfy $X_0 = 0$ almost surely will be denoted by $\mathcal{M}^{c,\text{loc}}$. It is easy to see that $\mathcal{M}^{c,\text{loc}}$ is a vector space (see Problem 3). It is also important to note that a local martingale may be integrable and yet fail to be a martingale (see Problem 4).

Now let us define the quadratic variation of a continuous local martingale $(X_t, \mathcal{F}_t) \in \mathcal{M}^{c,\text{loc}}$. We introduce the notation $X_t^{(n)} = X_{t\wedge\tau_n}$. Then, for $m \le n$, as in the example before Lemma 20.3,

$$\langle X^{(m)} \rangle_t = \langle X^{(n)} \rangle_{t\wedge\tau_m}.$$

This shows that $\langle X^{(m)} \rangle_t$ and $\langle X^{(n)} \rangle_t$ agree on the interval $0 \le t \le \tau_m(\omega)$ for almost all ω. Since $\tau_m \to \infty$ almost surely, we can define the limit $\langle X \rangle_t = \lim_{m\to\infty} \langle X^{(m)} \rangle_t$, which is a non-decreasing adapted process whose sample paths are continuous almost surely. The process $\langle X \rangle_t$ is called the quadratic variation of the local martingale X_t. This is justified by the fact that

$$(X^2 - \langle X \rangle)_{t\wedge\tau_n} = (X_t^{(n)})^2 - \langle X^{(n)} \rangle_t \in \mathcal{M}_2^c.$$

That is, $X_t^2 - \langle X \rangle_t$ is a local martingale. Let us show that the process $\langle X \rangle_t$ does not depend on the choice of the sequence of stopping times τ_n.

Lemma 20.21. *Let $X_t \in \mathcal{M}^{c,\text{loc}}$. There exists a unique (up to indistinguishability) non-decreasing adapted continuous process Y_t such that $Y_0 = 0$ almost surely and $X_t^2 - Y_t \in \mathcal{M}^{c,\text{loc}}$.*

Proof. The existence part was demonstrated above. Let us suppose that there are two processes Y_t^1 and Y_t^2 with the desired properties. Then $M_t = Y_t^1 - Y_t^2$ belongs to $\mathcal{M}^{c,\text{loc}}$ (since $\mathcal{M}^{c,\text{loc}}$ is a vector space) and is a process of bounded variation. Let τ_n be a non-decreasing sequence of stopping times which tend to infinity, such that $M_t^{(n)} = M_{t\wedge\tau_n}$ is a martingale for each n. Then $M_t^{(n)}$ is also a process of bounded variation. By Corollary 20.8, $M_t^{(n)} = 0$ for all t almost surely. Since $\tau_n \to \infty$, this implies that Y_t^1 and Y_t^2 are indistinguishable. □

The cross-variation of two local martingales can be defined by the same formula (20.1) as in the square-integrable case. It is also not difficult to see that $\langle X, Y\rangle_t$ is the unique (up to indistinguishability) adapted continuous process of bounded variation, such that $\langle X, Y\rangle_0 = 0$ almost surely, and $X_t Y_t - \langle X, Y\rangle_t \in \mathcal{M}^{c,\mathrm{loc}}$.

Let us now define the stochastic integral with respect to a continuous local martingale $M_t \in \mathcal{M}^{c,\mathrm{loc}}$. We can also extend the class of integrands. Namely, we shall say that $X_t \in \mathcal{P}^*$ if X_t is a progressively measurable process such that for every $0 \le t < \infty$,

$$\int_0^t X_s^2(\omega) d\langle M\rangle_s(\omega) < \infty \text{ almost surely.}$$

More precisely, we can view $\mathcal{P}^*$ as the set of equivalence classes of such processes, with two elements X_t^1 and X_t^2 representing the same class if and only if $\int_0^t (X_t^1 - X_t^2)^2 d\langle M\rangle_s = 0$ almost surely for every t.

Let us consider a sequence of stopping times $\tau_n : \Omega \to [0,\infty]$ with the following properties:

1. The sequence τ_n is non-decreasing and $\lim_{n\to\infty} \tau_n = \infty$ almost surely.
2. For each n, the process $M_t^{(n)} = M_{t\wedge\tau_n}$ is in $\mathcal{M}_2^c$.
3. For each n, the process $X_t^{(n)} = X_{t\wedge\tau_n}$ is in $\mathcal{L}^*(M^{(n)})$.

For example, such a sequence can be constructed as follows. Let τ_n^1 be a non-decreasing sequence such that $\lim_{n\to\infty} \tau_n^1 = \infty$ almost surely and the process $(X_{t\wedge\tau_n^1}, \mathcal{F}_t)$ is a martingale for each n. Define

$$\tau_n^2(\omega) = \inf\{t : \int_0^t X_s^2(\omega) d\langle M\rangle_s(\omega) = n\},$$

where the infimum of an empty set is equal to $+\infty$. It is clear that the sequence of stopping times $\tau_n = \tau_n^1 \wedge \tau_n^2$ has the properties (1)–(3).

Given a sequence τ_n with the above properties, a continuous local martingale $M_t \in \mathcal{M}^{c,\mathrm{loc}}$, and a process $X_t \in \mathcal{P}^*$, we can define

$$I_t^M(X) = \lim_{n\to\infty} I_t^{M^{(n)}}(X^{(n)}).$$

For almost all ω, the limit exists for all t. Indeed, by Lemma 20.18, almost surely,

$$I_t^{M^{(m)}}(X^{(m)}) = I_t^{M^{(n)}}(X^{(n)}), \quad 0 \le t \le \tau_m \wedge \tau_n.$$

Let us show that the limit does not depend on the choice of the sequence of stopping times, thus providing a correct definition of the integral with respect to a local martingale. If $\widetilde{\tau}_n$ and $\overline{\tau}_n$ are two sequences of stopping times with properties (1)–(3), and $\widetilde{M}_t^{(n)}$, $\widetilde{X}_t^{(n)}$, $\overline{M}_t^{(n)}$, and $\overline{X}_t^{(n)}$ are the corresponding processes, then

$$I_t^{\widetilde{M}^{(n)}}(\widetilde{X}^{(n)}) = I_t^{\overline{M}^{(n)}}(\overline{X}^{(n)}), \quad 0 \leq t \leq \widetilde{\tau}_n \wedge \overline{\tau}_n,$$

again by Lemma 20.18. Therefore, the limit in the definition of the integral $I_t^M(X)$ does not depend on the choice of the sequence of stopping times.

It is clear from the definition of the integral that $I_t^M(X) \in \mathcal{M}^{c,\mathrm{loc}}$, and that it is linear in the argument, that is, it satisfies (20.5) for any $X, Y \in \mathcal{P}^*$ and $a, b \in \mathbb{R}$. The formula for cross-variation of two integrals with respect to local martingales is the same as in the square-integrable case, as can be seen using localization. Namely, if $M_t, N_t \in \mathcal{M}^{c,\mathrm{loc}}$, $X_t \in \mathcal{P}^*(M)$, and $Y_t \in \mathcal{P}^*(N)$, then for almost all ω,

$$\langle I^M(X), I^N(Y)\rangle_t = \int_0^t X_s Y_s d\langle M, N\rangle_s, \quad 0 \leq t < \infty.$$

Similarly, by using localization, it is easy to see that (20.11) remains true if $M_t \in \mathcal{M}^{c,\mathrm{loc}}$, $Y_t \in \mathcal{P}^*(M)$, and $X_t \in \mathcal{P}^*(I^M(Y))$.

Remark 20.22. Let X_u, $s \leq u \leq t$, be such that the process $\widetilde{X}_u \in \mathcal{P}^*$, where $\widetilde{X}_u$ is equal to X_u for $s \leq u \leq t$, and to zero otherwise. In this case we can define $\int_s^t X_u dM_u = \int_0^t \widetilde{X}_u dM_u$ as in the case of square-integrable martingales.

20.7 Ito Formula

In this section we shall prove a formula which may be viewed as the analogue of the Fundamental Theorem of Calculus, but is now applied to martingale-type processes with unbounded first variation.

Definition 20.23. *Let X_t, $t \in \mathbb{R}^+$, be a process adapted to a filtration $\mathcal{F}_t$. Then $(X_t, \mathcal{F}_t)$ is a continuous semimartingale if X_t can be represented as*

$$X_t = X_0 + M_t + A_t, \tag{20.12}$$

where $M_t \in \mathcal{M}^{c,\mathrm{loc}}$, A_t is a continuous process adapted to the same filtration such that the total variation of A_t on each finite interval is bounded almost surely, and $A_0 = 0$ almost surely.

Theorem 20.24 (Ito Formula). *Let $f \in C^2(\mathbb{R})$ and let $(X_t, \mathcal{F}_t)$ be a continuous semimartingale as in (20.12). Then, for any $t \geq 0$, the equality*

$$f(X_t) = f(X_0) + \int_0^t f'(X_s)dM_s + \int_0^t f'(X_s)dA_s + \frac{1}{2}\int_0^t f''(X_s)d\langle M\rangle_s \tag{20.13}$$

holds almost surely.

Remark 20.25. The first integral on the right-hand side is a stochastic integral, while the other two integrals must be understood in the Lebesgue-Stieltjes sense. Since both sides are continuous functions of t for almost all ω, the processes on the left- and right-hand sides are indistinguishable.

Proof of Theorem 20.24. We shall prove the result under stronger assumptions. Namely, we shall assume that $M_t = W_t$ and that f is bounded together with its first and second derivatives. The proof in the general case is similar, but somewhat more technical. In particular, it requires the use of localization. Thus we assume that

$$X_t = X_0 + W_t + A_t,$$

and wish to prove that

$$f(X_t) = f(X_0) + \int_0^t f'(X_s)dW_s + \int_0^t f'(X_s)dA_s + \frac{1}{2}\int_0^t f''(X_s)ds. \quad (20.14)$$

Let $\sigma = \{t_0, t_1, \ldots, t_n\}$, $0 = t_0 \leq t_1 \leq \ldots \leq t_n = t$, be a partition of the interval $[0, t]$ into n subintervals. By the Taylor formula,

$$\begin{aligned} f(X_t) &= f(X_0) + \sum_{i=1}^n (f(X_{t_i}) - f(X_{t_{i-1}})) \\ &= f(X_0) + \sum_{i=1}^n f'(X_{t_{i-1}})(X_{t_i} - X_{t_{i-1}}) + \frac{1}{2}\sum_{i=1}^n f''(\xi_i)(X_{t_i} - X_{t_{i-1}})^2, \end{aligned} \quad (20.15)$$

where $\min(X_{t_{i-1}}, X_{t_i}) \leq \xi_i \leq \max(X_{t_{i-1}}, X_{t_i})$ is such that

$$f(X_{t_i}) - f(X_{t_{i-1}}) = f'(X_{t_{i-1}})(X_{t_i} - X_{t_{i-1}}) + \frac{1}{2} f''(\xi_i)(X_{t_i} - X_{t_{i-1}})^2.$$

Note that we can take $\xi_i = X_{t_{i-1}}$ if $X_{t_{i-1}} = X_{t_i}$. If $X_{t_{i-1}} \neq X_{t_i}$, we can solve the above equation for $f''(\xi_i)$, and therefore we may assume that $f''(\xi_i)$ is measurable.

Let $Y_s = f'(X_s)$, $0 \leq s \leq t$, and define the simple process Y_s^σ by

$$Y_s^\sigma = f'(X_0)\chi_{\{0\}}(s) + \sum_{i=1}^n f'(X_{t_{i-1}})\chi_{(t_{i-1}, t_i]}(s) \quad \text{for } 0 \leq s \leq t.$$

Note that $\lim_{\delta(\sigma)\to 0} Y_s^\sigma(\omega) = Y_s(\omega)$, where the convergence is uniform on $[0, t]$ for almost all ω since the process Y_s is continuous almost surely.

Let us examine the first sum on the right-hand side of (20.15),

$$\begin{aligned} &\sum_{i=1}^n f'(X_{t_{i-1}})(X_{t_i} - X_{t_{i-1}}) \\ &\quad = \sum_{i=1}^n f'(X_{t_{i-1}})(W_{t_i} - W_{t_{i-1}}) + \sum_{i=1}^n f'(X_{t_{i-1}})(A_{t_i} - A_{t_{i-1}}) \\ &\quad = \int_0^t Y_s^\sigma dW_s + \int_0^t Y_s^\sigma dA_s = S_1^\sigma + S_2^\sigma, \end{aligned}$$

where S_1^σ and S_2^σ denote the stochastic and the ordinary integral, respectively. Since

$$\mathrm{E}\left(\int_0^t (Y_s^\sigma - Y_s) dW_s\right)^2 = \mathrm{E}\int_0^t (Y_s^\sigma - Y_s)^2 ds \to 0,$$

we obtain

$$\lim_{\delta(\sigma)\to 0} S_1^\sigma = \lim_{\delta(\sigma)\to 0} \int_0^t Y_s^\sigma dW_s = \int_0^t Y_s dW_s \quad \text{in } L^2(\Omega, \mathcal{F}, \mathrm{P}).$$

We can apply the Lebesgue Dominated Convergence Theorem to the Lebesgue-Stieltjes integral (which is just a difference of two Lebesgue integrals) to obtain

$$\lim_{\delta(\sigma)\to 0} S_2^\sigma = \lim_{\delta(\sigma)\to 0} \int_0^t Y_s^\sigma dA_s = \int_0^t Y_s dA_s \quad \text{almost surely.}$$

Now let us examine the second sum on the right-hand side of (20.15):

$$\begin{aligned}\frac{1}{2}\sum_{i=1}^n f''(\xi_i)(X_{t_i} - X_{t_{i-1}})^2 &= \frac{1}{2}\sum_{i=1}^n f''(\xi_i)(W_{t_i} - W_{t_{i-1}})^2 \\ &\quad + \sum_{i=1}^n f''(\xi_i)(W_{t_i} - W_{t_{i-1}})(A_{t_i} - A_{t_{i-1}}) \qquad (20.16)\\ &\quad + \frac{1}{2}\sum_{i=1}^n f''(\xi_i)(A_{t_i} - A_{t_{i-1}})^2 = S_3^\sigma + S_4^\sigma + S_5^\sigma.\end{aligned}$$

The last two sums on the right-hand side of this formula tend to zero almost surely as $\delta(\sigma) \to 0$. Indeed,

$$\left|\sum_{i=1}^n f''(\xi_i)(W_{t_i} - W_{t_{i-1}})(A_{t_i} - A_{t_{i-1}}) + \frac{1}{2}\sum_{i=1}^n f''(\xi_i)(A_{t_i} - A_{t_{i-1}})^2\right|$$

$$\le \sup_{x\in\mathbb{R}} f''(x)\left(\max_{1\le i\le n}(|W_{t_i} - W_{t_{i-1}}|) + \frac{1}{2}\max_{1\le i\le n}(|A_{t_i} - A_{t_{i-1}}|)\right)\sum_{i=1}^n |A_{t_i} - A_{t_{i-1}}|,$$

which tends to zero almost surely since W_t and A_t are continuous and A_t has bounded variation.

It remains to deal with the first sum on the right-hand side of (20.16). Let us compare it with the sum

$$\widetilde{S}_3^\sigma = \frac{1}{2}\sum_{i=1}^n f''(X_{t_{i-1}})(W_{t_i} - W_{t_{i-1}})^2,$$

and show that the difference converges to zero in L^1. Indeed,

$$\mathrm{E}\left|\sum_{i=1}^n f''(\xi_i)(W_{t_i} - W_{t_{i-1}})^2 - \sum_{i=1}^n f''(X_{t_{i-1}})(W_{t_i} - W_{t_{i-1}})^2\right|$$

$$\leq (\mathrm{E}(\max_{1\leq i\leq n}(f''(\xi_i)-f''(X_{t_{i-1}}))^2)^{1/2}(\mathrm{E}(\sum_{i=1}^n(W_{t_i}-W_{t_{i-1}})^2)^2)^{1/2}.$$

The first factor here tends to zero since f'' is continuous and bounded. The second factor is bounded since

$$\mathrm{E}(\sum_{i=1}^n(W_{t_i}-W_{t_{i-1}})^2)^2 = 3\sum_{i=1}^n(t_i-t_{i-1})^2+\sum_{i\neq j}(t_i-t_{i-1})(t_j-t_{j-1})$$
$$\leq 3(\sum_{i=1}^n(t_i-t_{i-1}))(\sum_{j=1}^n(t_j-t_{j-1})) = 3t^2,$$

which shows that $(S_3^\sigma - \widetilde{S}_3^\sigma) \to 0$ in $L^1(\Omega,\mathcal{F},\mathrm{P})$ as $\delta(\sigma)\to 0$. Let us compare $\widetilde{S}_3^\sigma$ with the sum

$$\overline{S}_3^\sigma = \frac{1}{2}\sum_{i=1}^n f''(X_{t_{i-1}})(t_i-t_{i-1}),$$

and show that the difference converges to zero in L^2. Indeed, similarly to the proof of Lemma 18.24,

$$\mathrm{E}[\sum_{i=1}^n f''(X_{t_{i-1}})(W_{t_i}-W_{t_{i-1}})^2-\sum_{i=1}^n f''(X_{t_{i-1}})(t_i-t_{i-1})]^2$$
$$=\sum_{i=1}^n \mathrm{E}([f''(X_{t_{i-1}})^2][(W_{t_i}-W_{t_{i-1}})^2-(t_i-t_{i-1})]^2)$$
$$\leq \sup_{x\in\mathbb{R}}|f''(x)|^2(\sum_{i=1}^n \mathrm{E}(W_{t_i}-W_{t_{i-1}})^4+\sum_{i=1}^n(t_i-t_{i-1})^2)$$
$$=4\sup_{x\in\mathbb{R}}|f''(x)|^2\sum_{i=1}^n(t_i-t_{i-1})^2 \leq 4\sup_{x\in\mathbb{R}}|f''(x)|^2\max_{1\leq i\leq n}(t_i-t_{i-1})\sum_{i=1}^n(t_i-t_{i-1})$$
$$=4\sup_{x\in\mathbb{R}}|f''(x)|^2 t\delta(\sigma),$$

where the first equality is justified by

$$\mathrm{E}[f''(X_{t_{i-1}})((W_{t_i}-W_{t_{i-1}})^2-(t_i-t_{i-1}))$$
$$f''(X_{t_{j-1}})((W_{t_j}-W_{t_{j-1}})^2-(t_j-t_{j-1}))]$$
$$=\mathrm{E}[f''(X_{t_{i-1}})((W_{t_i}-W_{t_{i-1}})^2-(t_i-t_{i-1}))$$
$$\mathrm{E}(f''(X_{t_{j-1}})((W_{t_j}-W_{t_{j-1}})^2-(t_j-t_{j-1}))|\mathcal{F}_{j-1})] = 0 \quad \text{if} \quad i<j.$$

Thus, we see that $(\widetilde{S}_3^\sigma - \overline{S}_3^\sigma) \to 0$ in $L^2(\Omega,\mathcal{F},\mathrm{P})$ as $\delta(\sigma)\to 0$. It is also clear that

$$\lim_{\delta(\sigma)\to 0}\overline{S}_3^\sigma = \frac{1}{2}\int_0^t f''(X_s)ds \quad \text{almost surely.}$$

Let us return to formula (20.15), which we can now write as

$$f(X_t) = f(X_0) + S_1^\sigma + S_2^\sigma + (S_3^\sigma - \widetilde{S}_3^\sigma) + (\widetilde{S}_3^\sigma - \overline{S}_3^\sigma) + \overline{S}_3^\sigma + S_4^\sigma + S_5^\sigma .$$

Take a sequence $\sigma(n)$ with $\lim_{n\to\infty} \delta(\sigma(n)) = 0$. We saw that

$$\lim_{n\to\infty} S_1^{\sigma(n)} = \int_0^t f'(X_s)dW_s \ \text{ in } L^2(\Omega, \mathcal{F}, \mathrm{P}), \tag{20.17}$$

$$\lim_{n\to\infty} S_2^{\sigma(n)} = \int_0^t f'(X_s)dA_s \ \text{ almost surely}, \tag{20.18}$$

$$\lim_{n\to\infty} (S_3^{\sigma(n)} - \widetilde{S}_3^{\sigma(n)}) = 0 \ \text{ in } L^1(\Omega, \mathcal{F}, \mathrm{P}), \tag{20.19}$$

$$\lim_{n\to\infty} (\widetilde{S}_3^{\sigma(n)} - \overline{S}_3^{\sigma(n)}) = 0 \ \text{ in } L^2(\Omega, \mathcal{F}, \mathrm{P}), \tag{20.20}$$

$$\lim_{n\to\infty} \overline{S}_3^{\sigma(n)} = \frac{1}{2}\int_0^t f''(X_s)ds \ \text{ almost surely}, \tag{20.21}$$

$$\lim_{n\to\infty} S_4^{\sigma(n)} = \lim_{n\to\infty} S_5^{\sigma(n)} = 0 \ \text{ almost surely}. \tag{20.22}$$

We can replace the sequence $\sigma(n)$ by a subsequence for which all the equalities (20.17)–(20.22) hold almost surely. This justifies (20.14). □

Remark 20.26. The stochastic integral on the right-hand side of (20.13) belongs to $\mathcal{M}^{c,\mathrm{loc}}$, while the Lebesgue-Stieltjes integrals are continuous adapted processes with bounded variation. Therefore, the class of semimartingales is invariant under the composition with twice continuously differentiable functions.

Example. Let $f \in C^2(\mathbb{R})$, A_t and B_t be progressively measurable processes such that $\int_0^t A_s^2 ds < \infty$ and $\int_0^t |B_s| ds < \infty$ for all t almost surely, and X_t a semimartingale of the form

$$X_t = X_0 + \int_0^t A_s dW_s + \int_0^t B_s ds.$$

Applying the Ito formula, we obtain

$$f(X_t) = f(X_0) + \int_0^t f'(X_s)A_s dW_s + \int_0^t f'(X_s)B_s ds + \frac{1}{2}\int_0^t f''(X_s)A_s^2 ds,$$

where the relation $\int_0^t f'(X_s)d(\int_0^s A_u dW_u) = \int_0^t f'(X_s)A_s dW_s$ is justified by formula (20.11) applied to local martingales.

This is one of the most common applications of the Ito formula, particularly when the processes A_t and B_t can be represented as $A_t = \sigma(t, X_t)$ and $B_t = v(t, X_t)$ for some smooth functions σ and v, in which case X_t is called a diffusion process with time-dependent coefficients.

We state the following multi-dimensional version of the Ito formula, whose proof is very similar to that of Theorem 20.24.

Theorem 20.27. *Let $M_t = (M_t^1, \ldots, M_t^d)$ be a vector of continuous local martingales, that is $(M_t^i, \mathcal{F}_t)_{t\in\mathbb{R}^+}$ is a local martingale for each $1 \leq i \leq d$. Let $A_t = (A_t^1, \ldots, A_t^d)$ be a vector of continuous processes adapted to the same filtration such that the total variation of A_t^i on each finite interval is bounded almost surely, and $A_0^i = 0$ almost surely. Let $X_t = (X_t^1, \ldots, X_t^d)$ be a vector of adapted processes such that $X_t = X_0 + M_t + A_t$, and let $f \in C^{1,2}(\mathbb{R}^+ \times \mathbb{R}^d)$. Then, for any $t \geq 0$, the equality*

$$f(t, X_t) = f(0, X_0) + \sum_{i=1}^{d} \int_0^t \frac{\partial}{\partial x_i} f(s, X_s) dM_s^i + \sum_{i=1}^{d} \int_0^t \frac{\partial}{\partial x_i} f(s, X_s) dA_s^i$$
$$+ \frac{1}{2} \sum_{i,j=1}^{d} \int_0^t \frac{\partial^2}{\partial x_i \partial x_j} f(s, X_s) d\langle M^i, M^j \rangle_s + \int_0^t \frac{\partial}{\partial s} f(s, X_s) ds$$

holds almost surely.

Let us apply this theorem to a pair of processes $X_t^i = X_0^i + M_t^i + A_t^i$, $i = 1, 2$, and the function $f(x_1, x_2) = x_1 x_2$.

Corollary 20.28. *If $(X_t^1, \mathcal{F}_t)$ and $(X_t^2, \mathcal{F}_t)$ are continuous semimartingales, then*

$$X_t^1 X_t^2 = X_0^1 X_0^2 + \int_0^t X_s^1 dM_s^2 + \int_0^t X_s^1 dA_s^2$$
$$+ \int_0^t X_s^2 dM_s^1 + \int_0^t X_s^2 dA_s^1 + \langle M^1, M^2 \rangle_t.$$

Using the shorthand notation $\int_0^t Y_s dX_s = \int_0^t Y_s dM_s + \int_0^t Y_s dA_s$ for a process Y_s and a semimartingale X_s, we can rewrite the above formula as

$$\int_0^t X_s^1 dX_s^2 = X_t^1 X_t^2 - X_0^1 X_0^2 - \int_0^t X_s^2 dX_s^1 - \langle M^1, M^2 \rangle_s. \qquad (20.23)$$

This is the integration by parts formula for the Ito integral.

20.8 Problems

1. Prove that if X_t is a continuous non-random function, then the stochastic integral $I_t(X) = \int_0^t X_s dW_s$ is a Gaussian process.
2. Let W_t be a one-dimensional Brownian motion defined on a probability space $(\Omega, \mathcal{F}, \mathrm{P})$. Prove that there is a unique orthogonal random measure Z with values in $L^2(\Omega, \mathcal{F}, \mathrm{P})$ defined on a $([0,1], \mathcal{B}([0,1])$ such that $Z([s,t]) = W_t - W_s$ for $0 \leq s \leq t \leq 1$. Prove that

$$\int_0^1 \varphi(t)dZ(t) = \int_0^1 \varphi(t)dW_t$$

for any function φ that is continuous on $[0, 1]$.

3. Prove that if $X_t, Y_t \in \mathcal{M}^{c,\text{loc}}$, then $aX_t + bY_t \in \mathcal{M}^{c,\text{loc}}$ for any constants a and b.

4. Give an example of a local martingale which is integrable, yet fails to be a martingale.

5. Let W_t be a one-dimensional Brownian motion relative to a filtration $\mathcal{F}_t$. Let τ be a stopping time of $\mathcal{F}_t$ with $\mathrm{E}\tau < \infty$. Prove the Wald Identities

$$\mathrm{E}W_\tau = 0, \quad \mathrm{E}W_\tau^2 = \mathrm{E}\tau.$$

(Note that the Optional Sampling Theorem can not be applied directly since τ may be unbounded.)

6. Find the distribution function of the random variable $\int_0^1 W_t^n dW_t$, where W_t is a one-dimensional Brownian motion.