Notes on the Itô Calculus

Steven P. Lalley

May 15, 2012

1 Continuous-Time Processes: Progressive Measurability

1.1 Progressive Measurability

Measurability issues are a bit like plumbing – a bit ugly, and most of the time you would prefer that it remains hidden from view, but sometimes it is unavoidable that you actually have to open it up and work on it. In the theory of continuous–time stochastic processes, measurability problems are usually more subtle than in discrete time, primarily because sets of measures 0 can add up to something significant when you put together uncountably many of them. In stochastic integration there is another issue: we must be sure that any stochastic processes $X_t(\omega)$ that we try to integrate are *jointly* measurable in (t, ω), so as to ensure that the integrals are themselves random variables (that is, measurable in ω).

Assume that $(\Omega, \mathcal{F}, P)$ is a fixed probability space, and that all random variables are $\mathcal{F}-$measurable. A *filtration* is an increasing family $\mathbb{F} := \{\mathcal{F}_t\}_{t \in J}$ of sub-$\sigma-$algebras of $\mathcal{F}$ indexed by $t \in J$, where J is an interval, usually $J = [0, \infty)$. Recall that a stochastic process $\{Y_t\}_{t \geq 0}$ is said to be *adapted* to a filtration if for each $t \geq 0$ the random variable Y_t is $\mathcal{F}_t-$measurable, and that a filtration is *admissible* for a Wiener process $\{W_t\}_{t \geq 0}$ if for each $t > 0$ the post-t increment process $\{W_{t+s} - W_t\}_{s \geq 0}$ is independent of $\mathcal{F}_t$.

Definition 1. A stochastic process $\{X_t\}_{t \geq 0}$ is said to be *progressively measurable* if for every $T \geq 0$ it is, when viewed as a function $X(t, \omega)$ on the product space $([0, T]) \times \Omega$, measurable relative to the product $\sigma-$algebra $\mathcal{B}_{[0,T]} \times \mathcal{F}_T$.

Progressive measurability is the least we should expect for any stochastic process that we hope to integrate, because this is what is necessary for the integral over any time interval to be a random variable. If a process $\{X_t\}_{t \in J}$ is progressively measurable then it is necessarily adapted. Not every adapted process is progressively measurable, though, at least if you subscribe to the Axiom of Choice, as this implies the existence of a non-measurable function $f : \mathbb{R}_+ \to \mathbb{R}$. For any such function, the process $X(t, \omega) = f(t)$ is adapted but not progressively measurable. Thus, checking that a process is progressively measurable is in principle more difficult than checking that it is adapted, but there are some simple, useful sufficient conditions.

Lemma 1. *Limits of progressively measurable processes are progressively measurable. In addition, if $\{X_t\}_{t \in J}$ is an adapted process with right- or left-continuous paths, then it is progressively measurable*

Proof. Consider the case where the process has right-continuous paths (the case of left-continuous processes is similar). Fix $T > 0$; we must show that the function $X(t, \omega)$ is jointly measurable in t, ω relative to $\mathcal{B}_{[0,T]} \times \mathcal{F}_T$. For integers $m \geq 1$ and $0 \leq k \leq 2^m$ define

$$X_m(t, \omega) = X(kT/2^m, \omega) \quad \text{for } \ (k-1)T/2^m \leq t < kT/2^m.$$

Clearly, $X_m(t, \omega)$ is jointly measurable in (t, ω) relative to the product $\sigma-$algebra $\mathcal{B}_{[0,T]} \times \mathcal{F}_T$ (even though it is not adapted). Moreover, because the process $X(t, \omega)$ is right-continuous in t,

$$\lim_{m \to \infty} X_m(t, \omega) = X(t, \omega) \quad \text{for all } \ t \leq T, \omega \in \Omega. \tag{1}$$

Therefore, $X(t, \omega)$ is jointly measurable relative to $\mathcal{B}_{[0,T]} \times \mathcal{F}_T$. □

Example 1. Let W_t be a one-dimensional Wiener process, and $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$ a standard, admissible filtration. Is the process $Y(t) = \mathbf{1}\{W_t > 0\}$ progressively measurable? To see that it is, consider the sequence of processes $Y_n(t) = (W_t \wedge 1)_+^{1/n}$ where the subscript $+$ denotes positive part and $\wedge$ denotes min. For each integer $n \geq 1$ the process Y_n is progressively measurable (because it is adapted and has continuous paths), and $Y_n(t) \to Y(t)$ pointwise as $n \to$. NOTE: As we will see later, there are some very good reason for wanting to integrate the process $Y(t)$ against the underlying Wiener process W_t.

What happens when a process is evaluated at a stopping time?

Lemma 2. *Let $\tau \in J$ be a stopping time relative to the filtration $\mathbb{F}$ and let $\{X_t\}_{t \in J}$ be progressively measurable. Then $X_\tau \mathbf{1}_{\{\tau < \infty\}}$ is an $\mathcal{F}_\tau-$measurable random variable.*

Proof. EXERCISE. □

Definition 2. Fix an interval $J = [0, T]$ or $J = [0, \infty)$. The class $\mathcal{V}^p = \mathcal{V}^p_J$ consists of all progressively measurable processes $X_t = X(t, \omega)$ on $J \times \Omega$ such that

$$\int_J E|X_t|^p \, dt < \infty. \tag{2}$$

Lemma 3. *If X_s belongs to the class $\mathcal{V}^1_J$ then for almost every $\omega \in \Omega$ the function $t \mapsto X_t(\omega)$ is Borel measurable and integrable on J, and $X(t, \omega)$ is integrable relative to the product measure Lebesgue$\times P$. Furthermore, the process*

$$Y_t := \int_{s \leq t} X_s \, ds \tag{3}$$

is continuous in t (for almost every ω) and progressively measurable.

Proof. Assume first that the process X is bounded. Fubini's theorem implies that for every $\omega \in \Omega$, the function $X_t(\omega)$ is an integrable, Borel measurable function of t. Thus, the process Y_t is well-defined, and the dominated convergence theorem implies that Y_t is continuous in t. To prove that the process Y_t is progressively measurable it suffices, by Lemma 1, to check that it is adapted. This follows from Fubini's theorem.

The general case follows by a truncation argument. If $X \in \mathcal{V}^1$, then by Tonelli's theorem $X(t, \omega)$ is integrable relative to the product measure Lebesgue$\times P$, and Fubini's theorem implies that $X_t(\omega)$ is Borel measurable in t and Lebesgue integrable for $P-$almost every ω. Denote by X^m the truncation of X at $-m, m$. Then the dominated convergence theorem implies that for $P-$almost every ω the integrals

$$Y_t^m(\omega) := \int_{s\leq t} X_s^m(\omega)\, ds$$

converge to the integral of X_s. Define $Y_t(\omega) = 0$ for those pairs (t, ω) for which the sequence $Y_t^m(\omega)$ does not converge. Then the process $Y_t(\omega)$ is well-defined, and since it is the almost everywhere limit of the progressively measurable processes $Y_t^m(\omega)$, it is progressively measurable. □

1.2 Approximation of progressively measurable processes

The usual procedure for defining an integral is to begin by defining it for a class of simple integrands and then extending the definition to a larger, more interesting class of integrands by approximation. To define the Itô integral (section 3 below), we will extend from a simple class of integrands called *elementary processes*. A stochastic process $V(t) = V_t$ is called *elementary* relative to the filtration $\mathbb{F}$ if if it has the form

$$V_t = \sum_{j=0}^{K} \xi_j \mathbf{1}_{(t_j, t_{j+1}]}(t) \tag{4}$$

where $0 = t_0 < t_1 < \cdots < t_K < \infty$, and for each index j the random variable ξ_j is measurable relative to $\mathcal{F}_{t_j}$. Every elementary process is progressively measurable, and the class of elementary processes is closed under addition and multiplication (thus, it is an *algebra*).

Exercise 1. Let τ be a stopping time that takes values in a finite set $\{t_i\}$. (Such stopping times are called *elementary*.) Show that the process $V_t := \mathbf{1}_{(0,\tau]}(t)$ is elementary.

Lemma 4. *For any $p \in [1, \infty)$, the elementary processes are L^p-dense in the space $\mathcal{V}^p$. That is, for any $Y \in \mathcal{V}^p$ there is a sequence V_n of elementary functions such that*

$$E\int_J |Y(t) - V_n(t)|^p\, dt \longrightarrow 0 \tag{5}$$

The proof requires an intermediate approximation:

Lemma 5. *Let X be a progressively measurable process such that $|X(t, \omega)| \leq K$ for some constant $K < \infty$. Then there exists a sequence X^m of progressively measurable processes with continuous paths, and all uniformly bounded by K, such that*

$$X(t, \omega) = \lim_{m\to\infty} X^m(t, \omega) \quad a.e. \tag{6}$$

Note 1. OKSENDAL mistakenly asserts that the same is true for *adapted* processes — see Step 2 in his proof of Lemma 3.1.5. This is unfortunate, as this is the key step in the construction of the Itô integral.

Proof. This is a consequence of the Lebesgue differentiation theorem. Since X is bounded, Lemma 3 implies that the indefinite integral Y_t of X_s is progressively measurable and has continuous paths. Define

$$X_t^m = m(Y_t - Y_{(t-1/m)_+});$$

then X^m is progressively measurable, has continuous paths, and is uniformly bounded by K (because it is the average of X_s over the interval $[t-1/m, t]$). The Lebesgue differentiation theorem implies that $X^m(t,\omega) \to X(t,\omega)$ for almost every pair (t,ω). □

Proof of Lemma 4. Any process in the class $\mathcal{V}^p$ can be arbitrarily well-approximated in L^p norm by *bounded*, progressively measurable processes. (Proof: Truncate at $\pm m$, and use the dominated convergence theorem.) Thus, it suffices to prove the lemma for bounded processes. But any bounded process can be approximated by processes with continuous paths, by Lemma 5. Thus, it suffices to prove the lemma for bounded, progressively measurable processes with continuous sample paths.

Suppose, then, that $Y_t(\omega)$ is bounded and continuous in t. For each $n = 1, 2, \dots$, define

$$V_n(t) = \sum_{j=0}^{2^{2n}-1} Y(j/2^n)\mathbf{1}_{[j/2^n,(j+1)/2^n)}(t);$$

this is an elementary process, and because Y is bounded, $V_n \in \mathcal{V}^p$. Since Y is continuous in t for $t \leq 2^K$,

$$\lim_{n\to\infty} V_n(t) = Y_t,$$

and hence the L^p- convergence (5) follows by the bounded convergence theorem. □

2 Continuous Martingales

2.1 Review.

Let $\mathbb{F} := \{\mathcal{F}_t\}_{t\geq 0}$ be a filtration of the probability space $(\Omega, \mathcal{F}, P)$ and let $M_t = M(t)$ be a martingale relative to $\mathbb{F}$. Unless otherwise specified, assume that $M(t)$ is right-continuous with left limits at every t. (In fact, all of the martingales we will encounter in connection with Itô integrals will have versions with continuous paths.) For any stopping time τ denote by $\mathcal{F}_\tau$ the stopping field associated with τ, that is,

$$\mathcal{F}_\tau := \{A \in \mathcal{F} : A \cap \{\tau \leq t\} \in \mathcal{F}_t \quad \forall\, t \geq 0\}. \tag{7}$$

Proposition 1. *If $0 = \tau_0 \leq \tau_1 \leq \tau_2 \leq \cdots$ are stopping times, then the sequence $M(\tau_n)$ is a martingale relative to the discrete filtration $\mathcal{F}_{\tau_n}$* provided *either (a) the stopping times are bounded, or (b) the martingale M_t is uniformly integrable. In either case, the* Doob Optional Sampling Formula *holds for every τ_n:*

$$EM_{\tau_n} = EM_0. \tag{8}$$

Proposition 2. *If $\{M_t\}_{t\geq 0}$ is bounded in L^p for some $p > 1$ then it is uniformly integrable and closed in L^p, that is,*

$$\lim_{t\to\infty} M_t := M_\infty \tag{9}$$

exists a.s and in L^p, and for each $t < \infty$,

$$M_t = E(M_\infty \,|\, \mathcal{F}_t). \tag{10}$$

Note 2. This is not true for $p = 1$: the "double-or-nothing" martingale is an example of a nonnegative martingale that converges a.s. to zero, but not in L^1.

Proposition 3. *(Doob's Maximal Inequality) If M_t is closed in L^p for some $p \geq 1$ then*

$$P\{\sup_{t \in \mathbb{Q}_+} |M_t| \geq \alpha\} \leq E|M_\infty|^p/\alpha^p \quad \textit{and} \tag{11}$$

$$P\{\sup_{\substack{t \leq T \\ t \in \mathbb{Q}_+}} |M_t| \geq \alpha\} \leq E|M_T|^p/\alpha^p \tag{12}$$

Note 3. The restriction $t \in \mathbb{Q}_+$ is only to guarantee that the sup is a random variable; if the martingale is right-continuous then the sup may be replaced by $\sup_{t \geq 0}$. Also, the second inequality (12) follows directly from the first, because the martingale M_t can be "frozen" at its value M_T for all $t \geq T$.

2.2 Martingales with continuous paths

Definition 3. Let $\mathcal{M}_2$ be the linear space of all L^2-bounded, martingales M_t with *continuous* paths and initial values $M_0 = 0$. Since each such martingale $M = (M_t)_{0 \leq t < \infty}$ is L^2-closed, it has an L^2-limit M_∞. Define the $\mathcal{M}_2-$norm of M by $\|M\| = \|M_\infty\|_2$.

Proposition 4. *The space $\mathcal{M}_2$ is complete in the metric determined by the norm $\|\cdot\|$. That is, if $M_n = \{M_n(t)\}_{t \geq 0}$ is a Cauchy sequence in $\mathcal{M}_2$ then the sequence of random variables $M_n(\infty)$ converges in L^2 to some random variable $M(\infty)$, and the the martingale*

$$M(t) := E(M(\infty)|\mathcal{F}_t) \tag{13}$$

has a version with continuous paths.

Proof. This follows from Doob's maximal inequality. First, a sequence $\{M_n\}$ of L^2- martingales is Cauchy in $\mathcal{M}_2$ if and only if the sequence $\{M_n(\infty)\}$ is Cauchy in L^2, by definition of the norm. Consequently, the random variables $M_n(\infty)$ converge in L^2, because the space L^2 is complete. Let $M(\infty) = \lim_{n \to \infty} M_n(\infty)$, and define $M(t)$ by conditional expectation as in equation (13). Then for each t,

$$M(t) - M_n(t) = E(M(\infty) - M_n(\infty) \,|\, \mathcal{F}_t).$$

Without loss of generality, assume (by taking a subsequence if necessary) that the L^2-norm of $M(\infty) - M_n(\infty)$ is less than 4^{-n}.Then by the maximal inequality,

$$P\{\sup_{t \geq 0} |M_n(t) - M(t)| \geq 2^{-n}\} \leq 4^{-2n}/4^{-n} = 4^{-n}.$$

(The sup is over *rational* $t \geq 0$.) Since $\sum 4^{-n} < \infty$, the Borel-Cantelli lemma implies that w.p.1, for all sufficiently large n the maximal difference between M_n and M will be less than 2^{-n}. Hence, w.p.1 the sequence $M_n(t)$ converges uniformly for rational t to $M(t)$. Since each $M_n(t)$ is continuous in t, it follows that the limit $M(t)$ must also be continuous (more precisely, it has a version with continuous paths). $\square$

2.3 Martingales of bounded variation

Proposition 5. *Let $M(t)$ be a continuous-time martingale with continuous paths. If the paths of $M(t)$ are of bounded variation on a time interval J, then $M(t)$ is constant on J.*

Proof. Recall that a function $F(t)$ is of bounded variation on J if there is a finite constant C such that for every finite partition $\mathcal{P} = \{I_j\}$ of J into finitely many nonoverlapping intervals I_j,

$$\sum_j |\Delta_j(F)| \leq C$$

where $\Delta_j(F)$ denotes the increment of F over the interval I_j. The supremum over all partitions of J is called the *total variation* of F on J, and is denoted by $\|F\|_{TV}$. If F is continuous then the total variation of F on $[a, b] \subset J$ is increasing and continuous in b. Also, if F is continuous and J is compact then F is uniformly continuous, and so for any $\varepsilon > 0$ there exists $\delta > 0$ so that if $\max_j |I_j| < \delta$ then $\max_j |\Delta_j(F)| < \varepsilon$, and so

$$\sum_j |\Delta_j(F)|^2 \leq C\varepsilon.$$

It suffices to prove Proposition 5 for Intervals $J = [0, b]$. It also suffices to consider *bounded* martingales of *bounded* total variation on J. To see this, let $M(t)$ be an arbitrary continuous martingale and let $M_n(t) = M(t \wedge \tau_n)$, where τ_n is the first time that either $|M(t)| = n$ or such that the total variation $\|M\|_{TV}$ reaches n. Since $M(t)$ is continuous, it is bounded for t in any finite interval J, by Weierstrass' theorem. Consequently, for all sufficiently large n, the function $M_n(t)$ will coincide with $M(t)$; thus, if each $M_n(t)$ is continuous on J then so is $M(t)$.

Suppose then that $M(t)$ is a bounded, continuous martingale of total variation no larger than C on J. Then there is a nested sequence of partitions $\mathcal{P}_n$ with mesh tending to 0 such that

$$\sum_j |\Delta_j^n(M)|^2 \longrightarrow 0 \quad \text{almost surely.}$$

Because the increments of L^2 martingales over nonoverlapping time intervals are orthogonal (Exercise: why?), if $J = [a, b]$ then for any partition $\mathcal{P}_n$

$$E(M(b) - M(a))^2 = E \sum_j |\Delta_j^n(M)|^2.$$

But by the choice of the partitions $\mathcal{P}_n$, the last sum converges to zero a.s. The sums are uniformly bounded because M is uniformly bounded and has uniformly bounded total variation. Therefore, by the dominated convergence theorem,

$$E(M(b) - M(a))^2 = 0.$$

□

3 Itô Integral: Definition and Basic Properties

3.1 Elementary integrands

Let $W_t = W(t)$ be a (one-dimensional) Wiener process, and fix a standard, admissible filtration $\mathbb{F}$. Recall that a process V_t is called *elementary* if it has the form

$$V_t = \sum_{j=0}^{K} \xi_j \mathbf{1}_{(t_j, t_{j+1}]}(t) \tag{14}$$

where $0 = t_0 < t_1 < \cdots < t_K < \infty$, and for each index j the random variable ξ_j is measurable relative to $\mathcal{F}_{t_j}$.

Definition 4. For a simple process $\{V_t\}_{t \geq 0}$ satisfying equation (14), define the *Itô integral* $I_t(V) = \int_0^t V\, dW$ as follows. (Note: The alternative notation $I_t(V)$ is commonly used in the literature, and I will use it interchangeably with the integral notation.)

$$I_t(V) = \int_0^t V_s\, dW_s := \sum_{j=0}^{K-1} \xi_j (W(t_{j+1} \wedge t) - W(t_j \wedge t)) \tag{15}$$

Why is this a reasonable definition? The random step function $\theta(s)$ takes the (random) value ξ_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 ξ_j times as large as those of the Brownian motion; when one adds up these infinitesimal increments, one gets ξ_j times the total increment of the Brownian motion over this time period.

Properties of the Itô Integral:

(A) Linearity: $I_t(aV + bU) = aI_t(V) + bI_t(U)$.
(B) Measurability: $I_t(V)$ is adapted to $\mathbb{F}$.
(C) Continuity: $t \mapsto I_t(V)$ is continuous.

These are all immediate from the definition.

Proposition 6. *Assume that V_t is elementary with representation* (14)*, and assume that each of the random variables ξ_j has finite second moment. Then*

$$E(\int_0^t V\, dW) = 0 \quad \textit{and} \tag{16}$$

$$E(\int_0^t V\, dW)^2 = \int_0^t EV_s^2\, ds. \tag{17}$$

Proof. See Proposition!7 below. □

The equality (17) is of crucial importance – it asserts that the mapping that takes the process V to its Itô integral at any time t is an L^2-isometry relative to the L^2-norm for the product measure Lebesgue$\times P$. This will be the key to extending the integral to integrands in the class $\mathcal{V}^2$. The simple calculations that lead to (16) and (17) also yield the following useful information about the process $I_t(V)$:

Proposition 7. *Assume that V_t is elementary with representation* (14), *and assume that each of the random variables ξ_j has finite second moment. Then $I_t(V)$ is an L^2-martingale relative to $\mathbb{F}$. Furthermore, if*

$$[I(V)]_t := \int_0^t V_s^2\, ds; \tag{18}$$

then $I_t(V)^2 - [I(V)]_t$ is a martingale.

Note: The process $[I(V)]_t$ is called the *quadratic variation* of the martingale $I_t(V)$. The square bracket notation is standard in the literature.

Proof. First recall that a linear combination of martingales is a martingale, so to prove that $I_t(V)$ is a martingale it suffices to consider elementary functions V_t with just one step:

$$V_t = \xi \mathbf{1}_{(s,r]}(t)$$

with ξ measurable relative to $\mathcal{F}_s$. For such a process V the integral $I_t(V)$ is zero for all $t \leq s$, and $I_t(V) = I_r(V)$ for all $t \geq r$, so to show that $I_t(V)$ is a martingale it is only necessary to check that

$$\begin{aligned} &E(I_t(V)\,|\mathcal{F}_u) = I_u(V) \quad \text{for } s \leq u < t \leq r. \\ \iff \quad &E(\xi(W_t - W_r)\,|\,\mathcal{F}_u) = \xi(W_u - W_r). \end{aligned}$$

But this follows routinely from basic properties of conditional expectation, since ξ is measurable relative to $\mathcal{F}_r$ and W_t is a martingale with respect to $\mathbb{F}$.

It is only slightly more difficult to check that $I_t(V)^2 - [I(V)]_t$ is a martingale (you have to decompose a sum of squares). Let V_t be elementary, and assume that the random variables ξ_j in the representation (14) are in L^2. We must show that for every $s, t \geq 0$,

$$E(I_{t+s}(V)^2\,|\,\mathcal{F}_t) - E([I(V)]_{t+s}\,|\,\mathcal{F}_t) = I_t(V)^2 - [I(V)]_t.$$

It suffices to prove this for values of s such that $s \leq t_j - t_{j-1}$ (where the t_j are the discontinuity points in the representation (14)), by the tower property of conditional expectations. Thus, we may assume without loss of generality that V_r is constant on the interval $r \in [t, t+s]$, that is, $V_r = \xi$ where $\xi \in L^2$ and ξ is measurable with respect to $\mathcal{F}_t$. Under this assumption,

$$I_{t+s}(V) - I_t(V) = \xi(W_{t+s} - W_s).$$

Now $I_t(V)$ is measurable relative to $\mathcal{F}_t$, and hence, since the process $I_r(V)$ is a martingale (by the first part of the proof),

$$\begin{aligned} E(I_{t+s}(V)^2\,|\,\mathcal{F}_t) &= I_t(V)^2 + E((I_{t+s}(V) - I_t(V))^2\,|\,\mathcal{F}_t) + 2I_t(V)E((I_{t+s}(V) - I_t(V))\,|\,\mathcal{F}_t) \\ &= I_t(V)^2 + E((I_{t+s}(V) - I_t(V))^2\,|\,\mathcal{F}_t) \\ &= I_t(V)^2 + \xi^2 E((W_{t+s}(V) - W_t(V))^2\,|\,\mathcal{F}_t) \\ &= I_t(V)^2 + \xi^2 s \\ &= I_t(V)^2 + E([I(V)]_{t+s}\,|\,\mathcal{F}_t) - [I(V)]_t. \end{aligned}$$

$\square$

3.2 Extension to the class $\mathcal{V}_T^2$

Fix $T \leq \infty$. Recall from Lemma 4 that the class $\mathcal{V}^2 = \mathcal{V}_T^2$ consists of all progressively measurable processes V_t, defined for $t \leq T$, such that

$$\|V\|^2 = \|V\|_{\mathcal{V}_T}^2 = \int_0^T EV_s^2\, ds = E[V]_T < \infty. \tag{19}$$

Since $\mathcal{V}^2$ is the only one of the $\mathcal{V}^p$ spaces that will play a role in the following discussion, I may occasionally drop the superscript and just write $\mathcal{V} = \mathcal{V}_T$.

The norm is just the standard L^2-norm for the product measure Lebesgue$\times P$. Consequently, the class $\mathcal{V}_T^2$ is a Hilbert space; in particular, every Cauchy sequence in the space $\mathcal{V}_T^2$ has a limit. The class $\mathcal{E}_T$ of bounded elementary functions is a linear subspace of $\mathcal{V}_T^2$. Lemma 4 implies that $\mathcal{E}_T$ is a *dense* linear subspace.

Corollary 1. (Itô Isometry) *The Itô integral $I_t(V)$ defined by* (15) *extends to all integrands $V \in \mathcal{V}_T$ in such a way that for each $t \leq T$ the mapping $V \mapsto I_t(V)$ is a linear isometry from the space $\mathcal{V}_t$ to the L^2-space of square-integrable random variables. In particular, if V_n is any sequence of bounded elementary functions such that $\|V_n - V\| \to 0$, then for all $t \leq T$,*

$$I_t(V) = \int_0^t V\, dW := L^2 - \lim_{n\to\infty} \int_0^t V_n\, dW \tag{20}$$

exists and is independent of the approximating sequence V_n.

Proof. If $V_n \to V$ in the norm (19) then the sequence V_n is Cauchy with respect to this norm. Consequently, by Proposition 6, the sequence of random variables $I_t(V_n)$ is Cauchy in $L^2(P)$, and so it has an L^2-limit. Linearity and uniqueness of the limit both follow by routine L^2-arguments. □

This extended Itô integral inherits all of the properties of the Itô integral for elementary functions. Following is a list of these properties. Assume that $V \in \mathcal{V}_T$ and $t \leq T$.

Properties of the Itô Integral:

(A) Linearity: $I_t(aV + bU) = aI_t(V) + bI_t(U)$.
(B) Measurability: $I_t(V)$ is progressively measurable.
(C) Continuity: $t \mapsto I_t(V)$ is continuous (for some version).
(D) Mean: $EI_t(V) = 0$.
(E) Variance: $EI_t(V)^2 = \|V\|_{\mathcal{V}_t}^2$.
(F) Martingale Property: $\{I_t(V)\}_{t\leq T}$ is an L^2-martingale.
(G) Quadratic Martingale Property: $\{I_t(V)^2 - [I(V)]_t\}_{t\leq T}$ is an L^1-martingale, where

$$[I(V)]_t := \int_0^t V_s^2\, ds \tag{21}$$

All of these, with the exception of (C), follow routinely from (20) and the corresponding properties of the integral for elementary functions by easy arguments using DCT and the like (but you should fill in the details for (F) and (G)). Property (C) follows from Proposition 4 in Section 2, because for elementary V_n the process $I_t(V_n)$ has continuous paths.

3.3 Quadratic Variation and $\int_0^T W\,dW$

There are tools for calculating stochastic integrals that usually make it unnecessary to use the definition of the Itô integral 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. 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. The Quadratic Variation formula, in its simplest guise, is this:

Proposition 8. *For any $T > 0$,*

$$P - \lim_{n\to\infty} \sum_{k=0}^{2^n-1} (\Delta_k^n W)^2 = T, \tag{22}$$

where

$$\Delta_k^n W := W\left(\frac{kT+T}{2^n}\right) - W\left(\frac{kT}{2^n}\right).$$

Proof. For each fixed n, the increments Δ_k^n are independent, identically distributed Gaussian random variables with mean zero and variance $2^{-n}T$. Hence, the result follows from the WLLN for χ^2-random variables. □

Exercise 2. Prove that the convergence holds almost surely. HINT: Borel-Cantelli and exponential estimates.

Exercise 3. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function with compact support. Show that

$$\lim_{n\to\infty} \sum_{k=0}^{2^n-1} f(W(kT/2^n))(\Delta_k^n W)^2 = \int_0^T f(W(s))\,ds.$$

Exercise 4. Let $W_1(t)$ and $W_2(t)$ be independent Wiener processes. Prove that

$$P - \lim_{n\to\infty} \sum_{k=0}^{2^n-1} (\Delta_k^n W_1)(\Delta_k^n W_2) = 0.$$

HINT: $(W_1(t) + W_2(t))/\sqrt{2}$ is a standard Wiener process.

The Wiener process W_t is itself in the class $\mathcal{V}_T$, for every $T < \infty$, because

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

Thus, by Corollary 1, the integral $\int_0^T W\,dW$ is defined and an element of L^2. To evaluate it, we will use the most obvious approximation of W_s by elementary functions. For simplicity, set $T = 1$. Let $\theta_s^{(n)}$ be the elementary function 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 $W(k/2^n)$: that is,

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

Lemma 6. $\lim_{n\to\infty} \int_0^1 E(W_s - \theta_s^{(n)})^2\, ds = 0.$

Proof. Since the simple process $\theta_s^{(n)}$ takes the value $W(k/2^n)$ for all $s \in [k/2^n, (k+1)/2^n]$,

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

□

Corollary 1 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 elementary, 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":

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

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 (Proposition 8), and so converges, as $n \to \infty$, to 1. Therefore, $\int_0^1 W\, dW = (W_1^2 - 1)/2$. More generally,

$$\boxed{\int_0^T W_s\,dW_s = \frac{1}{2}(W_T^2 - T).} \tag{23}$$

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

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

Thus, formula (23) shows that the Itô calculus is fundamentally different than ordinary calculus.

3.4 Stopping Rule for Itô Integrals

Proposition 9. *Let $V_t \in \mathcal{V}_T$ and let $\tau \leq T$ be a stopping time relative to the filtration $\mathbb{F}$. Then*

$$\int_0^\tau V_s\,dW_s = \int_0^T V_s\mathbf{1}_{[0,\tau]}(s)\,dW_s. \tag{24}$$

In other words, if the Itô integral $I_t(V)$ is evaluated at the random time $t = \tau$, the result is a.s. the same as the Itô integral $I_T(V\mathbf{1}_{[0,\tau]})$ of the truncated process $V_s\mathbf{1}_{[0,\tau]}(s)$.

Proof. First consider the special case where both V_s and τ are elementary (in particular, τ takes values in a finite set). Then the truncated process $V_s\mathbf{1}_{[0,\tau]}(s)$ is elementary (Exercise 1 above), and so both sides of (24) can be evaluated using formula (15). It is routine to check that they give the same value (do it!).

Next, consider the case where V is elementary and $\tau \leq T$ is an arbitrary stopping time. Then there is a sequence $\tau_m \leq T$ of elementary stopping times such that $\tau_m \downarrow \tau$. By path-continuity of $I_t(V)$ (property (C) above),

$$\lim_{n\to\infty} I_{\tau_n}(V) = I_\tau(V).$$

On the other hand, by the dominated convergence theorem, the sequence $V\mathbf{1}_{[0,\tau_n]}$ converges to $V\mathbf{1}_{[0,\tau]}$ in $\mathcal{V}_T-$norm, so by the Itô isometry,

$$L^2 - \lim_{n\to\infty} I_T(V\mathbf{1}_{[0,\tau_n]}) = I_T(V\mathbf{1}_{[0,\tau]}).$$

Therefore, the equality (24) holds, since it holds for each τ_m.

Finally, consider the general case $V \in \mathcal{V}_T$. By Proposition **??**, there is a sequence V_n of bounded elementary functions such that $V_n \to V$ in the $\mathcal{V}_T-$norm. Consequently, by the dominated convergence theorem, $V_n\mathbf{1}_{[\mathbf{0},\tau]} \to V\mathbf{1}_{[\mathbf{0},\tau]}$ in $\mathcal{V}_T-$norm, and so

$$I_T(V_n\mathbf{1}_{[\mathbf{0},\tau]}) \longrightarrow I_T(V\mathbf{1}_{[\mathbf{0},\tau]})$$

in L^2, by the Itô isometry. But on the other hand, Doob's Maximal Inequality (see Proposition 4), together with the Itô isometry, implies that

$$\max_{t\leq T} |I_t(V_n) - I_t(V)| \longrightarrow 0$$

in probability. The equality (24) follows. □

Corollary 2. *(Localization Principle) Let $\tau \leq T$ be a stopping time. Suppose that $V, U \in \mathcal{V}_T$ are two processes that agree up to time τ, that is, $V_t \mathbf{1}_{[0,\tau]}(t) = U_t \mathbf{1}_{[0,\tau]}(t)$. Then*

$$\int_0^\tau U\, dW = \int_0^\tau V\, dW. \tag{25}$$

Proof. Immediate from Proposition 9. □

3.5 Extension to the Class $\mathcal{W}_T$

Fix $T \leq \infty$. Define $\mathcal{W} = \mathcal{W}_T$ to be the class of all progressively measurable processes $V_t = V(t)$ such that

$$P\left\{\int_0^T V_s^2\, ds < \infty\right\} = 1 \tag{26}$$

Proposition 10. *Let $V \in \mathcal{W}_T$, and for each $n \geq 1$ define $\tau_n = T \wedge \inf\{t : \int_0^t V_s^2\, ds \geq n\}$. Then for each n the process $V(t)\mathbf{1}_{[0,\tau_n]}(t)$ is an element of $\mathcal{V}_T$, and*

$$\lim_{n\to\infty} \int_0^t V_s \mathbf{1}_{[0,\tau_n]}(s)\, dW_s := \int_0^t V_s\, dW_s := I_t(V) \tag{27}$$

exists almost surely and varies continuously with $t \leq T$. The process $\{I_t(V)\}_{t\leq T}$ is called the Itô integral process *associated to the integrand V.*

Proof. First observe that $\lim_{n\to\infty} \tau_n = T$ almost surely; in fact, with probability one, for all but finitely many n it will be the case that $\tau_n = T$. Let $G_n = \{\tau_n = T\}$. By the Localization Principle (Corollary 2and the Stopping Rule, for all n, m,

$$\int_0^{t\wedge\tau_n} V_s \mathbf{1}_{[0,\tau_{n+m}]}(s)\, dW_s = \int_0^t V_s \mathbf{1}_{[0,\tau_n]}(s)\, dW_s.$$

Consequently, on the event G_n,

$$\int_0^t V_s \mathbf{1}_{[0,\tau_{n+m}]}(s)\, dW_s = \int_0^t V_s \mathbf{1}_{[0,\tau_n]}(s)\, dW_s$$

for all $m = 1, 2, \ldots$. Therefore, the integrals stabilize on the event G_n, for all $t \leq T$. Since the events G_n converge up to an event of probability one, it follows that the integrals stabilize a.s. Continuity in t follows because each of the approximating integrals is continuous. □

Caution: The Itô integral defined by (27) does not share all of the properties of the Itô integral for integrands of class $\mathcal{V}_T$. In particular, the integrals may not have finite first moments; hence they are no longer necessarily martingales; and there is no Itô isometry.

4 The Itô Formula

4.1 Itô formula for Wiener functionals

The cornerstone of stochastic calculus is the Itô Formula, the stochastic analogue of the Fundamental Theorem of (ordinary) calculus. The simplest form is this:

Theorem 1. *(Univariate Itô Formula) Let $u(t,x)$ be twice continuously differentiable in x and once continuously differentiable in t. If W_t is a standard Wiener process, then*

$$\boxed{u(t,W_t)-u(0,0)=\int_0^t u_s(s,W_s)\,ds+\int_0^t u_x(s,W_s)\,dW_s+\frac{1}{2}\int_0^t u_{xx}(s,W_s)\,ds.} \tag{28}$$

Proof. See section 4.5 below. □

One of the reasons for developing the Itô integral for filtrations larger than the minimal filtration is that this allows us to use the Itô calculus for functions and processes defined on several independent Wiener processes. Recall that a $k-$dimensional Wiener process is an $\mathbb{R}^k-$vector-valued process

$$\mathbf{W}(t)=(W_1(t),W_2(t),\dots,W_k(t)) \tag{29}$$

whose components $W_i(t)$ are mutually independent one-dimensional Wiener processes. Assume that $\mathbb{F}$ is a filtration that is admissible for each component process, that is, such that each process $W_i(t)$ is a martingale relative to $\mathbb{F}$. Then a progressively measurable process V_t relative to $\mathbb{F}$ can be integrated against any one of the Wiener processes $W_i(t)$. If $\mathbf{V}(t)$ is itself a vector-valued process each of whose components $V_i(t)$ is in the class $\mathcal{W}_T$, then define

$$\int_0^t \mathbf{V}\cdot d\mathbf{W}=\sum_{i=1}^k\int_0^t V_i(s)\,dW_i(s) \tag{30}$$

When there is no danger of confusion I will drop the boldface notation.

Theorem 2. *(Multivariate Itô Formula) Let $u(t,\mathbf{x})$ be twice continuously differentiable in each x_i and once continuously differentiable in t. If W_t is a standard $k-$dimensional Wiener process, then*

$$\boxed{u(t,\mathbf{W}_t)-u(0,\mathbf{0})=\int_0^t u_s(s,\mathbf{W}_s)\,ds+\int_0^t \nabla_{\mathbf{x}}u(s,\mathbf{W}_s)\cdot\,d\mathbf{W}_s+\frac{1}{2}\int_0^t \Delta_{\mathbf{x}}u(s,\mathbf{W}_s)\,ds.} \tag{31}$$

Here $\nabla_{\mathbf{x}}$ and $\Delta_{\mathbf{x}}$ denote the gradient and Laplacian operators in the $\mathbf{x}-$variables, respectively.

Proof. This is essentially the same as the proof in the univariate case. □

Example 2. First consider the case of one variable, and let $u(t,x)=x^2$. Then $u_{xx}=2$ and $u_t=0$, and so the Itô formula gives another derivation of formula (23). Actually, the Itô formula will be proved in general by mimicking the derivation that led to (23), using a two-term Taylor series approximation for the increments of $u(t,W)_t)$ over short time intervals. □

Example 3. (Exponential Martingales.) Fix $\theta \in \mathbb{R}$, and let $u(t,x) = \exp\{\theta x - \theta^2 t/2\}$. It is readily checked that $u_t + u_{xx}/2 = 0$, so the two ordinary integrals in the Itô formula cancel, leaving just the stochastic integral. Since $u_x = \theta u$, the Itô formula gives

$$Z^\theta(t) = 1 + \int_0^t \theta Z^\theta(s)\, dW(s) \tag{32}$$

where

$$Z^\theta(t) := \exp\left\{\theta W(t) - \theta^2 t/2\right\}.$$

Thus, the exponential martingale $Z_\theta(t)$ is a solution of the linear stochastic differential equation $dZ_t = \theta Z_t\, dW_t$. □

Example 4. A function $u : \mathbb{R}^k \to \mathbb{R}$ is called *harmonic* in a domain D (an open subset of $\mathbb{R}^k$) if it satisfies the Laplace equation $\Delta u = 0$ at all points of D. Let u be a harmonic function on $\mathbb{R}^k$ that is twice continuously differentiable. Then the multivariable Itô formula implies that if W_t is a $k-$dimensional Wiener process,

$$u(W_t) = u(W_0) + \int_0^t \nabla u(W_s)\, dW_s.$$

It follows, by localization, that if τ is a stopping time such that $\nabla u(W_s)$ is bounded for $s \le \tau$ then $u(W_{t\wedge\tau})$ is an L^2 martingale. □

Exercise 5. Check that in dimension $d \ge 3$ the *Newtonian potential* $u(x) = |x|^{-d+2}$ is harmonic away from the origin. Check that in dimension $d = 2$ the *logarithmic potential* $u(x) = \log|x|$ is harmonic away from the origin.

4.2 Itô processes

An *Itô process* is a solution of a stochastic differential equation. More precisely, an Itô process is an $\mathbb{F}-$progressively measurable process X_t that can be represented as

$$X_t = X_0 + \int_0^t A_s\, ds + \int_0^t V_s \cdot dW_s \quad \forall\, t \le T, \tag{33}$$

or equivalently, in differential form,

$$dX(t) = A(s)\, ds + V(s) \cdot dW(t). \tag{34}$$

Here $W(t)$ is a $k-$dimensional Wiener process; $V(s)$ is a $k-$dimensional vector-valued process with components $V_i \in \mathcal{W}_T$; and A_t is a progressively measurable process that is integrable (in t) relative to Lebesgue measure with probability 1, that is,

$$\int_0^T |A_s|\, ds < \infty \quad \text{a.s.} \tag{35}$$

If $X_0 = 0$ and the integrand $A_s = 0$ for all s, then call $X_t = I_t(V)$ an *Itô integral process*. Note that every Itô process has (a version with) continuous paths. Similarly, a *$k-$dimensional Itô* *process* is a vector-valued process X_t with representation (33) where U, V are vector-valued and W is a $k-$dimensional Wiener process relative to $\mathbb{F}$. (Note: In this case $\int V\, dW$ must be interpreted as $\int V \cdot dW$.) If X_t is an Itô process with representation (33) (either univariate or multivariate), its *quadratic variation* is defined to be the process

$$[X]_t := \int_0^t |V_s|^2 \, ds. \tag{36}$$

If $X_1(t)$ and $X_2(t)$ are Itô processes relative to the same driving $d-$dimensional Wiener process, with representations (in differential form)

$$dX_i(t) = A_i(s)\, ds + \sum_{j=1}^{d} V_{ij}(t)\, dW(t), \tag{37}$$

then the *quadratic covariation* of X_1 and X_2 is defined by

$$d[X_i, X_j]_t := \sum_{l=1}^{d} V_{il}(t) V_{jl}(t)\, dt. \tag{38}$$

Theorem 3. *(Univariate Itô Formula) Let $u(t,x)$ be twice continuously differentiable in x and once continuously differentiable in t, and let $X(t)$ be a univariate Itô process. Then*

$$\boxed{du(t, X(t)) = u_t(t, X(t))\, dt + u_x(t, X(t))\, dX(t) + \frac{1}{2} u_{xx}(t, X(t))\, d[X]_t}. \tag{39}$$

Note: It should be understood that the differential equation in (39) is shorthand for an integral equation. Since u and its partial derivatives are assumed to be continuous, the ordinary and stochastic integrals of the processes on the right side of (39) are well-defined up to any finite time t. The differential $dX(t)$ is interpreted as in (34).

Theorem 4. *(Multivariate Itô Formula) Let $u(t,x)$ be twice continuously differentiable in $x \in \mathbb{R}^k$ and once continuously differentiable in t, and let $X(t)$ be a $k-$dimensional Itô process whose components $X_i(t)$ satisfy the stochastic differential equations* (37)*. Then*

$$\boxed{du(t, X(t)) = u_s(s, X(s))\, ds + \nabla_x u(s, X_s)\, dX(s) + \frac{1}{2} \sum_{i,j=1}^{k} u_{x_i, x_j}(s, X(s))\, d[X_i, X_j](s)} \tag{40}$$

Note: Unlike the Multivariate Itô Formula for functions of Wiener processes (Theorem 2 above), this formula includes mixed partials.

Theorems3 and **??** can be proved by similar reasoning as in sec. 4.5 below. Alternatively, they can be deduced as special cases of the general Itô formula for Itô integrals relative to continuous local martingales. (See notes on web page.)

4.3 Example: Ornstein-Uhlenbeck process

Recall that the Ornstein-Uhlenbeck process with mean-reversion parameter $\alpha > 0$ is the mean zero Gaussian process X_t whose covariance function is $EX_sX_t = \exp\{-\alpha|t-s|\}$. This process is the continuous-time analogue of the autoregressive-1 process, and is a weak limit of suitably scaled *AR* processes. It occurs frequently as a weak limit of stochastic processes with some sort of mean-reversion, for much the same reason that the classical harmonic oscillator equation (Hooke's Law) occurs in mechanical systems with a restoring force. The natural stochastic analogue of the harmonic oscillator equation is

$$dX_t = -\alpha X_t \, dt + dW_t; \tag{41}$$

α is called the *relaxation parameter*. To solve equation (41), set $Y_t = e^{\alpha t} X_t$ and use the Itô formula along with (41) to obtain

$$dY_t = e^{\alpha t} \, dW_t.$$

Thus, for any initial value $X_0 = x$ the equation (41) has the unique solution

$$X_t = X_0 e^{-\alpha t} + e^{-\alpha t} \int_0^t e^{\alpha s} \, dW_s. \tag{42}$$

It is easily checked that the Gaussian process defined by this equation has the covariance function of the Ornstein-Uhlenbeck process with parameter α. Since Gaussian processes are determined by their means and covariances, it follows that the process X_t defined by (42) is a *stationary* Ornstein-Uhlenbeck process, provided the initial value X_0 is chosen to be a standard normal variate independent of the driving Brownian motion W_t.

4.4 Example: Brownian bridge

Recall that the standard Brownian bridge is the mean zero Gaussian process $\{Y_t\}_{0 \leq t \leq 1}$ with covariance function $EY_sY_t = s(1-t)$ for $0 < s \leq t < 1$. The Brownian bridge is the continuum limit of scaled simple random walk conditioned to return to 0 at time $2n$. But simple random walk conditioned to return to 0 at time $2n$ is equivalent to the random walk gotten by sampling *without replacement* from a box with n tickets marked $+1$ and n marked -1. Now if $S_{[nt]} = k$, then there will be an excess of k tickets marked -1 left in the box, and so the next step is a biased Bernoulli. This suggests that, in the continuum limit, there will be an instantaneous drift whose direction (in the (t, x) plane) points to $(1, 0)$. Thus, let W_t be a standard Brownian motion, and consider the stochastic differential equation

$$dY_t = -\frac{Y_t}{1-t} dt + dW_t \tag{43}$$

for $0 \leq t \leq 1$. To solve this, set $U_t = f(t)Y_t$ and use (43) together with the Itô formula to determine which choice of $f(t)$ will make the dt terms vanish. The answer is $f(t) = 1/(1-t)$ (easy exercise), and so

$$d((1-t)^{-1}Y_t) = (1-t)^{-1} \, dW_t.$$

Consequently, the unique solution to equation (43) with initial value $Y_0 = 0$ is given by

$$Y_t = (1-t)\int_0^t (1-s)^{-1}\,dW_s. \tag{44}$$

It is once again easily checked that the stochastic process Y_t defined by (44) is a mean zero Gaussian process whose covariance function matches that of the standard Brownian bridge. Therefore, the solution of (43) with initial condition $Y_0 = 0$ is a standard Brownian bridge.

4.5 Proof of the univariate Itô formula

For ease of notation, I will consider only the case where the driving Wiener process is $1-$dimensional; the argument in the general case is similar. First, I claim that it suffices to prove the result for functions u with compact support. This follows by a routine argument using the Stopping Rule and Localization Principle for Itô integrals: let D_n be an increasing sequence of open sets in $\mathbb{R}_+ \times \mathbb{R}$ that exhaust the space, and let τ_n be the first time that X_t exits the region D_n. Then by continuity, $u(t\wedge\tau_n, X(t\wedge\tau_n)) \to u(t, X(t))$ as $n \to \infty$, and

$$\int_0^{\tau\wedge\tau_n} \longrightarrow \int_0^t$$

for each of the integrals in (39). Thus, the result (39) will follows from the corresponding formula with t replaced by $t\wedge\tau_n$. For this, the function u can be replaced by a function $\tilde{u}$ with compact support such that $\tilde{u} = u$ in D_n, by the Localization Principle. (Note: the Localization Lemma in the Appendix to the notes on harmonic functions implies that there is a C^∞ function ψ_n with compact support that takes values between 0 and 1 and is identically 1 on D_n. Set $\hat{u} = u\psi_n$ to obtain a C^2 function with compact support that agrees with u on D_n.) Finally, if (39) can be proved with u replaced by $\tilde{u}$, then it will hold with t replaced by $t\wedge\tau_n$, using the Stopping Rule again. Thus, we may now assume that the function u has compact support, and therefore that it and its partial derivatives are bounded and *uniformly* continuous.

Second, I claim that it suffices to prove the result for *elementary* Itô processes, that is, processes X_t of the form (33) where V_s and A_s are *bounded, elementary* processes. This follows by a routine approximation argument, because any Itô process $X(t)$ can be approximated by elementary Itô processes.

It remains to prove (39) for elementary Itô processes X_t and functions u of compact support. Assume that X has the form (33) with

$$A_s = \sum \zeta_j \mathbf{1}_{[t_j,t_{j+1}]}(s),$$
$$V_s = \sum \xi_j \mathbf{1}_{[t_j,t_{j+1}]}(s)$$

where ζ_j, ξ_j are bounded random variables, both measurable relative to $\mathcal{F}_{t_j}$. Now it is clear that to prove the Itô formula (39) it suffices to prove it for $t \in (t_j, t_{j+1})$ for each index j. But this is essentially the same as proving that for an elementary Itô process X of the form

(33) with $A_s = \zeta \mathbf{1}_{[a,b]}(s)$ and $V_s = \xi \mathbf{1}_{[a,b]}(s)$, and ζ, ξ measurable relative to $\mathcal{F}_a$,

$$u(t, X_t) - u(a, X_a) = \int_a^t u_s(s, X_s)\, ds + \int_a^t u_x(s, X_s)\, dX_s + \frac{1}{2} \int_a^t u_{xx}(s, X_s)\, d[X]_s$$

for all $t \in (a, b)$. Fix $a < t$ and set $T = t - a$. Define

$$\begin{aligned}
&\Delta_k^n X := X(a + (k+1)T/2^n) - X(a + kT/2^n), \\
&\Delta_k^n U := U(a + (k+1)T/2^n) - U(a + kT/2^n), \quad \text{where} \\
&U(s) := u(s, X(s));\ U_s(s) = u_s(s, X(s));\ U_x(s) = u_x(s, X(s));\ \text{etc.}
\end{aligned}$$

Notice that because of the assumptions on A_s and V_s,

$$\Delta_k^n X = \zeta 2^{-n} T^{-1} + \xi \Delta_k^n W \tag{45}$$

Now by Taylor's theorem,

$$\begin{aligned}
u(t, X_t) - u(a, X_a) = &\sum_{k=0}^{2^{n-1}} \Delta_k^n U \qquad (46) \\
= &2^{-n} T^{-1} \sum_{k=0}^{2^{n-1}} U_s(kT/2^n) + \sum_{k=0}^{2^{n-1}} U_x(kT/2^n) \Delta_k^n X \\
&+ \sum_{k=0}^{2^{n-1}} U_{xx}(kT/2^n)(\Delta_k^n X)^2/2 + \sum_{k=0}^{2^{n-1}} R_k^n
\end{aligned}$$

where the remainder term R_k^n satisfies

$$|R_k^n| \le \varepsilon_n (2^{-2n} + (\Delta_k^n X)^2) \tag{47}$$

and the constants ε_n converge to zero as $n \to \infty$. (Note: This uniform bound on the remainder terms follows from the assumption that $u(s, x)$ is $C^{1 \times 2}$ and has compact support, because this ensures that the partial derivatives u_s and u_{xx} are uniformly continuous and bounded.)

Finally, let's see how the four sums on the right side of (46) behave as $n \to \infty$. First, because the partial derivative $u_s(s, x)$ is uniformly continuous and bounded, the first sum is just a Riemann sum approximation to the integral of a continuous function; thus,

$$\lim_{n \to \infty} 2^{-n} T^{-1} \sum_{k=0}^{2^{n-1}} U_s(kT/2^n) = \int_a^t u_s(s, X_s)\, ds.$$

Next, by (45), the second sum also converges, since it can be split into a Riemann sum for a Riemann integral and an elementary approximation to an Itô integral:

$$\lim_{n \to \infty} \sum_{k=0}^{2^{n-1}} U_x(kT/2^n) \Delta_k^n X = \int_a^t u_x(s, X_s)\, dX_s.$$

The third sum is handled using Proposition 8 on the quadratic variation of the Wiener process, and equation (45) to reduce the quadratic variation of X to that of W (Exercise: Use the fact that u_{xx} is uniformly continuous and bounded to fill in the details):

$$\lim_{n\to\infty}\sum_{k=0}^{2^{n-1}} U_{xx}(kT/2^n)(\Delta_k^n X)^2/2 = \frac{1}{2}\int_a^t u_{xx}(s, X_s)d[X]_s.$$

Finally, by (47) and Proposition 8,

$$\lim_{n\to\infty}\sum_{k=0}^{2^{n-1}} R_k^n = 0.$$

□

5 Complex Exponential Martingales and their Uses

Assume in this section that $W_t = (W_t^1, W_t^2, \ldots, W_t^d)$ is a $d-$dimensional Brownian motion started at the origin, and let $\mathrm{F} = \{\mathcal{F}_t\}_{t\geq 0}$ be the minimal filtration.

5.1 Exponential Martingales

Let V_t be a progressively measurable, $d-$dimensional process such that for each $T < \infty$ the process V_t is in the class $\mathcal{W}_T$, that is,

$$P\left\{\int_0^T \|V_s\|^2\,ds < \infty\right\} = 1.$$

Then the Itô formula (39) implies that the (complex-valued) exponential process

$$Z_t := \exp\left\{i\int_0^t V_s\cdot\,dW_s + \frac{1}{2}\int_0^t \|V_s\|^2\,ds\right\},\quad t\leq T, \tag{48}$$

satisfies the stochastic differential equation

$$dZ_t = iZ_tV_t\cdot\,dW_t. \tag{49}$$

This alone does not guarantee that the process Z_t is a martingale, because without further assumptions the integrand Z_tV_t might not be in the class $\mathcal{V}_T^2$. Of course, if the integrand V_t is uniformly bounded for $t\leq T$ then so is Z_t, and so the stochastic differential equation (49) exhibits Z_t as the Itô integral of a process in $\mathcal{V}_T^2$, which implies that $\{Z_t\}_{t\leq T}$ is a martingale. This remains true under weaker hypotheses on V_t:

Proposition 11. *Assume that for each $T < \infty$,*

$$E\exp\left\{\int_0^T \|V_s\|^2\,ds\right\} < \infty \tag{50}$$

Then the process Z_t defined by (48) *is a martingale, and in particular,*

$$EZ_T = 1\quad\textit{for each}\;\; T<\infty. \tag{51}$$

Proof. Set

$$X_t = X(t) = \int_0^t V_s \cdot dW_s,$$

$$[X]_t = [X](t) = \int_0^t \|V_s\|^2 \, ds, \quad \text{and}$$

$$\tau_n = \inf\{t \, : \, [X]_t = n\}$$

Since $Z(t \wedge \tau_n) V(t \wedge \tau_n)$ is uniformly bounded for each n, the process $Z(t \wedge \tau_N)$ is a bounded martingale. But

$$|Z(t \wedge \tau_n)| = \exp\{iX(t \wedge \tau_n) + [X](t \wedge \tau_n)/2\} \leq \exp\{[X]_t/2\},$$

so the random variables $Z(t \wedge \tau_n)$ are all dominated by the L^1 random variable $\exp\{[X]_t/2\}$ (note that the hypothesis (50) is the same as the assertion that $\exp\{[X]_t/2\}$ has finite first moment). Therefore the dominated convergence theorem for conditional expectations implies that the process $Z(t)$ is a martingale. □

5.2 Radial part of a $d-$dimensional Brownian motion

Proposition 12. *If Θ_t is any progressively measurable process taking values in the unit sphere of $\mathbb{R}^d$ then the Itô process*

$$X_t = X(t) = \int_0^t \Theta_s \cdot dW_s$$

is a standard one-dimensional Brownian motion.

Proof. Since X_t has continuous paths (recall that all Itô integral processes do), it suffices to show that X_t has stationary, independent increments and that the marginal distribution of X_t is Normal-$(0, t)$. Both of these tasks can be done simultaneously via characteristic functions (Fourier transforms), by showing that for all choices $0 = t_0 < t_1 < \cdots < t_k$ and all $\beta_1, \beta_2, \ldots, \beta_k \in \mathbb{R}$,

$$E \exp\left\{ \sum_{j=1}^k i\beta_j (X(t_j) - X(t_{j-1})) + \sum_{j=1}^k \beta_j^2 (t_j - t_{i-1})/2 \right\} = 1.$$

Define

$$\begin{aligned} V_t &= \beta_j \Theta(t) \quad \text{if } \ t_{j-1} < t \leq t_j \quad \text{and} \\ V_t &= 0 \quad \text{if } \ t > t_k; \end{aligned}$$

then

$$E \exp\left\{ \sum_{j=1}^k i\beta_j (X(t_j) - X(t_{j-1})) + \sum_{j=1}^k \beta_j^2 (t_j - t_{i-1})/2 \right\}$$
$$= E \exp\left\{ i \int_0^{t_k} V_s \, dW_s + \frac{1}{2} \int_0^{t_k} V_s^2 \, ds \right\}.$$

The process V_t is clearly bounded in norm (by $\max|\beta_i|$), so Proposition **??** implies that the process

$$Z_t := \exp\left\{i \int_0^t V_s\, dW_s + \frac{1}{2}\int_0^t V_s^2\, ds\right\}$$

is a martingale, and it follows that

$$EZ_{t_k} = 1.$$

□

A *Bessel process* with dimension parameter d is a solution (or a process whose law agrees with that of a solution) of the stochastic differential equation

$$dX_t = \frac{d-1}{2X_t}\, dt + dW_t, \tag{52}$$

where W_t is a standard one-dimensional Brownian motion. The problems of existence and uniqueness of solutions to the Bessel SDEs (52) will be addressed later. The next result shows that solutions exist when d is a positive integer.

Proposition 13. *Let $W_t = (W_t^1, W_t^2, \dots, W_t^d)$ be a $d-$dimensional Brownian motion started at a point $x \neq 0$, and let $R_t = |W_t|$ be the modulus of W_t. Then R_t is a Bessel process of dimension d started at $|x|$.*

Proof. The process R_t is gotten by applying a smooth (everywhere except at the origin) real-valued function to $d-$dimensional Brownian motion. Since $d-$dimensional Brownian motion started at a point $x \neq 0$ will never visit the origin, Ito's formula applies, and (after a brief adventure in multivariate calculus) shows that for any $t > 0$

$$R_t - R_0 = \int_\varepsilon^t \Theta_s \cdot dW_s + \int_0^t (d-1)/(2R_s)\, ds$$

By Proposition 12, the first integral determines a standard, one-dimensional Brownian motion. □

Remark 1. In section 6 we will use the stochastic differential equation (52) to show that the law of one-dimensional Brownian motion conditioned to stay positive forever coincides with that of the radial part R_t of a $3-$dimensional Brownian motion.

5.3 Time Change for Itô Integral Processes

Let $X_t = I_t(V)$ be an Itô integral process, where the integrand V_s is a $d-$dimensional, progressively measurable process in the class $\mathcal{W}_T$. One should interpret the magnitude $|V_t|$ as representing instantaneous *volatility* – in particular, the conditional distribution of the increment $X(t+\delta t) - X(t)$ given the value $|V_t| = \sigma$ is approximately, for $\delta t \to 0$, the normal distribution with mean zero and variance $\sigma\delta t$. One may view this in one of two ways: (1) the volatility $|V_t|$ is a *damping* factor – that is, it multiplies the next Wiener increment by $|V_t|$; alternatively, (2) $|V_t|$ is a time regulation factor, either slowing or speeding the normal rate at which variance is accumulated. The next theorem makes this latter viewpoint precise:

Theorem 5. *Every Itô integral process is a time-changed Wiener process. More precisely, let $X_t = I_t(V)$ be an Itô integral process with quadratic variation $[X]_t = \int_0^t |V_s|\,ds$. Assume that $[X]_t < \infty$ for all $t < \infty$, but that $[X]_\infty = \lim_{t\to\infty}[X]_t = \infty$ with probability one. For each $s \geq 0$ define*

$$\tau(s) = \inf\{t > 0 \,:\, [X]_t = s\}. \tag{53}$$

Then the process

$$\tilde{W}(s) = X(\tau(s)), \quad s \geq 0 \tag{54}$$

is a standard Wiener process.

Note: A more general theorem of P. LÉVY asserts that *every* continuous martingale (not necessarily adapted to a Wiener filtration) is a time-changed Wiener process.

Proof. The strategy is the same as in the proof of Proposition 12: since $\tilde{W}_s$ has continuous paths, it suffices to show that it has stationary, independent increments, and that the marginal distribution of $\tilde{W}_t$ is, for any $t > 0$, normal with mean 0 and variance t.

By Proposition 11, for any $\theta \in \mathbb{R}$ and any $s > 0$ the process

$$Z_\theta(t) := \exp\{i\theta X(t \wedge \tau(s)) + \theta^2 [X]_{t\wedge\tau(s)}/2\}$$

is a bounded martingale. Hence, by the Bounded Convergence Theorem,

$$E\exp\{i\theta X(\tau(s)) + \theta^2 s/2\} = 1.$$

This implies that $\tilde{W}(s) = X(\tau(s))$ has the characteristic function of the $N(0, s)$ distribution. A variation of this argument (Exercise!) shows that the process $\tilde{W}(s)$ has stationary, independent increments. □

Corollary 3. *Let $X_t = I_t(V)$ be an Itô integral process with quadratic variation $[X]_t$, and let $\tau(s)$ be defined by* (53). *Then for each $\alpha > 0$,*

$$P\{\max_{t\leq\tau(s)} |X_t| \geq \alpha\} \leq 2P\{W_s \geq \alpha\}. \tag{55}$$

Consequently, the random variable $\max_{t\leq\tau(s)} |X_t|$ has finite moments of all order, and even a finite moment generating function.

Proof. The maximum of $|X(t)|$ up to time $\tau(s)$ coincides with the maximum of the Wiener process $\tilde{W}$ up to time s, so the result follows from the Reflection Principle. □

5.4 Itô Representation Theorem

Theorem 6. *Let W_t be a $d-$dimensional Wiener process and let $\mathbb{F}^W = (\mathcal{F}_t^W)_{0\leq t<\infty}$ be the minimal filtration for W_t. Then for any $\mathcal{F}_T^W-$measurable random variable Y with mean zero and finite variance, there exists a ($d-$dimensional vector-valued) process V_t in the class $\mathcal{V}_T$ such that*

$$Y = I_T(V) = \int_0^T V_s \cdot dW_s. \tag{56}$$

Corollary 4. *If $\{M_t\}_{t \leq T}$ is an L^2- bounded martingale relative to the minimal filtration $\mathbb{F}^W$ of a Wiener process, then $M_t = I_t(V)$ a.s. for some process V_t in the class $\mathcal{V}_T$, and consequently M_t has a version with continuous paths.*

Proof of the Corollary. Assume that $T < \infty$; then $Y := M_T$ satisfies the hypotheses of Theorem 6, and hence has representation (56). For any integrand V_t of class $\mathcal{V}_T$, the Itô integral process $I_t(V)$ is an L^2-martingale, and so by (56),

$$M_t = E(M_T \,|\, \mathcal{F}_t^W) = E(Y \,|\, \mathcal{F}_t^W) = I_t(V) \quad \text{a.s.}$$

$\square$

Proof of Theorem 6. It is enough to consider random variables Y of the form

$$Y = f(W(t_1), W(t_2), \dots, W(t_I)) \tag{57}$$

because such random variables are dense in $L^2(\Omega, \mathcal{F}_T^W, P)$. If there is a random variable Y of the form (57) that is *not* a stochastic integral, then (by orthogonal projection) there exists such a Y that is uncorrelated with every Y' of the form (57) that *is* a stochastic integral. I will show that if Y is a mean zero, finite-variance random variable of the form (57) that is uncorrelated with every random variable Y' of the same form that is also a stochastic integral, then $Y = 0$ a.s. By (49) above (or alternatively see Example 3 in section 4), for all $\theta_j \in \mathbb{R}^d$ the random variable

$$Y' = \exp\left\{\sum_{j=1}^{J} \theta_j W(t_j)\right\}$$

is a stochastic integral. Clearly, Y' has finite variance. Thus, by hypothesis, for all $\theta_j \in \mathbb{R}^d$,

$$Ef(W(t_1), W(t_2), \dots, W(t_J)) \exp\left\{\sum_{j=1}^{J} i\langle\theta_j, W(t_j)\rangle\right\} = 0. \tag{58}$$

This implies that the random variable Y must be 0 with probability one. To see this, consider the (signed) measure μ on $\mathbb{R}^{Jd}$ defined by

$$d\mu(x) = f(x)P\{(W(t_1), W(t_2), \dots, W(t_J)) \in dx\};$$

then equation (58) implies that the Fourier transform of μ is identically 0, and this in turn implies that $\mu = 0$. $\square$

Exercise 6. Does an L^1-bounded martingale $\{M_t\}_{t \leq T}$ necessarily have a version with continuous paths?

5.5 Hermite Functions and Hermite Martingales

The *Hermite functions* $H_n(x,t)$ are polynomials in the variables x and t that satisfy the backward heat equation $H_t + H_{xx}/2 = 0$. As we have seen (e.g., in connection with the exponential function $\exp\{\theta x - \theta^2 t/2\}$), if $H(x,t)$ satisfies the backward heat equation, then when the Itô Formula is applied to $H(W_t,t)$, the ordinary integrals cancel, leaving only the Itô integral; and thus, $H(W_t,t)$ is a (local) martingale. Consequently, the Hermite functions provide a sequence of polynomial martingales. The first few Hermite functions are

$$\begin{aligned} H_0(x,t) &= 1, \\ H_1(x,t) &= x, \\ H_2(x,t) &= x^2 - t, \\ H_3(x,t) &= x^3 - 3xt, \\ H_4(x,t) &= x^4 - 6x^2t + 3t^2. \end{aligned} \tag{59}$$

The formal definition is by a generating function:

$$\sum_{n=0}^{\infty} H_n(x,t)\frac{\theta^n}{n!} = \exp\{\theta x - \theta^2 t/2\} \tag{60}$$

Exercise 7. Show that the Hermite functions satisfy the two-term recursion relation $H_{n+1} = xH_n - ntH_{n-1}$. Conclude that every term of H_{2m} is a constant times $x^{2m}t^{n-m}$ for some $0 \leq m \leq n$, and that the lead term is the monomial x^{2n}. Conclude also that each H_n solves the backward heat equation.

Proposition 14. *Let V_s be a* bounded, *progressively measurable process, and let $X_t = I_t(V)$ be the Itô integral process with integrand V. Then for each $n \geq 0$, the processes $H_n(X_t, [X]_t)$ is a martingale.*

Proof. Since each $H_n(x,t)$ satisfies the backward heat equation, the Itô formula implies that $H(X_{t\wedge\tau}, [X]_{t\wedge\tau})$ is a martingale for each stopping time $\tau = \tau(m) =$ first t such that either $|X_t| = m$ or $[X]_t = m$. If V_s is bounded, then Corollary 3 guarantees that for each t the random variables $H(X_{t\wedge\tau(m)}, [X]_{t\wedge\tau(m)})$ are dominated by an integrable random variable. Therefore, the DCT for conditional expectations implies that $H_n(X_t, [X]_t)$ is a martingale. □

5.6 Moment Inequalities

Corollary 5. *Let V_t be a progressively measurable process in the class $\mathcal{W}_T$, and let $X_t = I_t(V)$ be the associated Itô integral process. Then for every integer $m \geq 1$ and every time $T < \infty$, there exist constants $C_m < \infty$ such that*

$$EX_T^{2m} \leq C_m E[X]_T^m. \tag{61}$$

Note: Burkholder, Davis, and Gundy have proved maximal inequalities that are considerably stronger than this, but the arguments are not elementary.

Proof. First, it suffices to consider the case where the integrand V_s is uniformly bounded. To see this, define in general the truncated integrands $V_s^{(m)} := V(s) \wedge m$; then

$$\lim_{m\to\infty} I_t(V^{(m)}) = I_t(V) \quad \text{a.s., and}$$
$$\lim_{m\to\infty} \uparrow [I(V^{(m)})]_t = [I(V)]_t.$$

Hence, if the result holds for each of the truncated integrands $V^{(m)}$, then it will hold for V, by Fatou's Lemma and the Monotone Convergence Theorem.

Assume, then, that V_s is uniformly bounded. The proof of (61) in this case is by induction on m. First note that, because V_t is assumed bounded, so is the quadratic variation $[X]_T$ at any finite time T, and so $[X]_T$ has finite moments of all orders. Also, if V_t is bounded then it is an element of $\mathcal{V}_T$, and so the Itô isometry implies that

$$EX_T^2 = E[X]_T.$$

This takes care of the case $m = 1$. The induction argument uses the Hermite martingales $H_{2m}(X_t, [X]_t)$ (Proposition 14). By Exercise 7, the lead term (in x) of the polynomial $H_{2m}(x,t)$ is x^{2m}, and the remaining terms are all of the form $a_{m,k}x^{2k}t^{m-k}$ for $k < m$. Since $H_{2m}(X_0, [X]_0) = 0$, the martingale identity implies

$$\begin{aligned}
EX_T^{2m} &= -\sum_{k=0}^{m-1} a_{m,k} EX_T^{2k}[X]_T^{m-k} \quad \Longrightarrow \\
EX_T^{2m} &\leq A_{2m} \sum_{k=0}^{m-1} EX_T^{2k}[X]_T^{m-k} \quad \Longrightarrow \\
EX_T^{2m} &\leq A_{2m} \sum_{k=0}^{m-1} (EX_T^{2m})^{k/m} (E[X]_T^m)^{1-k/m},
\end{aligned}$$

the last by Holder's inequality. Note that the constants $A_{2m} = \max |a_{m,k}|$ are determined by the coefficients of the Hermite function H_{2m}. Now divide both sides by EX_T^{2m} to obtain

$$1 \leq A_{2m} \sum_{k=0}^{m-1} \left(\frac{E[X]_T^m}{EX_T^{2m}} \right)^{1-k/m}.$$

The inequality (61) follows. □

6 Girsanov's Theorem

6.1 Change of measure

Let $(\Omega, \{\mathcal{F}_t\}_{t\geq 0}, P)$ be a filtered probability space. If Z_T is a nonnegative, $\mathcal{F}_T-$ measurable random variable with expectation 1 then it is a *likelihood ratio*, that is, the measure Q on $\mathcal{F}_T$ defined by

$$Q(F) := E_P \mathbf{1}_F Z_T \tag{62}$$

is a *probability* measure, and the likelihood ratio (Radon-Nikodym derivative) of Q relative to P is Z_T. It is of interest to know how the measure Q restricts to the $\sigma-$algebra $\mathcal{F}_t \subset \mathcal{F}_T$ when $t < T$.

Proposition 15. *Let Z_T be an $\mathcal{F}_T-$measurable, nonnegative random variable such that $E^P Z_T = 1$, and let $Q = Q_T$ be the probability measure on $\mathcal{F}_T$ with likelihood ratio Z_T relative to P. Then for any $t \in [0, T]$ the restriction Q_t of Q to the $\sigma-$algebra $\mathcal{F}_t$ has likelihood ratio*

$$Z_t := E^P(Z_T|\mathcal{F}_t). \tag{63}$$

Proof. The random variable Z_t defined by (63) is nonnegative, $\mathcal{F}_t-$measurable, and integrates to 1, so it is the likelihood ratio of a probability measure on $\mathcal{F}_t$. For any event $A \in \mathcal{F}_t$,

$$Q(A) := E^P Z_T \mathbf{1}_A = E^P E^P(Z_T|\mathcal{F}_t)\mathbf{1}_A$$

by definition of conditional expectation, so

$$Z_t = (dQ/dP)_{\mathcal{F}_t}.$$

□

6.2 Example: Brownian motion with drift

Assume now that W_t is a standard one-dimensional Brownian motion, with admissible filtration $\{\mathcal{F}_t\}_{t\geq 0}$ and let $\theta \in \mathbb{R}$ be a fixed constant. Recall that the exponential process

$$Z_t^\theta := \exp\{\theta W_t - \theta^2 t/2\} \tag{64}$$

is a nonnegative martingale with initial value 1. Thus, for each $T < \infty$ the random variable Z_T is the likelihood ratio of a probability measure $Q = Q_T^\theta$ on $\mathcal{F}_T$, defined by (62).

Theorem 7. *Under $Q = Q_T^\theta$, the process $\{W_t\}_{t\leq T}$ is a Brownian motion with drift θ, equivalently, the process $\{\tilde{W}_t = W_t - \theta t\}_{t\leq T}$ is a standard Brownian motion.*

Proof. Under P the process $W_t - \theta t$ has continuous paths, almost surely. Since Q is absolutely continuous with respect to P, it follows that under Q the process $W_t - \theta t$ has continuous paths. Thus, to show that this process is a Brownian motion it is enough to show that it has the right finite-dimensional distributions. This can be done by calculating the joint moment generating functions of the increments: fix $0 = t_0 < t_1 < \cdots < t_J = T$, and set

$$\begin{aligned}
\Delta t_j &= t_j - t_{j-1}, \\
\Delta W_j &= W(t_j) - W(t_{j-1}), \\
\Delta \tilde{W}_j &= \tilde{W}(t_j) - \tilde{W}(t_{j-1});
\end{aligned}$$

then

$$\begin{aligned}
E_Q \exp\left\{\sum_{i=1}^J \alpha_j \Delta\tilde{W}_j\right\} &= \exp\left\{-\theta\sum_{j=1}^J \alpha_j \Delta t_j\right\} E_P \exp\left\{\sum_{i=1}^J (\alpha_j+\theta)\Delta W_j - \theta^2 T/2\right\} \\
&= \exp\left\{-\theta^2 T/2 - \theta\sum_{j=1}^J \alpha_j \Delta t_j\right\} E_P \exp\left\{\sum_{i=1}^J (\alpha_j+\theta)\Delta W_j\right\} \\
&= \exp\left\{-\theta^2 T/2 - \theta\sum_{j=1}^J \alpha_j \Delta t_j\right\} \exp\left\{\sum_{i=1}^J (\alpha_j+\theta)^2\Delta t_j/2\right\} \\
&= \exp\left\{\sum_{i=1}^J \alpha_j^2 \Delta t_j/2\right\}
\end{aligned}$$

□

By Proposition 15, the family of measures $\{Q_T^\theta\}_{T\geq 0}$ is consistent in the sense that the restriction of Q_{T+S} to the $\sigma-$algebra $\mathcal{F}_T$ is just Q_T. It is a routine exercise in measure theory to show that these measures extend to a probability measure $Q^\theta = Q^\theta_\infty$ on the smallest $\sigma-$algebra $\mathcal{F}_\infty$ that contains $\cup_{t\geq 0}\mathcal{F}_t$. It is important to note that, although each Q_T is absolutely continuous relative to P_T, the extension Q_∞ is *singular* relative to P_∞: this is because the strong law of large numbers for sums of i.i.d. standard normals implies that

$$\begin{aligned}
W_n/n &\to 0 \quad \text{almost surely } P \quad \text{but} \\
W_n/n &\to \theta \quad \text{almost surely } Q.
\end{aligned}$$

6.3 The Girsanov formula

The Girsanov theorem is a far-reaching extension of Theorem 7 that describes the change of measure needed to transform Brownian motion to Brownian motion plus a *progressively measurable* drift. Let $(\Omega, \mathcal{F}\}, P)$ be a probability space that supports a $d-$dimensional Brownian motion W_t, and let $\{\mathcal{F}_t\}_{t\geq 0}$ be an admissible filtration. Let $\{V_t\}_{t\geq 0}$ be a progressively measurable process relative to the filtration, and assume that for each $T < \infty$,

$$P\left\{\int_0^T |V_s|^2\, ds < \infty\right\} = 1.$$

Then the Itô integral $I_t(V)$ is well-defined for every $t \geq 0$, and by a routine application of the Itô formula, the process

$$Z_t = \exp\left\{\int_0^t V_s \cdot\, dW_s - \int_0^t V_s^2\, ds/2\right\} = \exp\{I_t(V) - [I_t(V)]/2\} \tag{65}$$

satisfies the stochastic differential equation

$$dZ_t = Z_t V_t \cdot dW_t \quad \Longleftrightarrow \quad Z_t - Z_0 = \int_0^t Z_s V_s \cdot dW_s. \tag{66}$$

Proposition 16. *If the process $\{V_t\}_{t\geq 0}$ is bounded then $\{Z_t\}_{t\geq 0}$ is a martingale, and consequently, for each $T < \infty$, $EZ_T = 1$.*

Remark 2. The hypothesis that V_t be bounded can be weakened substantially: a theorem of Novikov asserts that Z_t is a martingale if for every $T < \infty$,

$$E \exp\left\{\frac{1}{2}\int_0^T |V_s|^2\, ds\right\} < \infty. \tag{67}$$

Proof of Proposition 16. Since the stochastic differential equation (66) has no dt term, it suffices to show that for each $T < \infty$ the process $Z_t V_t$ is in the integrability class $\mathcal{V}_T^2$, and since the process V_t is bounded, it suffices to show that for each $T < \infty$,

$$\int_0^T EZ_t^2\, dt < \infty. \tag{68}$$

Clearly,

$$Z_t^2 \leq \exp\left\{2\int_0^t V_s \cdot dW_s\right\}.$$

By the time-change theorem for Itô integral processes, the process $2I_t(V)$ in the last exponential is a time-changed Brownian motion, in particular, the process $\tilde{W}_s = I_{\tau(s)}(V)$ is a Brownian motion in s, where $\tau(s)$ is defined by (53). Because the process V_t is bounded, there exists $C < \infty$ such that the accumulated quadratic variation $[I_t(V)]$ is bounded by Ct, for all t. Consequently,

$$\int_0^t V_s \cdot dW_s \leq \max_{s\leq Ct} \tilde{W}_s := \tilde{M}_{Ct},$$

and so by Brownian scaling,

$$EZ_t^2 \leq E\exp\{2\sqrt{Ct}\tilde{M}_1\}.$$

By the reflection principle, this last expectation can be bounded as follows:

$$\begin{aligned} E\exp\{2\sqrt{Ct}\tilde{M}_1\} &= 2\int_0^\infty e^{2\sqrt{Ct}y} 2e^{-y^2/2}\, dy/\sqrt{2\pi} \\ &\leq 2\int_{-\infty}^\infty e^{2\sqrt{Ct}y} 2e^{-y^2/2}\, dy/\sqrt{2\pi} \\ &= 2\exp\{2Ct\}. \end{aligned}$$

It is now apparent that (68) holds. □

Proposition 16 asserts that if the process V_t is bounded then for each $T < \infty$ the random variable $Z(T)$ is a likelihood ratio, that is, a nonnegative random variable that integrates to 1. We have already noted that boundedness of V_t is not *necessary* for $EZ_T = 1$, which is all that is needed to ensure that

$$Q(F) = E_P(Z(T)\mathbf{1}_F) \tag{69}$$

defines a new probability measure on $(\Omega, \mathcal{F}_T)$. Girsanov's theorem describes the distribution of the stochastic process $\{W(t)\}_{t\geq 0}$ under this new probability measure. Define

$$\tilde{W}(t) = W(t) - \int_0^t V_s \, ds \tag{70}$$

Theorem 8. *(Girsanov) Assume that under P the process $\{W_t\}_{t\geq 0}$ is a $d-$dimensional Brownian motion with admissible filtration $\mathbb{F} = \{\mathcal{F}_t\}_{t\geq 0}$, and that the exponential process $\{Z_t\}_{t\leq T}$ defined by* (65) *is a martingale relative to $\mathbb{F}$ under P. Define Q on $\mathcal{F}_T$ by equation* (69)*. Then under the probability measure Q, the stochastic process $\left\{\tilde{W}(t)\right\}_{0\leq t\leq T}$ is a standard Wiener process.*

Proof. To show that the process $\tilde{W}_t$, under Q, is a standard Wiener process, it suffices to show that it has independent, normally distributed increments with the correct variances. For this, it suffices to show *either* that the joint moment generating function *or* the joint characteristic function (under Q) of the increments

$$\tilde{W}(t_1), \tilde{W}(t_2) - \tilde{W}(t_1), \cdots, \tilde{W}(t_n) - \tilde{W}(t_{n-1})$$

where $0 < t_1 < t_2 < \cdots < t_n$, is the same as that of n independent, normally distributed random variables with expectations 0 and variances $t_1, t_2 - t_1, \ldots$, that is, either

$$E_Q \exp\left\{\sum_{k=1}^n \alpha_k(\tilde{W}(t_k) - \tilde{W}(t_{k-1}))\right\} = \prod_{k=1}^n \exp\left\{\alpha_k^2(t_k - t_{k-1})\right\} \quad \text{or} \tag{71}$$

$$E_Q \exp\left\{\sum_{k=1}^n i\theta_k(\tilde{W}(t_k) - \tilde{W}(t_{k-1}))\right\} = \prod_{k=1}^n \exp\left\{-\theta_k^2(t_k - t_{k-1})\right\}. \tag{72}$$

Special Case: Assume that the integrand process V_s is bounded. In this case it is easiest to use moment generating functions. Consider for simplicity the case $n = 1$: To evaluate the expectation E_Q on the left side of (71), we rewrite it as an expectation under P, using the basic likelihood ratio identity relating the two expectation operators:

$$\begin{aligned}
E_Q \exp\left\{\alpha\tilde{W}(t)\right\} &= E_Q \exp\left\{\alpha W(t) - \alpha\int_0^t V_s \, ds\right\} \\
&= E_P \exp\left\{\alpha W(t) - \alpha\int_0^t V_s \, ds\right\} \exp\left\{\int_0^t V_s \, dW_s - \int_0^t V_s^2 \, ds/2\right\} \\
&= E_P \exp\left\{\int_0^t (\alpha + V_s) \, dW_s - \int_0^t (2\alpha V_s + V_s^2) \, ds/2\right\} \\
&= e^{\alpha^2 t/2} E_P \exp\left\{\int_0^t (\alpha + V_s) \, dW_s - \int_0^t (\alpha + V_s)^2 \, ds/2\right\} \\
&= e^{\alpha^2 t},
\end{aligned}$$

as desired. In the last step we used the fact that the exponential integrates to one. This follows from Novikov's theorem, because the hypothesis that the integrand V_s is bounded guarantees that Novikov's condition (67) is satisfied by $(\alpha + V_s)$. A similar calculation shows that (71) holds for $n > 1$.

General Case: Unfortunately, the final step in the calculation above cannot be justified in general. However, a similar argument can be made using characteristic functions rather than moment generating functions. Once again, consider for simplicity the case $n = 1$:

$$
\begin{aligned}
E_Q \exp\left\{i\theta \tilde{W}(t)\right\} &= E_Q \exp\left\{i\theta W(t) - i\theta \int_0^t V_s\,ds\right\} \\
&= E_P \exp\left\{i\theta W(t) - i\theta \int_0^t V_s\,ds\right\} \exp\left\{\int_0^t V_s\,dW_s - \int_0^t V_s^2\,ds/2\right\} \\
&= E_P \exp\left\{\int_0^t (i\theta + V_s)\,dW_s - \int_0^t (2i\theta V_s + V_s^2)\,ds/2\right\} \\
&= E_P \exp\left\{\int_0^t (i\theta + V_s)\,dW_s - \int_0^t (i\theta + V_s)^2\,ds/2\right\} e^{-\theta^2 t/2} \\
&= e^{-\theta^2 t/2},
\end{aligned}
$$

which is the characteristic function of the normal distribution $N(0, t)$. To justify the final equality, we must show that

$$E_P \exp\{X_t - [X]_t/2\} = 1$$

where

$$X_t = \int_0^t (i\theta + V_s)\,dW_s \quad \text{and} \quad [X]_t = \int_0^t (i\theta + V_s)^2\,ds$$

Itô's formula implies that

$$d \exp\{X_t - [X]_t/2\} = \exp\{X_t - [X]_t/2\}\,(i\theta + V_t)\,dW_t,$$

and so the martingale property will hold up to any stopping time τ that keeps the integrand on the right side bounded. Define stopping times

$$\tau(n) = \inf\{s \,:\, |X|_s = n \text{ or } [X]_s = n \text{ or } |V_s| = n\};$$

then for each $n = 1, 2, \dots,$

$$E_P \exp\{X_{t\wedge\tau(n)} - [X]_{t\wedge\tau(n)}/2\} = 1$$

As $n \to \infty$, the integrand converges pointwise to $\exp\{X_t - [X]_t/2\}$, so to conclude the proof it suffices to verify uniform integrability. For this, observe that for any $s \leq t$,

$$|\exp\{X_s - [X]_s/2\}| \leq e^{\theta s} Z_s \leq e^{\theta t} Z_s$$

By hypothesis, the process $\{Z_s\}_{s\leq t}$ is a positive martingale, and consequently the random variables $Z_{t\wedge\tau(n)}$ are uniformly integrable. This implies that the random variables $|\exp\{X_{t\wedge\tau(n)} - [X]_{t\wedge\tau(n)}/2\}|$ are also uniformly integrable. $\square$

6.4 Example: Ornstein-Uhlenbeck revisited

Recall that the solution X_t to the linear stochastic differential equation (41) with initial condition $X_0 = x$ is an Ornstein-Uhlenbeck process with mean-reversion parameter α. Because the stochastic differential equation (41) is of the same form as equation (70), Girsanov's theorem implies that a change of measure will convert a standard Brownian motion to an Ornstein-Uhlenbeck process with initial point 0. The details are as follows: Let W_t be a standard Brownian motion defined on a probability space $(\Omega, \mathcal{F}, P)$, and let $\mathbb{F} = \{\mathcal{F}_t\}_{t \geq 0}$ be the associated Brownian filtration. Define

$$N_t = -\alpha \int_0^t W_s \, dW_s, \tag{73}$$

$$Z_t = \exp\{N_t - [N]_t/2\}, \tag{74}$$

and let $Q = Q_T$ be the probability measure on $\mathcal{F}_T$ with likelihood ratio Z_T relative to P. Observe that the quadratic variation $[N]_t$ is just the integral $\int_0^t W_s^2 \, ds$. Theorem 8 asserts that, under the measure Q, the process

$$\tilde{W}_t := W_t + \int_0^t \alpha W_s \, ds,$$

for $0 \leq t \leq T$, is a standard Brownian motion. But this implies that the process W itself must solve the stochastic differential equation

$$dW_t = -\alpha W_t \, dt + d\tilde{W}_t$$

under Q. It follows that $\{W_t\}_{0 \leq t \leq T}$ is, under Q, an Ornstein-Uhlenbeck process with mean-reversion parameter α and initial point x. Similarly, the shifted process $x + W_t$ is, under Q, an Ornstein-Uhlenbeck process with initial point x.

It is worth noting that the likelihood ratio Z_t may be written in an alternative form that contains no Itô integrals. To do this, use the Itô formula on the quadratic function $u(x) = x^2$ to obtain $W_t^2 = 2I_t(W) + t$; this shows that the Itô integral in the definition (73) of N_t may be replaced by $W_t^2/2 - t/2$. Hence, the likelihood ratio Z_t may be written as

$$Z_t = \exp\{-\alpha W_t^2/2 + \alpha t/2 - \alpha^2 \int_0^t W_s^2 \, ds/2\}. \tag{75}$$

A physicist would interpret the quantity in the exponential as the *energy* of the path W in a quadratic potential well. According to the fundamental postulate of statistical physics (see FEYNMAN, *Lectures on Statistical Physics*), the probability of finding a system in a given configuration σ is proportional to the exponential $e^{-H(\sigma)/kT}$, where $H(\sigma)$ is the energy (Hamiltonian) of the system in configuration σ. Thus, a physicist might view the Ornstein-Uhlenbeck process as describing fluctuations in a quadratic potential. (In fact, the Ornstein-Uhlenbeck process was originally invented to describe the instantaneous *velocity* of a particle undergoing rapid collisions with molecules in a gas. See the book by Edward Nelson on Brownian motion for a discussion of the physics of the Brownian motion process.)

The formula (75) also suggests an interpretation of the change of measure in the language of acceptance/rejection sampling: Run a standard Brownian motion for time T to obtain a path $x(t)$; then "accept" this path with probability proportional to

$$\exp\{-\alpha x_T^2/2 - \alpha^2 \int_0^T x_s^2\, ds/2\}.$$

The random paths produced by this acceptance/rejection scheme will be distributed as Ornstein-Uhlenbeck paths with initial point 0. This suggests (and it can be proved, but this is beyond the scope of these notes) that the Ornstein-Uhlenbeck measure on $C[0,T]$ is the weak limit of a sequence of discrete measures μ_n that weight random walk paths according to their potential energies.

Problem 1. Show how the Brownian bridge can be obtained from Brownian motion by change of measure, and find an expression for the likelihood ratio that contains no Itô integrals.

6.5 Example: Brownian motion conditioned to stay positive

For each $x \in \mathbb{R}$, let P^x be a probability measure on $(\Omega, \mathcal{F})$ such that under P^x the process W_t is a one-dimensional Brownian motion with initial point $W_0 = x$. (Thus, under P^x the process $W_t - x$ is a standard Brownian motion.) For $a < b$ define

$$T = T_{a,b} = \min\{t \geq 0 \,:\, W_t \notin (a,b)\}.$$

Proposition 17. *Fix $0 < x < b$, and let $T = T_{0,b}$. Let Q^x be the probability measure obtained from P^x by conditioning on the event $\{W_T = b\}$ that W reaches b before 0, that is,*

$$Q^x(F) := P^x(F \cap \{W_T = b\})/P^x\{W_T = b\}. \tag{76}$$

Then under Q^x the process $\{W_t\}_{t \leq T}$ has the same distribution as does the solution of the stochastic differential equation

$$dX_t = X_t^{-1}\, dt + dW_t, \quad X_0 = x \tag{77}$$

under P^x. In other words, conditioning on the event $W_T = b$ has the same effect as adding the location-dependent drift $1/X_t$.

Proof. Q^x is a measure on the $\sigma-$algebra $\mathcal{F}_T$ that is absolutely continuous with respect to P^x. The likelihood ratio dQ^x/dP^x on $\mathcal{F}_T$ is

$$Z_T := \frac{\mathbf{1}\{W_T = b\}}{P^x\{W_T = b\}} = \frac{b}{x}\mathbf{1}\{W_T = b\}.$$

For any (nonrandom) time $t \geq 0$, the likelihood ratio $Z_{t\wedge T} = dQ^x/dP^x$ on $\mathcal{F}_{T\wedge t}$ is gotten by computing the conditional expectation of Z_T under P^x (Proposition 15). Since Z_T is a function only of the endpoint W_T, its conditional expectation on $\mathcal{F}_{T\wedge t}$ is the same as the conditional expectation on $\sigma(W_{T\wedge t})$, by the (strong) Markov property of Brownian motion. Moreover, this conditonal expectation is just $W_{T\wedge t}/b$ (gambler's ruin!). Thus,

$$Z_{T\wedge t} = W_{T\wedge t}/x.$$

This doesn't at first sight appear to be of the exponential form required by the Girsanov theorem, but actually it is: by the Itô formula,

$$Z_{T\wedge t} = \exp\{\log(W_{T\wedge t}/x)\} = \exp\left\{\int_0^{T\wedge t} W_s^{-1}\, dW_s - \int_0^{T\wedge t} W_s^{-2}\, ds/2\right\}.$$

Consequently, the Girsanov theorem (Theorem 8) implies that under Q^x the process $W_{t\wedge T} - \int_0^{T\wedge t} W_s^{-1}\, ds$ is a Brownian motion. This is equivalent to the assertion that under Q^x the process $W_{T\wedge t}$ behaves as a solution to the stochastic differential equation (77). □

Observe that the stochastic differential equation (77) does not involve the stopping place b. Thus, Brownian motion conditioned to hit b before 0 can be constructed by running a Brownian motion conditioned to hit $b+a$ before 0, and stopping it at the first hitting time of b. Alternatively, one can construct a Brownian motion conditioned to hit b before 0 by solving[1] the stochastic differential equation (77) for $t \geq 0$ and stopping it at the first hit of b. Now observe that if W_t is a Brownian motion started at any fixed s, the events $\{W_{T_{0,b}} = b\}$ are decreasing in b, and their intersection over all $b > 0$ is the event that $W_t > 0$ for all t and W_t visits all $b > x$. (This is, of course, an event of probability zero.) For this reason, probabilists refer to the process X_t solving the stochastic differential equation (77) as *Brownian motion conditioned to stay positive*.

CAUTION: Because the event that $W_t > 0$ for all $t > 0$ has probability zero, we cannot define "Brownian motion conditioned to stay positive" using conditional probabilities directly in the same way (see equation (76)) that we defined "Brownian motion conditioned to hit b before 0". Moreover, whereas Brownian motion conditioned to hit b before 0 has a distribution that is absolutely continuous relative to that of Brownian motion, Brownian motion conditioned to stay positive for all $t > 0$ has a distribution (on $C[0,\infty)$) that is necessarily *singular* relative to the law of Brownian motion.

7 Local Time and the Tanaka Formula

7.1 Occupation Measure of Brownian Motion

Let W_t be a standard one-dimensional Brownian motion and let $\mathbb{F} := \{\mathcal{F}_t\}_{t\geq 0}$ the associated Brownian filtration. A sample path of $\{W_s\}_{0\leq s\leq t}$ induces, in a natural way, an *occupation measure* Γ_t on (the Borel field of) $\mathbb{R}$:

$$\Gamma_t(A) := \int_0^t \mathbf{1}_A(W_s)\, ds. \tag{78}$$

Theorem 9. *With probability one, for each $t < \infty$ the occupation measure Γ_t is absolutely continuous with respect to Lebesgue measure, and its Radon-Nikodym derivative $L(t;x)$ is jointly continuous in t and x.*

[1]We haven't yet established that the SDE (77) has a soulution for all $t \geq 0$. This will be done later.

The occupation density $L(t;x)$ is known as the *local time* of the Brownian motion at x. It was first studied by Paul Lévy, who showed — among other things — that it could be defined by the formula

$$L(t;x) = \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_0^t \mathbf{1}_{[x-\varepsilon, x+\varepsilon]}(W_s)\, ds. \tag{79}$$

Joint continuity of $L(t;x)$ in t, x was first proved by Trotter. The modern argument that follows is based on an integral representation discovered by Tanaka.

7.2 Tanaka's Formula

Theorem 10. *The process $L(t;x)$ defined by*

$$2L(t;x) = |W_t - x| - |x| - \int_0^t \operatorname{sgn}(W_s - x)\, dW_s \tag{80}$$

is almost surely nondecreasing, jointly continuous in t, x, and constant on every open time interval during which $W_t \neq x$.

For the proof that $L(t;x)$ is continuous (more precisely, that it has a continuous version) we will need two auxiliary results. The first is the Burkholder-Davis-Gundy inequality (Corollary 5 above). The second, the Kolmogorov-Chentsov criterion for path-continuity of a stochastic process, we have seen earlier:

Proposition 18. *(Kolmogorov-Chentsov) Let $Y(t)$ be a stochastic process indexed by a $d-$dimensional parameter t. Then $Y(t)$ has a version with continuous paths if there exist constants $p, \delta > 0$ and $C < \infty$ such that for all s, t,*

$$E|Y(t) - Y(s)|^p \leq C|t - s|^{d+\delta}. \tag{81}$$

Proof of Theorem 10: Continuity. Since the process $|W_t - x| - |x|$ is jointly continuous, the Tanaka formula (80) implies that it is enough to show that

$$Y(t;x) := \int_0^t \operatorname{sgn}(W_s - x)\, dW_s$$

is jointly continuous in t and x. For this, we appeal to the Kolmogorov-Chentsov theorem: This asserts that to prove continuity of $Y(t;x)$ in t, x it suffices to show that for some $p \geq 1$ and $C, \delta > 0$,

$$E|Y(t;x) - Y(t;x')|^p \leq C|x - x'|^{2+\delta} \quad \text{and} \tag{82}$$

$$E|Y(t;x) - Y(t';x)|^p \leq C|t - t'|^{2+\delta}. \tag{83}$$

I will prove only (82), with $p = 6$ and $\delta = 1$; the other inequality, with the same values of p, δ, is similar. Begin by observing that for $x < x'$,

$$\begin{aligned} Y(t;x) - Y(t;x') &= \int_0^t \{\operatorname{sgn}(W_s - x) - \operatorname{sgn}(W_s - x')\}\, dW_s \\ &= 2\int_0^t \mathbf{1}_{(x,x')}(W_s)\, dW_s \end{aligned}$$

This is a martingale in t whose quadratic variation is flat when W_s is not in the interval (x, x') and grows linearly in time when $W_s \in (x, x')$. By Corollary 5,

$$
\begin{aligned}
&E|Y(t;x) - Y(t;x')|^{2m} \\
&\qquad \leq C_m 2^{2m} E\left(\int_0^t \mathbf{1}_{(x,x')}(W_s)\, ds\right)^m \\
&\qquad \leq C_m 2^{3m} |x - x'|^m m! \int_{t_1=0}^{t} \int_{t_2=0}^{t-t_1} \cdots \int_{t_m=0}^{t-t_{m-1}} \frac{1}{\sqrt{t_1 t_2 \cdots t_m}}\, dt_1 dt_2 \cdots dt_m \\
&\qquad \leq C' |x - x'|^m.
\end{aligned}
$$

□

Proof of Theorem 10: Conclusion. It remains to show that $L(t;x)$ is nondecreasing in t, and that it is constant on any time interval during which $W_t \neq x$. Observe that the Tanaka formula (80) is, in effect, a variation of the Itô formula for the absolute value function, since the sgn function is the derivative of the absolute value everywhere except at 0. This suggests that we try an approach by approximation, using the usual Itô formula to a smoothed version of the absolute value function. To get a smoothed version of $|\cdot|$, convolve with a smooth probability density with support contained in $[-\varepsilon, \varepsilon]$. Thus, let φ be an even, C^∞ probability density on $\mathbb{R}$ with support contained in $(-1, 1)$ (see the proof of Lemma **??** in the Appendix below), and define

$$
\begin{aligned}
\varphi_n(x) &:= n\varphi(nx); \\
\psi_n(x) &:= -1 + 2\int_{-\infty}^{x} \varphi_n(y)\, dy; \quad \text{and} \\
F_n(x) &:= |x| \quad \text{for } |x| > 1 \\
&:= 1 + \int_{-1/n}^{x} \psi_n(z)\, dz \quad \text{for } |x| \leq 1.
\end{aligned}
$$

Note that $\int_{-1/n}^{1/n} \psi_n = 0$, because of the symmetry of φ about 0. (EXERCISE: Check this.) Consequently, F_n is C^∞ on $\mathbb{R}$, and agrees with the absolute value function outside the interval $[-n^{-1}, n^{-1}]$. The first derivative of F_n is ψ_n, and hence is bounded in absolute value by 1; it follows that $F_n(y) \to |y|$ for all $y \in \mathbb{R}$. The second derivative $F_n'' = 2\varphi_n$. Therefore, Itô's formula implies that

$$
F_n(W_t - x) - F_n(-x) = \int_0^t \psi_n(W_s - x)\, dW_s + \frac{1}{2}\int_0^t 2\varphi_n(W_s - x)\, ds. \tag{84}
$$

By construction, $F_n(y) \to |y|$ as $n \to \infty$, so the left side of (84) converges to $|W_t - x| - |x|$. Now consider the stochastic integral on the right side of (84): As $n \to \infty$, the function ψ_n converges to sgn, and in fact ψ_n coincides with sgn outside of $[-n^{-1}, n^{-1}]$; moreover, the difference $|\operatorname{sgn} - \psi_n|$ is bounded. Consequently, as $n \to \infty$,

$$
\lim_{n \to \infty} \int_0^t \psi_n(W_s - x)\, dW_s = \int_0^t \operatorname{sgn}(W_s - x)\, dW_s,
$$

because the quadratic variation of the difference converges to 0. (EXERCISE: Check this.) This proves that two of the three quantities in (84) converge as $n \to \infty$, and so the third must also converge:

$$L(t;x) = \lim_{n\to\infty} \int_0^t \varphi_n(W_s - x)\, ds := L_n(t;x) \tag{85}$$

Each $L_n(t;x)$ is nondecreasing in t, because $\varphi_n \geq 0$; therefore, $L(t;x)$ is also nondecreasing in t. Finally, since $\varphi_n = 0$ except in the interval $[-n^{-1}, n^{-1}]$, each of the processes $L_{n+m}(t;x)$ is constant during time intervals when $W_t \notin [-n^{-1}, n^{-1}]$. Hence, $L(t;x)$ is constant on time intervals during which $W_s \neq x$. □

Proof of Theorem 9. It suffices to show that for every *continuous* function $f : \mathbb{R} \to \mathbb{R}$ with compact support,

$$\int_0^t f(W_s)\, ds = \int_{\mathbb{R}} f(x) L(t;x)\, dx \quad \text{a.s.} \tag{86}$$

Let φ be, as in the proof of Theorem 10, a C^∞ probability density on $\mathbb{R}$ with support $[-1, 1]$, and let $\varphi_n(x) = n\varphi(nx)$; thus, φ_n is a probability density with support $[-1/n, 1/n]$. By Corollary **??**,

$$\varphi_n * f \to f$$

uniformly as $n \to \infty$. Therefore,

$$\lim_{n\to\infty} \int_0^t \varphi_n * f(W_s)\, ds = \int_0^t f(W_s)\, ds.$$

But

$$\begin{aligned}
\int_0^t \varphi_n * f(W_s)\, ds &= \int_0^t \int_{\mathbb{R}} f(x)\varphi_n(W_s - x)\, dx ds \\
&= \int_{\mathbb{R}} f(x) \int_0^t \varphi_n(W_s - x)\, ds dx \\
&= \int_{\mathbb{R}} f(x) L_n(t;x)\, dx
\end{aligned}$$

where $L_n(t;x)$ is defined in equation (85) in the proof of Theorem 10. Since $L_n(t;x) \to L(t;x)$ as $n \to \infty$, equation 86 follows. □

7.3 Skorohod's Lemma

Tanaka's formula (80) relates three interesting processes: the local time process $L(t;x)$, the reflected Brownian motion $|W_t - x|$, and the stochastic integral

$$X(t;x) := \int_0^t \operatorname{sgn}(W_s - x)\, dW_s. \tag{87}$$

Proposition 19. *For each $x \in \mathbb{R}$ the process $\{X(t;x)\}_{t\geq 0}$ is a standard Wiener process.*

Proof. Since $X(t;x)$ is given as the stochastic integral of a bounded integrand, the time-change theorem implies that it is a time-changed Brownian motion. To show that the time change is trivial it is enough to check that the accumulated quadratic variation up to time t is t. But the quadratic variation is

$$\begin{aligned} [X]_t &= \int_0^t \operatorname{sgn}(W_s - x)^2 \, ds \leq t, \quad \text{and} \\ E[X]_t &= \int_0^t P\{W_s \neq x\} \, ds = t. \end{aligned}$$

□

Thus, the process $|x| + X(t;x)$ is a Brownian motion started at $|x|$. Tanaka's formula, after rearrangement of terms, shows that this Brownian motion can be decomposed as the difference of the nonnegative process $|W_t - x|$ and the nondecreasing process $2L(t;x)$:

$$|x| + X(t;x) = |W_t - x| - 2L(t;x). \tag{88}$$

This is called *Skorohod's equation*. Skorohod discovered that there is only one such decomposition of a continuous path, and that the terms of the decomposition have a peculiar form:

Lemma 7. *(Skorohod) Let $w(t)$ be a continuous, real-valued function of $t \geq 0$ such that $w(0) \geq 0$. Then there exists a unique pair of real-valued continuous functions $x(t)$ and $y(t)$ such that*

(a) $x(t) = w(t) + y(t)$;

(b) $x(t) \geq 0$ *and* $y(0) = 0$;

(c) $y(t)$ *is nondecreasing and is constant on any time interval during which* $x(t) > 0$.

The functions $x(t)$ and $y(t)$ are given by

$$y(t) = (-\min_{s \leq t} w(s)) \vee 0 \quad \textit{and} \quad x(t) = w(t) + y(t). \tag{89}$$

Proof. First, let's verify that the functions $x(t)$ and $y(t)$ defined by (89) satisfy properties (a)–(c). It is easy to see that if $w(t)$ is continuous then its minimum to time t is also continuous; hence, both $x(t)$ and $y(t)$ are continuous. Up to the first time $t = t_0$ that $w(t) = 0$, the function $y(s)$ will remain at the value 0, and so $x(s) = w(s) \geq 0$ for all $s \leq t_0$. For $t \geq t_0$,

$$y(t) = -\min_{s \leq t} w(s) \quad \text{and} \quad x(t) = w(t) - \min_{s \leq t} w(s);$$

hence, $x(t) \geq 0$ for all $t \geq t_0$. Clearly, $y(t)$ never decreases; and after time t_0 it increases only when $w(t)$ is at its minimum, so $x(t) = 0$. Thus, (a)–(c) hold for the functions $x(t)$ and $y(t)$.

Now suppose that (a)–(c) hold for some other functions $\tilde{x}(t)$ and $\tilde{y}(t)$. By hypothesis, $y(0) = \tilde{y}(0) = 0$, so $x(0) = \tilde{x}(0) = w(0) \geq 0$. Suppose that at some time $t_* \geq 0$,

$$\tilde{x}(t_*) - x(t_*) = \varepsilon > 0.$$

Then up until the next time $s_* > t_*$ that $\tilde{x} = 0$, the function $\tilde{y}$ must remain constant, and so $\tilde{x} - w$ must remain constant. But $x - w = y$ never decreases, so up until time s_* the difference $\tilde{x} - x$ cannot increase; at time s_* (if this is finite) the difference $\tilde{x} - x$ must be ≤ 0, because $\tilde{x}(s_*) = 0$. Since $\tilde{x}(t) - x(t)$ is a continuous function of t that begins at $\tilde{x}(0) - x(0) = 0$, it follows that the difference $\tilde{x}(t) - x(t)$ can never exceed ε. But $\varepsilon > 0$ is arbitrary, so it must be that

$$\tilde{x}(t) - x(t) \leq 0 \;\; \forall\, t \geq 0.$$

The same argument applies when the roles of $\tilde{x}$ and x are reversed. Therefore, $x \equiv \tilde{x}$. □

Corollary 6. *(Lévy) For $x \geq 0$, the law of the vector-valued process $(|W_t - x|, L(t;x))$ coincides with that of $(W_t + x - M(t;x)^-, M(t;x)^-)$, where*

$$M(t;x)^- := \min_{s \leq t}(W_s + x) \wedge 0. \tag{90}$$

7.4 Application: Extensions of Itô's Formula

Theorem 11. *(Extended Itô Formula) Let $u : \mathbb{R} \to \mathbb{R}$ be twice differentiable (but not necessarily C^2). Assume that $|u''|$ is integrable on any compact interval. Let W_t be a standard Brownian motion. Then*

$$u(W_t) = u(0) + \int_0^t u'(W_s)\, dW_s + \frac{1}{2}\int_0^t u''(W_s)\, ds. \tag{91}$$

Proof. **Exercise.** Hint: First show that it suffices to consider the case where u has compact support. Next, let φ_δ be a C^∞ probability density with support $[-\delta, \delta]$. By Lemma **??**, $\varphi_\delta * u$ is C^∞, and so Theorem 1 implies that the Itô formula (91) is valid when u is replaced by $u * \varphi_\delta$. Now use Theorem 9 to show that as $\delta \to 0$,

$$\int_0^t u * \varphi_\delta''(W_s)\, ds \longrightarrow \int_0^t u''(W_s)\, ds.$$

□

Tanaka's theorem shows that there is at least a reasonable substitute for the Itô formula when $u(x) = |x|$. This function fails to have a second derivative at $x = 0$, but it does have the property that its derivative is nondecreasing. This suggest that the Itô formula might generalize to convex functions u, as these also have nondecreasing (left) derivatives.

Definition 5. A function $u : \mathbb{R} \to \mathbb{R}$ is *convex* if for any real numbers $x < y$ and any $s \in (0,1)$,

$$u(sx + (1-s)y) \leq su(x) + (1-s)u(y). \tag{92}$$

Lemma 8. *If u is convex, then it has a right derivative $D^+(x)$ at all but at most countably many points $x \in \mathbb{R}$, defined by*

$$D^+u(x) := \lim_{\varepsilon \to 0+} \frac{u(x+\varepsilon) - u(x)}{\varepsilon}. \tag{93}$$

The function $D^+u(x)$ is nondecreasing and right continuous in x, and the limit (93) *exists everywhere except at those x where $D^-u(x)$ has a jump discontinuity.*

Proof. Exercise — or check any reasonable real analysis textbook, e.g. ROYDEN, ch. 5. □

Since D^+u is nondecreasing and right continuous, it is the cumulative distribution function of a Radon measure[2] μ on $\mathbb{R}$, that is,

$$\begin{aligned} D^+u(x) &= D^+u(0) + \mu((0,x]) \quad \text{for } x > 0 \\ &= D^+u(0) - \mu((x,0]) \quad \text{for } x \leq 0. \end{aligned} \tag{94}$$

Consequently, by the Lebesgue differentiation theorem and an integration by parts (Fubini),

$$\begin{aligned} u(x) &= u(0) + D^+u(0)x + \int_{[0,x]} (x-y)\,d\mu(y) \quad \text{for } x > 0 \\ &= u(0) + D^+u(0)x + \int_{[x,0]} (y-x)\,d\mu(y) \quad \text{for } x \leq 0. \end{aligned} \tag{95}$$

Observe that this exhibits u as a mixture of absolute value functions $a_y(x) := |x-y|$. Thus, given the Tanaka formula, the next result should come as no surprise.

Theorem 12. *(Generalized Itô Formula) Let $u : \mathbb{R} \to \mathbb{R}$ be convex, with right derivative D^+u and second derivative measure μ, as above. Let W_t be a standard Brownian motion and let $L(t;x)$ be the associated local time process. Then*

$$u(W_t) = u(0) + \int_0^t D^+u(W_s)\,dW_s + \int_{\mathbb{R}} L(t;x)\,d\mu(x). \tag{96}$$

Proof. Another exercise. (Smooth; use Itô; take limits; use Tanaka and Trotter.) □

[2] A *Radon measure* is a Borel measure that attaches finite mass to any compact interval.