# LECTURE 6: THE ITÔ CALCULUS

## 1. INTRODUCTION: GEOMETRIC BROWNIAN MOTION

According to LÉVY's representation theorem, quoted at the beginning of the last lecture, every continuous–time martingale with continuous paths and finite quadratic variation is a *time–changed* Brownian motion. Thus, we expect discounted price processes in arbitrage–free, continuous–time markets to be *time–changed* Brownian motions. However, Brownian motion (with no time change) is not an appropriate model for the behavior of stock price processes. Why not? Among other reasons is this: If $\{W^x(t)\}_{t\geq 0}$ is a Brownian motion started at $x > 0$, then eventually the path $W(t)$ will drift below $0$.[1] But the share prices of stocks traded on major exchanges do not drop below $0$.

So what is a more appropriate model for stock price behavior? Consider the issue of *expected return* on investment. If an investor has \$100 to invest, he/she should, at the very least, expect the same return on the \$100 investment when the share price is \$10 as when the share price is \$5. (If the expected return were higher for \$5 shares than for \$10 shares, the shareholders would *split* the \$10 shares into twice as many \$5 shares, thus increasing their expected return by a mere accounting trick!) Similarly, the *uncertainty* about the return on a \$100 investment should be the same for \$5 shares as for \$10 shares. Thus, fluctuations in share price should be twice as great when the share price is twice as high, three times as great when the share price is three times as high, etc. This argument suggests that a better model for stock price behavior over time might be

$$(1) \qquad \frac{dS_t}{S_t} = \mu\, dt + \sigma\, dW_t,$$

where $\mu$ represents the instantaneous rate of return on a riskless asset, $\sigma$ represents the *volatility* of the stock, and $dW_t$ represents something like the infinitesimal change in a Brownian motion over the next instant of time. This is the stock price model used by BLACK and SCHOLES in their work on the pricing of call and put options — the stochastic processes $S_t$ governed by equation (1) are called *geometric Brownian motions*. Unfortunately, equation (1) cannot be interpreted as an ordinary differential equation, because Brownian paths $W_t$ are not differentiable. It was precisely for the purpose of dealing with differential equations incorporating stochastic differentials that ITÔ developed what is now called the **Itô calculus**.

## 2. THE ITÔ INTEGRAL

In ordinary calculus, the (Riemann) integral is defined by a limiting procedure. One first defines the integral of a *step function*, in such a way that the integral represents the "area beneath the graph". Then one extends the definition to a larger class of functions (the *Riemann–integrable* functions) by approximation: the integral of a function $f$ is defined to be the limit of the integrals of step functions which converge (in a certain sense) to $f$.

Thus, it is natural that the Itô integral is also defined by an approximation procedure. In the Itô theory, step functions are replaced by *simple processes*, which are nothing more than random step functions. The integral is then extended to larger classes of processes by approximation. For

---

[1] We even know the distribution of the amount of time that elapses before the first passage of $W(t)$ to $0$.

our purposes, the details of the approximation procedure are uninteresting, and we shall (for the most part) omit them. However, it is important to understand how the integral is defined for the simple processes, because nearly all of the basic properties of the Itô integral are established by first considering simple processes, then extending to more general integrands " by approximation".

2.1. **Integration of simple processes.** Let $\{W_t\}_{t\geq 0}$ be a standard Brownian motion, and let $\{\mathcal{F}_t\}_{t\geq 0}$ be its filtration. Recall that a stochastic process $\{\theta_t\}_{t\geq 0}$ is said to be *adapted* to the filtration $\{\mathcal{F}_t\}_{t\geq 0}$ if, for every $s \geq 0$, the random variable $\theta_s$ is measurable relative to the $\sigma-$algebra $\mathcal{F}_s$. (This means, roughly, that for each $s \geq 0$ the value of the random variable $\theta_s$, although random, is completely determined by the Brownian path $\{W_t\}_{0\leq t\leq s}$ up to time $s$.) An adapted process $\{\theta_t\}_{t\geq 0}$ is said to be *simple* if it is a random step function: This means that, for some finite sequence of times $0 < t_1 < t_2 < \cdots < t_m < \infty$ and random variables $\xi_0, \xi_1, \ldots, \xi_{m-1}$ such that $\xi_j$ is measurable relative to $\mathcal{F}_{t_j}$,

$$(2) \qquad \theta(s) = \sum_{j=0}^{m-1} \xi_j \mathbf{1}_{(t_j, t_{j+1}]}(s).$$

**Definition 1.** *For a simple process $\{\theta_t\}_{t\geq 0}$ satisfying equation (2), define the* Itô *integral $\int \theta_s \, dW_s$ as follows:*

$$(3) \qquad \int \theta_s \, dW_s = \sum_{j=0}^{m-1} \xi_j (W(t_{j+1}) - W(t_j))$$

Why is this a reasonable definition? The random step function $\theta(s)$ takes the (random) value $\xi_j$ between times $t_{j-1}$ and $t_j$. Thus, for all times $s \in (t_{j-1}, t_j]$, the random infinitesimal increments $\theta_s \, dW_s$ should be $\xi_j$ times as large as those of the Brownian motion; when one adds up these infinitesimal increments, one gets $\xi_j$ times the total increment of the Brownian motion over this time period.

Notice that the random "fluctuation rates" $\xi_j$ in the sum (3) are independent of the Brownian increments $W(t_{j+1}) - W(t_j)$ that they multiply. This is a consequence of the independent increments property of Brownian motion: $\xi_j$, being measurable relative to $\mathcal{F}_{t_j}$, is a function of the Brownian path up to time $t_j$, which is independent of all future increments. This independence property is behind the following calculation, which is of fundamental importance.

**Proposition 1.** *The mean and variance of the stochastic integral $\int \theta_s \, dW_s$ are*

$$(4) \qquad E\left(\int \theta_s \, dW_s\right) = 0$$

$$(5) \qquad E\left(\int \theta_s \, dW_s\right)^2 = \int E\theta_s^2 \, ds$$

*Proof.* **Exercise!** □

Two other properties of the Itô integral should be noted. First, it is *linear* in the integrand, that is, if $\{\theta_s\}$ and $\{\eta_s\}$ are two simple processes and $a, b \in \mathbb{R}$ are scalars, then

$$(6) \qquad \int (a\theta_s + b\eta_s) \, dW_s = a \int \theta_s \, dW_s + b \int \eta_s \, dW_s.$$

Second, if $\theta(s)$ is a simple function satisfying equation (2), then $\int \theta_s \, dW_s$, as defined by (3) is measurable relative to the $\sigma-$algebra $\mathcal{F}_{t_m}$.

2.2. **Integration of $\mathcal{H}^2$ Processes.** One extends the definition of the Itô integral to other integrands by approximation: one approximates the integrand $\theta_s$ by a simple process, then uses the Itô integral of the simple process, defined above, as an approximation to the Itô integral of $\{\theta_s\}$. The most important class of integrands are the $\mathcal{H}^2$ processes. An adapted process $\{\theta_s\}$ is a member of the class $\mathcal{H}^p$ (where $1 \le p < \infty$) if

$$(7) \qquad \int E|\theta_s|^p \, ds < \infty.$$

**Proposition 2.** *Every adapted process of class $\mathcal{H}^p$ may be approximated arbitrarily closely in $\mathcal{H}^p$ distance by simple processes. For an $\mathcal{H}^2$ process $\{\theta_s\}$, this means that there is a sequence of simple processes $\{\theta_s^{(n)}\}$ such that*

$$(8) \qquad \lim_{n \to \infty} \int E(\theta_s - \theta_s^{(n)})^2 \, ds = 0.$$

*Proof.* See KARATZAS & SHREVE, section 3.2, for the rather long and arduous proof. □

**Definition 2.** *Let $\{\theta_s\}$ be a member of the class $\mathcal{H}_2$, and let $\{\theta_s^{(n)}\}$ be any sequence of simple processes such that (8) holds. Define*

$$(9) \qquad \int \theta_s \, dW_s = \lim_{n \to \infty} \int \theta_s^{(n)} \, dW_s$$

**Proposition 3.** *The limit in (9) exists (in the $L^2$ sense) and is independent of the sequence of approximating simple processes.*

The proof is outlined below. The proposition guarantees that the Itô integral is well-defined; moreover, it asserts that *any* approximating sequence of simple processes may be used. In those (rare) instances where one calculates an Itô integral directly, using the definition above (see section 3 for an example), one usually uses the most obvious approximation.

*Proof of Proposition 3.* The proof is based on proposition 1 above, and standard elementary facts about the space $L^2$ of random variables with finite second moments. The argument (in brief) for the existence of the limit is as follows: Suppose that $\{\theta_s^{(n)}\}$ is a sequence of simple processes such that (8) holds. Then for any $\varepsilon > 0$, there exists an integer $n_\varepsilon$ sufficiently large that, for all $n \ge n_\varepsilon$,

$$\int E|\theta_s^{(n)} - \theta_s|^2 \, ds < \varepsilon.$$

Hence, by the triangle inequality for the $L^2$−metric, if $n, m \ge n_\varepsilon$ then

$$\left( \int E|\theta_s^{(n)} - \theta_s^{(m)}|^2 \, ds \right)^{1/2} \le \left( \int E|\theta_s^{(n)} - \theta_s|^2 \, ds \right)^{1/2} + \left( \int E|\theta_s^{(m)} - \theta_s|^2 \, ds \right)^{1/2} < 2\sqrt{\varepsilon}.$$

But Proposition 1, together with the linearity of the integral for simple processes, implies that

$$E|\int \theta_s^{(n)} \, dW_s - \int \theta_s^{(m)} \, dW_s|^2 = \int E|\theta_s^{(n)} - \theta_s^{(m)}|^2 \, ds < 4\varepsilon.$$

This shows that the sequence $\int \theta_s^{(n)} \, dW_s$ of stochastic integrals is a Cauchy sequence of random variables (in $L^2$). Since Cauchy sequences in $L^2$ have limits[2], it follows that

$$\lim_{n \to \infty} \int \theta_s^{(n)} \, dW_s$$

exists. A similar argument shows that the limit is the same for *all* approximating sequences of simple processes.

**Proposition 4.** *For any process $\{\theta_s\}$ of class $\mathcal{H}_2$, the mean and variance of the stochastic integral $\int \theta_s \, dW_s$ are given by equations (4)–(5) of Proposition 1. Moreover, the stochastic integral is linear, in the sense of equation (6) above.*

*Proof.* One deduces this from the corresponding facts about integrals of simple processes by an approximation argument. Details are omitted (but are not hard). $\square$

2.3. **Limits of Integration.** The implicit limits of integration in the stochastic integral $\int \theta_s \, dW_s$ are 0 and $\infty$. It is useful to define stochastic integrals with other limits of integration. This is done as follows:

$$(10) \qquad \int_a^b \theta_s \, dW_s := \int \theta_s^{(a,b]} \, dW_s$$

where $\{\theta_s^{(a,b]}\}$ is the adapted $\mathcal{H}^2$ function defined by

$$\theta_s^{(a,b]} = \theta_s \qquad \text{if } a < s \le b$$
$$= 0 \qquad \text{otherwise.}$$

**Proposition 5.** *Let $\{\theta_s\}$ be an adapted process of class $\mathcal{H}^2$, and set $I_t = \int_0^t \theta_s \, dW_s$. Then the stochastic process $\{I_t\}_{t \ge 0}$ is a martingale with continuous paths[3], and*

$$(11) \qquad I_t = E\left( \int \theta_s \, dW_s \,\middle|\, \mathcal{F}_t \right).$$

*Proof.* If $\{\theta_t\}$ is a simple process, then the martingale property may be proved directly from the definition (3), using basic properties of conditional expectation and the independent increments property of Brownian motion. (**Exercise:** Do this!) One then deduces the general case by a routine limiting argument. Continuity of paths (for an appropriate version) is harder to prove – we omit this. $\square$

---

[2]The completeness of the metric space $L^2$ is called the *Riesz-Fisher* theorem. See, for example, ROYDEN, *Real Analysis*, Ch. 6, or RUDIN, *Real & Complex Analysis*, Ch. 3.

[3]More precisely, there is a version with continuous paths. Because this will not be of importance in the subsequent developments, we shall not discuss the mathematical niceties here. As usual, see KARATZAS & SHREVE for the sordid details.

2.4. **Further Extensions of the Itô Integral.** It is occasionally necessary to integrate adapted processes that do not belong to the class $\mathcal{H}^2$. This may be accomplished by various other means of approximation. One important class of integrands is the set of adapted processes $\{\theta(s)\}_{s \geq 0}$ satisfying

(12)
$$P\{\int_0^\infty \theta(s)^2 \, ds < \infty\} = 1;$$

we shall refer to this class as $\mathcal{H}_{\mathrm{LOC}}^2$, for "locally" $\mathcal{H}^2$. The classes $\mathcal{H}_{\mathrm{LOC}}^p$, for $p \geq 1$, are defined similarly. Among the naturally occurring adapted processes $\{\theta(s)\}_{s \geq 0}$ in the class $\mathcal{H}_{\mathrm{LOC}}^2$ that are not (at least not necessarily) in the class $\mathcal{H}^2$ are those of the form

$$\theta(t) = u(W_t, t),$$

where $u(x, t)$ is any function of $x \in \mathbb{R}$ and $t \geq 0$ that is jointly continuous in its two arguments $t, x$. Such processes occur regularly in stochastic calculus and its applications.

The Itô integral is extended from $\mathcal{H}^2$ integrands to $\mathcal{H}_{\mathrm{LOC}}^2$ integrands by a technique called (fittingly enough) *localization*. This technique consists of successive truncations at suitable stopping times: Let $\{\theta(s)\}_{s \geq 0}$ be an element of the class $\mathcal{H}_{\mathrm{LOC}}^2$, and for each integer $n \geq 1$ define a stopping time $T_n \leq \infty$ by

$$T_n = \min\{t \geq 0 : \int_0^t \theta(s)^2 \, ds = n\},$$

with the convention that $T_n = \infty$ if there is no such $t$, and define an adapted process $\theta^{(n)}(t)$

$$\theta^{(n)}(t) = \theta(t) \qquad \text{for } t \leq T_n \text{ and}$$
$$\theta^{(n)}(t) = 0 \qquad \text{for } t > T_n$$

Observe that the hypothesis (12) implies that, with probability one, $T_n \to \infty$. Consequently, the sequence of processes $\theta^{(n)}(s)$ converge to $\theta(s)$ in the following sense: for each $t < \infty$, with probability one, $\theta^{(n)}(t) = \theta(t)$ for all sufficiently large $n$. Notice also that each process $\theta^{(n)}(s)$ is a member of the class $\mathcal{H}^2$, because

$$E \int_0^\infty \theta^{(n)}(s)^2 \, ds = E \int_0^{T_n} \theta(s)^2 \, ds \leq En = n < \infty.$$

Thus, the Itô integrals $\int \theta^{(n)}(s) \, dW_s$ are well-defined. It may be shown that, with probability one, these Itô integrals converge as $n \to \infty$ to a finite (random) limit. Thus, we may define

(13)
$$\int \theta(s) \, dW_s = \lim_{n \to \infty} \int \theta^{(n)}(s) \, dW_s.$$

Caution must be exercised in using the Itô integral defined by equation (13) for integrands in the class $\mathcal{H}_{\mathrm{LOC}}^2$. Although the Itô integrals $\int \theta^{(n)}(s) \, dW_s$ converge with probability one, they need not converge in mean square ($L^2$), nor need the limit random variable $\int \theta(s) \, dW_s$ have finite expectation. Thus, the relation (11) need not be valid for integrands of the class $\mathcal{H}_{\mathrm{LOC}}^2$; furthermore, the process

$$I_t := \int \theta(s) \mathbf{1}_{[0,t]}(s) \, dW_s$$

may fail to be a martingale. In general, the best that can be said is that, for each $n \geq 1$, the process

$$I_t^{(n)} := \int \theta(s) \mathbf{1}_{[0, t \wedge T_n]}(s) \, dW_s$$

is a martingale.

## 3. An Example: $\int_0^1 W\, dW$

There are tools for calculating stochastic integrals that usually make it unnecessary to use Definition 2 directly. The most useful of these, the *Itô formula*, will be discussed in the following sections. It is instructive, however, to do one explicit calculation using only the definition 2. This calculation will show (i) that the Fundamental Theorem of Calculus does not hold for Itô integrals; and (ii) the central importance of the Quadratic Variation formula in the Itô calculus.

Consider the process

$$(14) \qquad\qquad \theta_s = W_s \mathbf{1}\{s \le 1\}.$$

This process is obviously adapted to the natural filtration. It is also in the class $\mathcal{H}^2$, because

$$\int_0^\infty E\theta_s^2\, ds = \int_0^1 EW_s^2\, ds = \int_0^1 s\, ds = \frac{1}{2} < \infty.$$

The obvious way to approximate $\theta_s$ is by using the simple function $\theta_s^{(n)}$ whose jumps are at the dyadic rationals $1/2^n, 2/2^n, 3/2^n, \ldots$, and whose value in the interval $[k/2^n, (k+1)/2^n)$ is $\theta(k/2^n)$: that is,

$$\theta_s^{(n)} = \sum_{k=0}^{2^n} \theta(k/2^n) \mathbf{1}_{[k/2^n,(k+1)/2^n)}(s).$$

**Lemma 1.** $\lim_{n\to\infty} \int_0^\infty E(\theta_s - \theta_s^{(n)})^2\, ds = 0.$

*Proof.* Since the simple process $\theta_s^{(n)}$ takes the value $\theta(k/2^n)$ for all $s$ between $k/2^n$ and $(k+1)/2^n$, we have

$$
\begin{aligned}
\int_0^\infty E(\theta_s - \theta_s^{(n)})^2\, ds &= \sum_{k=0}^{2^n-1} \int_{k/2^n}^{(k+1)/2^n} E(\theta_s - \theta_s^{(n)})^2\, ds \\
&= \sum_{k=0}^{2^n-1} \int_{k/2^n}^{(k+1)/2^n} E(W_s - W_{k/2^n})^2\, ds \\
&= \sum_{k=0}^{2^n-1} \int_{k/2^n}^{(k+1)/2^n} (s - (k/2^n))\, ds \\
&\le \sum_{k=0}^{2^n-1} 2^{-2n} = 2^n/2^{2n} \longrightarrow 0
\end{aligned}
$$

$\square$

Definition 2 now implies that the stochastic integral $\int \theta_s\, dW_s$ is the limit of the stochastic integrals $\int \theta_s^{(n)}\, dW_s$. Since $\theta_s^{(n)}$ is a simple process, its stochastic integral is defined to be

$$\int \theta_s^{(n)}\, dW_s = \sum_{k=0}^{2^n-1} W_{k/2^n}(W_{(k+1)/2^n} - W_{k/2^n}).$$

To evaluate this sum, we use the technique of "summation by parts" (the discrete analogue of integration by parts). Here, the technique takes the form of observing that the sum can be modified

slightly to give a sum that "telescopes":

$$W_1^2 = \sum_{k=0}^{2^n-1} (W_{(k+1)/2^n}^2 - W_{k/2^n}^2)$$

$$= \sum_{k=0}^{2^n-1} (W_{(k+1)/2^n} - W_{k/2^n})(W_{(k+1)/2^n} + W_{k/2^n})$$

$$= \sum_{k=0}^{2^n-1} (W_{(k+1)/2^n} - W_{k/2^n})(W_{k/2^n} + W_{k/2^n})$$

$$+ \sum_{k=0}^{2^n-1} (W_{(k+1)/2^n} - W_{k/2^n})(W_{(k+1)/2^n} - W_{k/2^n})$$

$$= 2\sum_{k=0}^{2^n-1} W_{k/2^n}(W_{(k+1)/2^n} - W_{k/2^n})$$

$$+ \sum_{k=0}^{2^n-1} (W_{(k+1)/2^n} - W_{k/2^n})^2$$

The first sum on the right side is $2\int \theta_s^{(n)}\, dW_s$, and so converges to $2\int_0^1 W_s\, dW_s$ as $n \to \infty$. The second sum is the same sum that occurs in the Quadratic Variation Formula (Lecture 5), and so converges, as $n \to \infty$, to 1. Therefore,

(15)
$$\boxed{\int_0^1 W_s\, dW_s = \frac{1}{2}(W_1^2 - 1).}$$

Note that if the Itô integral obeyed the Fundamental Theorem of Calculus, then the value of the integral would be

$$\int_0^t W_s\, dW_s = \int_0^t W(s)W'(s)\, ds = \left.\frac{W_s^2}{2}\right|_0^t = \frac{W_t^2}{2}$$

Thus, formula (15) shows that the Itô calculus is fundamentally different than ordinary calculus.

## 4. Itô's Formula

The cornerstone of stochastic calculus is the *Itô formula*. Like the Fundamental Theorem of Calculus, the Itô formula takes many different forms, and holds at several levels of generality. The simplest version is as follows:

**Theorem 1.** *Let $u(x,t)$ be a function of $x \in \mathbb{R}$ and $t \geq 0$ that is twice continuously differentiable in $x$ and once continuously differentiable in $t$, and let $W_t$ be a Wiener process (Brownian motion). Denote by $u_t, u_x$, and $u_x x$ the first and second partial derivatives of $u$ with respect to the variables $t$ and $x$. Then*

(16)
$$u(W_t, t) - u(0,0) = \int_0^t u_x(W_s, s)\, dW_s + \int_0^t u_t(W_s, s)\, ds + \frac{1}{2}\int_0^t u_{xx}(W_s, s)\, ds.$$

A sketch of the proof will be given in section 8, and a more general form of the theorem will be discussed in section 5. In a later lecture, we shall see that there is also a muiltivariable version of the theorem.

**Example 1:** Let

(17)
$$u(x,t) = x^2 - t; \qquad \text{then}$$
$$u_t(x,t) = -1;$$
$$u_x(x,t) = 2x; \qquad \text{and}$$
$$u_{xx}(x,t) = 2.$$

Consequently, by the Itô formula (note that $u_t + (1/2)u_{xx} = 0$),

(18)
$$W_t^2 - t = 2\int_0^t W_s \, dW_s,$$

as was shown directly in the previous section.

**Example 2:** Let $\alpha, \beta \in \mathbb{R}$ be fixed nonrandom scalars. Let

(19)
$$u(x,t) = \exp\{\alpha x + \beta t\}; \qquad \text{then}$$
$$u_t(x,t) = \beta u(x,t),$$
$$u_x(x,t) = \alpha u(x,t), \qquad \text{and}$$
$$u_{xx}(x,t) = \alpha^2 u(x,t).$$

Thus, the Itô formula implies

(20)
$$u(W_t,t) = 1 + \alpha \int_0^t u(W_t,t) \, dW_t + (\beta + \frac{1}{2}\alpha^2) \int_0^t u(W_t,t) \, dt.$$

Set $S_t = u(W_t,t)$; then equation (20), in differential form, becomes

(21)
$$\frac{dS_t}{S_t} = \alpha dW_t + (\beta + \frac{1}{2}\alpha^2)dt.$$

In particular, if $\alpha = \sigma$ and $\beta = \mu - \sigma^2/2$ then $S_t$ is a solution of the stochastic differential equation (1). Also, if $\alpha = \theta$ and $\beta = -\theta^2/2$ for some scalar $\theta \in \mathbb{R}$ then equation (21) simplifies to

$$\frac{dS_t}{S_t} = \alpha dW_t, \qquad \text{equivalently,}$$

$$\exp\{\theta W_t - \theta^2 t/2\} = 1 + \theta \int_0^t \exp\{\theta W_s - \theta^2 s/2\} \, dW_s.$$

By Proposition 5, it follows that the process $\exp\{\theta W_t - \theta^2 t/2\}$ is a martingale relative to the standard filtration.

## 5. ITÔ'S FORMULA: THE GENERAL CASE

**Definition 3.** *An **Itô process** is a stochastic process of the form*

(22)
$$Z_t = Z_0 + \int_0^t A_s \, ds + \int_0^t B_s \, dW_s$$

*where for each $t < \infty$ the process $\{A_s \mathbf{1}_{(0,t]}(s)\}$ is in the class $\mathcal{H}^1_{\text{LOC}}$ and the process $\{B_s \mathbf{1}_{(0,t]}(s)\}$ is in the class $\mathcal{H}^2_{\text{LOC}}$.*

The Itô processes are the processes that solve stochastic differential equations. In particular, if the Itô process $\{Z_t\}$ satisfies the integral equation (22), then it solves the stochastic differential equation (SDE)

$$(23) \qquad\qquad dZ_t = A_t\, dt + B_t\, dW_t.$$

Now that we have defined the Itô integral against $dW$, we can easily extend the definition to allow integration against $dZ$ for any Itô process $Z$, using (23): for an adapted process $\{\theta_s\}$, define

$$(24) \qquad\qquad \int_0^t \theta_s\, dZ_s = \int_0^t \theta_s A_s\, ds + \int_0^t \theta_s B_s\, dW_s.$$

If for each $t < \infty$ the process $\{\theta_s A_s \mathbf{1}_{[0,t]}(s)\}$ is in the class $\mathcal{H}^1_{\mathrm{LOC}}$ and $\{\theta_s B_s \mathbf{1}_{[0,t]}(s)\}$ is in the class $\mathcal{H}^2_{\mathrm{LOC}}$ then both of the integrals on the right side of this equation are well–defined.

**Theorem 2.** *(Itô's Formula) Let $\{Z_t\}$ be an Itô process satisfying the SDE (23), and let $u(x,t)$ be twice continuously differentiable in $x$ and once continuously differentiable in $t$, and define $U_t = u(Z_t, t)$. Then $\{U_t\}$ is an Itô process that satisfies the SDE*

$$(25) \qquad dU_t = u_x(Z_t, t)\, dZ_t + u_t(Z_t, t)\, dt + \frac{1}{2} u_{xx}(Z_t, t) B_t^2\, dt$$

$$= u_x(Z_t, t) B_t\, dW_t + \left( u_x(Z_t, t) A_t + \frac{1}{2} u_{xx}(Z_t, t) B_t^2 + u_t(Z_t, t) \right) dt$$

*Proof.* See section 8 below for a sketch. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

## 6. When is an Itô Process a Martingale?

**Proposition 6.** *Let $\{Z_t\}$ be an Itô process satisfying the Itô integral equation (22), where $Z_0$ has finite first moment, $\{A_t\}$ is of class $\mathcal{H}^1$, and $\{B_t\}$ is of class $\mathcal{H}^2$. Then $\{Z_t\}$ is a martingale (relative to the Brownian filtration) if and only if $A_t = 0$ almost surely, for almost every $t \geq 0$.*

*Proof.* Suppose first that $A_t = 0$ almost surely, for all $t \geq 0$. Then $Z_t - Z_0 = \int_0^t B_s\, dW_s$, and so by Proposition 5, the process $\{Z_t - Z_0\}_{t \geq 0}$ is a martingale. It follows that the process $\{Z_t\}_{t \geq 0}$ is a martingale, by the linearity of conditional expectation.

The converse is a bit trickier. Suppose that the process $\{Z_t\}_{t \geq 0}$ is a martingale. Since the process $\int_0^t B_s\, dW_s$ is also a martingale, by Proposition 5, and since the difference between two martingales is a martingale (why?), it follows that

$$\int_0^t A_s\, ds$$

is a martingale. This means that, for any $t, u \geq 0$,

$$\int_0^t A_s \, ds = E\left(\int_0^{t+u} A_s \, ds \mid \mathcal{F}_t\right) \qquad \Longrightarrow$$

$$\int_0^t A_s \, ds = E\left(\int_0^t A_s \, ds + \int_t^{t+u} A_s \, ds \mid \mathcal{F}_t\right) \qquad \Longrightarrow$$

$$\int_0^t A_s \, ds = \int_0^t A_s \, ds + E\left(\int_t^{t+u} A_s \, ds \mid \mathcal{F}_t\right) \qquad \Longrightarrow$$

$$0 = E\left(\int_t^{t+u} A_s \, ds \mid \mathcal{F}_t\right) \qquad \Longrightarrow$$

$$0 = \int_t^{t+u} E(A_s \mid \mathcal{F}_t) \, ds \qquad \Longrightarrow$$

$$0 = \frac{1}{u} \int_t^{t+u} E(A_s \mid \mathcal{F}_t) \, ds$$

Somehow we must conclude from this that $A_s = 0$ almost surely for every $s$. There are several ways to do this. The shortest is to use the *Lebesgue differentiation theorem* (see, for example, ROYDEN, ch. 5), which, in essence, states that the Fundamental Theorem of Calculus remains valid, suitably interpreted, for integrands of class $L^1$. Since the last displayed equation holds for every $t$ and every $u > 0$, it holds also in the limit as $u \to 0+$. The Lebesgue differentiation theorem[4] tells us what the limit must be: that

$$\lim_{u \to 0+} \frac{1}{u} \int_t^{t+u} E(A_s \mid \mathcal{F}_t) \, ds = E(A_t \mid \mathcal{F}_t) = A_t$$

almost surely, for almost every $t$. It follows that $A_t = 0$ almost surely, for almost every $t \geq 0$. $\qquad \square$

### 7. EXERCISES

1. Evaluate $\int_0^1 W_s^3 \, dW_s$.

2. Use the Itô formula to show that, for any integer $k \geq 2$,

$$EW(t)^k = \frac{1}{2} k(k-1) \int_0^t EW(s)^{k-2} \, ds,$$

and use this to evaluate the first 23 moments of the normal distribution.

3. Use the Itô formula to show that, for any *nonrandom*, continuously differentiable function $f(t)$,

$$\int_0^t f(s) \, dW_s = f(t)W(t) - \int_0^t f'(s) W_s ds.$$

4. **Heat Martingales.** Suppose that $u(x, t)$ is a bounded solution to the (reverse) heat equation $u_{xx}/2 + u_t = 0$ for $x \in \mathbb{R}$ and $0 \leq t \leq T$. Define

$$Z_t = u(W_t, t) \qquad \text{for } 0 \leq t \leq T.$$

Show that $\{Z_t\}_{0 \leq t \leq T}$ is a martingale relative to the (truncated) Brownian filtration $\{\mathcal{F}_t\}_{0 \leq t \leq T}$. Conclude that

$$u(x, t) = Eu(W_{T-t} + x, T).$$

---

[4] That the integrand is $L^1$ almost surely follows from the hypothesis that the process $\{A_t\}_{t \geq 0}$ is of class $\mathcal{H}^1$.

5. **Stock Price Processes.** Let $\{S_t\}_{t\geq 0}$ be the price process of a STOCK in a market with a riskless asset BOND whose price at time $t$ is $B_t = e^{rt}$. Suppose that $\{S_t\}$ obeys the SDE (1) for some parameters $\mu$ and $\sigma > 0$. Show that $\mu = r$.

## 8. APPENDIX: DERIVATION OF THE ITÔ FORMULA

For simplicity, let $t = 1$, and assume that the partial derivatives $u_t$ and $u_{xx}$ are uniformly bounded and uniformly continuous. (The theorem is true without these hypotheses, but the proof then requires further approximation arguments, which I deem not very enlightening). For each $n \geq 1$, let $\mathcal{D}_n = \mathcal{D}_n[0,1]$ be the $n$th dyadic partition of the unit interval, that is, the partition into intervals $[t_j, t_{j+1}]$, where $t_j = j/2^n$.

**Lemma 2.** *For any uniformly continuous, nonnegative, uniformly bounded function $g(x,t)$,*

$$(26) \qquad \lim_{n\to\infty} \sum_{j=0}^{2^n-1} g(W(t_j), t_j)((W(t_{j+1}) - W(t_j))^2 = \int_0^1 g(W_t, t)\, dt \qquad and$$

$$(27) \qquad \lim_{n\to\infty} \sum_{j=0}^{2^n-1} o(((W(t_{j+1}) - W(t_j))^2) = 0$$

NOTE: The notation $o(y)$ indicates that the term is of smaller order of magnitude than $y$ as $n \to \infty$, that is, that $o(y)/y \to 0$ as $n \to \infty$. The proof of this lemma is deferred to the end of the section.

ANOTHER NOTE: Relation (26) is a generalization of the Quadratic Variation Formula proved in Lecture 5. It is the mathematical basis for the heuristic equation

$$(dW_t)^2 = dt$$

often encountered in formal calculations.

*Proof of the Itô Formula (Sketch).*
Expressing $u(W_1, 1) - u(0,0)$ as a telescoping sum yields

$$u(W_1, 1) - u(0,0) = \sum_{j=1}^{2^n} (u(W(t_j), t_j) - u(W(t_{j-1}), t_{j-1}))$$

$$= \sum_{j=1}^{2^n} (u(W(t_j), t_j) - u(W(t_j), t_{j-1}))$$

$$+ \sum_{j=1}^{2^n} (u(W(t_j), t_{j-1}) - u(W(t_{j-1}), t_{j-1}))$$

$$:= S_1^{(n)} + S_2^{(n)}.$$

The terms of the first sum $S_1^{(n)}$ may be approximated by using Taylor's series for the function $u(x,t)$, applied in the variable $t$:

$$u(W(t_j), t_j) - u(W(t_j), t_{j-1}) = u_t(W(t_j), t_j)(t_j - t_{j-1}) + o(t_j - t_{j-1})$$

$$= u_t(W(t_j), t_j)2^{-n} + o(2^{-n}).$$

The error term $o(2^{-n})$ is uniformly small, in view of our assumption that the first partial derivative $u_t(x,t)$ is uniformly bounded; consequently, if these $2^n$ errors are summed over all $j$, the resulting

error is $o(1)$. Summing the approximating terms on $j$ gives a Riemann sum for the ordinary Riemann integral $\int_0^1 u_s(W_s, s) \, ds$:

$$S_1^{(n)} = 2^{-n} \sum_{j=1}^{2^n} u_t(W(t_j), t_j) + o(1) \longrightarrow \int_0^1 u_s(W_s, s) \, ds.$$

The terms of the second sum $S_2^{(n)}$ may also be approximated by using the Taylor's series for the function $u(x, t)$, this time in the variable $x$, and this time using two terms of the series to keep the errors small:

$$
\begin{aligned}
u(W(t_j), t_{j-1}) - u(W(t_{j-1}), t_{j-1}) \\
&= u_x(W(t_{j-1}), t_{j-1})(W(t_j) - W(t_{j-1})) \\
&\quad + (1/2)u_{xx}(W(t_{j-1}), t_{j-1})(W(t_j) - W(t_{j-1}))^2 \\
&\quad + o((W(t_j) - W(t_{j-1}))^2).
\end{aligned}
$$

By Lemma 2 and the definition of the Itô integral, as $n \to \infty$,

$$\sum_{j=1}^{2^n} u_x(W(t_{j-1}), t_{j-1})(W(t_j) - W(t_{j-1})) \longrightarrow \int_0^1 u_x(W(s), s) \, dW_s,$$

$$\sum_{j=1}^{2^n} u_{xx}(W(t_{j-1}), t_{j-1})(W(t_j) - W(t_{j-1}))^2 \longrightarrow \int_0^1 u_{xx}(W_s, s) \, ds, \text{ and}$$

$$\sum_{j=1}^{2^n} o((W(t_j) - W(t_{j-1}))^2) \longrightarrow 0.$$

It follows that

$$
\begin{aligned}
u(W_1, 1) - u(0, 0) &= \lim_{n \to \infty} (S_1^{(n)} + S_2^{(n)}) \\
&= \int_0^1 u_s(W_s, s) \, ds + \int_0^1 u_x(W(s), s) \, dW_s \\
&\quad + \frac{1}{2} \int_0^1 u_{xx}(W_s, s) \, ds.
\end{aligned}
$$

*Proof of Lemma 2 (Sketch)* If the function $g(x, t)$ were *constant*, then this would follow immediately from the Quadratic Variation Formula proved in Lecture 5. For functions $g$ that are merely continuous (and uniformly bounded), the result also follows from the Quadratic Variation Formula, but a more delicate argument is needed. In brief, the idea is this: Although $g(W_t, t)$ is not constant, it is *nearly* constant over short stretches of time, by continuity. In particular, for any $\varepsilon > 0$ there exists $\delta > 0$ such that, with probability at least $1 - \varepsilon$,

$$|s - t| < \delta \qquad \Longrightarrow \qquad |g(W_t, t) - g(W_s, s)| < \varepsilon.$$

Without loss of generality, we may take $\delta = 1/2^m$ for some integer $m$. Now, over each time interval of length $2^{-m}$, we approximate $g(W_t, t)$ by its value at the beginning of the time interval: For

$n > m$,

$$(28) \quad \sum_{j=0}^{2^n-1} g(W(t_j), t_j)((W(t_{j+1}) - W(t_j))^2$$

$$= \sum_{i=0}^{2^m-1} g(W(i/2^m)) \left( \sum_{j=0}^{2^{n-m}-1} ((W((i/2^m) + (j/2^n)) - W((i/2^m) + (j/2^n) - (1/2^n)))^2 \right)$$

$$+ \text{ error,}$$

where

$$\text{error} \leq \varepsilon \sum_{j=0}^{2^n-1} ((W(t_{j+1}) - W(t_j))^2 \approx \varepsilon.$$

Note that the last approximation is valid as $n \to \infty$, by the Quadratic Variation Formula. But observe also that each of the inner sums in (28) is also of the type considered in the Quadratic Variation Formula, and so, as $n \to \infty$, each of these inner sums converges to $1/2^m$. Thus, for each fixed $m$,

$$\limsup_{n\to\infty} | \sum_{j=0}^{2^n-1} g(W(t_j), t_j)((W(t_{j+1}) - W(t_j))^2 - 2^{-m} \sum_{i=0}^{2^m-1} g(W(i/2^m))| \leq \varepsilon.$$

Now

$$\lim_{m\to\infty} 2^{-m} \sum_{i=0}^{2^m-1} g(W(i/2^m)) = \int_0^1 g(W_t, t)\, dt$$

Thus,

$$\limsup_{n\to\infty} | \sum_{j=0}^{2^n-1} g(W(t_j), t_j)((W(t_{j+1}) - W(t_j))^2 - \int_0^1 g(W_t, t)\, dt| \leq \varepsilon.$$

Since $\varepsilon > 0$ may be chosen arbitrarily small, the result (26) follows. The second result (27) may be proved by similar arguments.