CHAPTER 3

Stochastic Integration

1. Motivation

Suppose $\Delta(t)$ is your position at time t on a security whose price is $S(t)$. If you only trade this security at times $0 = t_0 < t_1 < t_2 < \cdots < t_n = T$, then the change in the value of your wealth up to time T is given by

$$X(t_n) - X(0) = \sum_{i=0}^{n-1} \Delta(t_i)(S(t_{i+1}) - S(t_i))$$

If you are trading this continuously in time, you'd expect that a "simple" limiting procedure should show that your wealth is given by the *Riemann-Stieltjes* integral:

$$X(T) - X(0) = \lim_{\|P\|\to 0} \sum_{i=0}^{n-1} \Delta(t_i)(S(t_{i+1}) - S(t_i)) = \int_0^T \Delta(t)\, dS(t)\,.$$

Here $P = \{0 = t_0 < \cdots < t_n = T\}$ is a partition of $[0,T]$, and $\|P\| = \max\{t_{i+1} - t_i\}$.

This has been well studied by mathematicians, and it is well known that for the above limiting procedure to "work directly", you need S to have finite *first variation*. Recall, the *first variation* of a function is defined to be

$$V_{[0,T]}(S) \stackrel{\text{def}}{=} \lim_{\|P\|\to 0} \sum_{i=0}^{n-1} |S(t_{i+1}) - S(t_i)|\,.$$

It turns out that almost *any* continuous martingale S *will not* have finite first variation. Thus to define integrals with respect to martingales, one has to do something 'clever'. It turns out that if X is adapted and S is an martingale, then the above limiting procedure works, and this was carried out by Itô (and independently by Doeblin).

2. The First Variation of Brownian motion

We begin by showing that the first variation of Brownian motion is infinite.

PROPOSITION 2.1. *If W is a standard Brownian motion, and $T > 0$ then*

$$\lim_{n\to\infty} \boldsymbol{E} \sum_{k=0}^{n-1} \Big| W\Big(\frac{k+1}{n}\Big) - W\Big(\frac{k}{n}\Big) \Big| = \infty\,.$$

REMARK 2.2. In fact

$$\lim_{n\to\infty} \sum_{k=0}^{n-1} \Big| W\Big(\frac{k+1}{n}\Big) - W\Big(\frac{k}{n}\Big) \Big| = \infty \quad \text{almost surely,}$$

but this won't be necessary for our purposes.

PROOF. Since $W((k+1)/n) - W(k/n) \sim N(0, 1/n)$ we know

$$\boldsymbol{E}\Big| W\Big(\frac{k+1}{n}\Big) - W\Big(\frac{k}{n}\Big) \Big| = \int_{\mathbb{R}} |x|\, G\Big(\frac{1}{n}, x\Big)\, dx = \frac{C}{\sqrt{n}}\,,$$

where

$$C = \int_{\mathbb{R}} |y| e^{-y^2/2} \frac{dy}{\sqrt{2\pi}} = \boldsymbol{E}|N(0,1)|\,.$$

Consequently

$$\sum_{k=0}^{n-1} \boldsymbol{E}\Big| W\Big(\frac{k+1}{n}\Big) - W\Big(\frac{k}{n}\Big) \Big| = \frac{Cn}{\sqrt{n}} \xrightarrow{n\to\infty} \infty\,. \qquad \square$$

3. Quadratic Variation

It turns out that the *second* variation of any *square integrable* martingale is almost surely finite, and this is the key step in constructing the Itô integral.

DEFINITION 3.1. Let M be any process. We define the *quadratic variation* of M, denoted by $[M, M]$ by

$$[M, M](T) = \lim_{\|P\|\to 0} \sum_{i=0}^{n-1} (\Delta_i M)^2\,,$$

where $P = \{0 = t_1 < t_1 \cdots < t_n = T\}$ is a partition of $[0,T]$, and $\Delta_i M = M(t_{i+1}) - M(t_i)$.

PROPOSITION 3.2. *If W is a standard Brownian motion, then $[W, W](T) = T$ almost surely.*

PROOF. For simplicity, let's assume $t_i = Ti/n$. Note

$$\sum_{i=0}^{n-1} (\Delta_i W)^2 - T = \sum_{i=0}^{n-1} \Big(W\Big(\frac{(i+1)T}{n}\Big) - W\Big(\frac{iT}{n}\Big) \Big)^2 - T = \sum_{i=0}^{n-1} \xi_i\,,$$

where

$$\xi_i \stackrel{\text{def}}{=} (\Delta_i W)^2 - \frac{T}{n} = \Big(W\Big(\frac{(i+1)T}{n}\Big) - W\Big(\frac{iT}{n}\Big) \Big)^2 - \frac{T}{n}\,.$$

Note that ξ_i's are i.i.d. with distribution $N(0, T/n)^2 - T/n$, and hence

$$\boldsymbol{E}\xi_i = 0 \qquad \text{and} \qquad \operatorname{Var} \xi_i = \frac{T^2(\boldsymbol{E}N(0,1)^4 - 1)}{n^2}.$$

Consequently

$$\operatorname{Var}\Big(\sum_{i=0}^{n-1} \xi_i \Big) = \frac{T^2(\boldsymbol{E}N(0,1)^4 - 1)}{n} \xrightarrow{n\to\infty} 0\,,$$

which shows

$$\sum_{i=0}^{n-1} \Big(W\Big(\frac{(i+1)T}{n}\Big) - W\Big(\frac{iT}{n}\Big) \Big)^2 - T = \sum_{i=0}^{n-1} \xi_i \xrightarrow{n\to\infty} 0\,. \qquad \square$$

COROLLARY 3.3. *The process $M(t) \stackrel{\text{def}}{=} W(t)^2 - [W, W](t)$ is a martingale.*

Proof. We see

$$\boldsymbol{E}(W(t)^2 - t \mid \mathcal{F}_s) = \boldsymbol{E}((W(t)-W(s))^2 + 2W(s)(W(t)-W(s)) + W(s)^2 \mid \mathcal{F}_s) - t$$
$$= W(s)^2 - s$$

and hence $\boldsymbol{E}(M(t) \mid \mathcal{F}_s) = M(s)$. □

The above wasn't a co-incidence. This property in fact characterizes the quadratic variation.

Theorem 3.4. *Let M be a continuous martingale with respect to a filtration $\{\mathcal{F}_t\}$. Then $\boldsymbol{E}M(t)^2 < \infty$ if and only if $\boldsymbol{E}[M,M](t) < \infty$. In this case the process $M(t)^2 - [M,M](t)$ is also a martingale with respect to the same filtration, and hence $\boldsymbol{E}M(t)^2 - \boldsymbol{E}M(0)^2 = \boldsymbol{E}[M,M](t)$.*

The above is in fact a characterization of the quadratic variation of martingales.

Theorem 3.5. *If $A(t)$ is any continuous, increasing, adapted process such that $A(0) = 0$ and $M(t)^2 - A(t)$ is a martingale, then $A = [M,M]$.*

The proof of these theorems are a bit technical and go beyond the scope of these notes. The results themselves, however, are extremely important and will be used subsequently.

Remark 3.6. The intuition to keep in mind about the first variation and the quadratic variation is the following. Divide the interval $[0,T]$ into $T/\delta t$ intervals of size δt. If X has finite first variation, then on each subinterval $(k\delta t, (k+1)\delta t)$ the increment of X should be of order δt. Thus adding $T/\delta t$ terms of order δt will yield something finite.

On the other hand if X has finite quadratic variation, on each subinterval $(k\delta t, (k+1)\delta t)$ the increment of X should be of order $\sqrt{\delta t}$, so that adding $T/\delta t$ terms of the *square* of the increment yields something finite. Doing a quick check for Brownian motion (which has finite quadratic variation), we see

$$\boldsymbol{E}|W(t+\delta t) - W(t)| = \sqrt{\delta t}\,\boldsymbol{E}|N(0,1)|\,,$$

which is in line with our intuition.

Remark 3.7. If a continuous process has finite first variation, its quadratic variation will necessarily be 0. On the other hand, if a continuous process has finite (and non-zero) quadratic variation, its first variation will necessary be infinite.

4. Construction of the Itô integral

Let W be a standard Brownian motion, $\{\mathcal{F}_t\}$ be the Brownian filtration and D be an adapted process. We think of $D(t)$ to represent our position at time t on an asset whose spot price is $W(t)$.

Lemma 4.1. *Let $P = \{0 = t_0 < t_1 < t_2 < \cdots\}$ be an increasing sequence of times, and assume D is constant on $[t_i, t_{i+1})$ (i.e. the asset is only traded at times $t_0, \ldots, t_n$). Let $I_P(T)$, defined by*

$$I_P(T) = \sum_{i=0}^{n-1} D(t_i)\Delta_i W + D(t_n)(W(T) - W(t_n)) \quad \text{if } T \in [t_n, t_{n+1})\,.$$

be your cumulative winnings up to time T. As before $\Delta_i W \stackrel{\text{def}}{=} W(t_{i+1}) - W(t_i)$. Then,

$$\boldsymbol{E}I_P(T)^2 = \boldsymbol{E}\Big[\sum_{i=0}^{n} D(t_i)^2(t_{i+1}-t_i) + D(t_n)^2(T-t_n)\Big] \quad \text{if } T \in [t_n, t_{n+1})\,. \tag{4.1}$$

Moreover, I_P is a martingale and

$$[I_P, I_P](T) = \sum_{i=0}^{n-1} D(t_i)^2(t_{i+1}-t_i) + D(t_n)^2(T-t_n) \quad \text{if } T \in [t_n, t_{n+1})\,. \tag{4.2}$$

This lemma, as we will shortly see, is the key to the construction of stochastic integrals (called Itô integrals).

Proof. We first prove (4.1) with $T = t_n$ for simplicity. Note

$$\boldsymbol{E}I_P(t_n)^2 = \sum_{i=0}^{n-1} \boldsymbol{E}D(t_i)^2(\Delta_i W)^2 + 2\sum_{j=0}^{n-1}\sum_{i=0}^{j-1} \boldsymbol{E}D(t_i)D(t_j)(\Delta_i W)(\Delta_j W) \tag{4.3}$$

By the tower property

$$\boldsymbol{E}D(t_i)^2(\Delta_i W)^2 = \boldsymbol{E}\boldsymbol{E}\big(D(t_i)^2(\Delta_i W)^2 \bigm| \mathcal{F}_{t_i}\big)$$
$$= \boldsymbol{E}D(t_i)^2\boldsymbol{E}\big((W(t_{i+1}) - W(t_i))^2 \bigm| \mathcal{F}_{t_i}\big) = \boldsymbol{E}D(t_i)^2(t_{i+1}-t_i)\,.$$

Similarly we compute

$$\boldsymbol{E}D(t_i)D(t_j)(\Delta_i W)(\Delta_j W) = \boldsymbol{E}\boldsymbol{E}\big(D(t_i)D(t_j)(\Delta_i W)(\Delta_j W) \bigm| \mathcal{F}_{t_j}\big)$$
$$= \boldsymbol{E}D(t_i)D(t_j)(\Delta_i W)\boldsymbol{E}\big((W(t_{j+1}) - W(t_j)) \bigm| \mathcal{F}_{t_j}\big) = 0\,.$$

Substituting these in (4.3) immediately yields (4.1) for $t_n = T$.

The proof that I_P is an martingale uses the same "tower property" idea, and is left to the reader to check. The proof of (4.2) is also similar in spirit, but has a few more details to check. The main idea is to let $A(t)$ be the right hand side of (4.2). Observe A is clearly a continuous, increasing, adapted process. Thus, if we show $M^2 - A$ is a martingale, then using Theorem 3.5 we will have $A = [M,M]$ as desired. The proof that $M^2 - A$ is an martingale uses the same "tower property" idea, but is a little more technical and is left to the reader. □

Note that as $\|P\| \to 0$, the right hand side of (4.2) converges to the standard Riemann integral $\int_0^T D(t)^2\,dt$. Itô realised he could use this to prove that I_P itself converges, and the limit is now called the Itô integral.

Theorem 4.2. *If $\int_0^T D(t)^2\,dt < \infty$ almost surely, then as $\|P\| \to 0$, the processes I_P converge to a* continuous process *I denoted by*

$$I(T) \stackrel{\text{def}}{=} \lim_{\|P\|\to 0} I_P(T) \stackrel{\text{def}}{=} \int_0^T D(t)\,dW(t)\,. \tag{4.4}$$

This is known as the Itô integral of D with respect to W. *If further*

$$\boldsymbol{E}\int_0^T D(t)^2\,dt < \infty\,, \tag{4.5}$$

then the process $I(T)$ is a martingale *and the quadratic variation $[I,I]$ satisfies*

$$[I,I](T) = \int_0^T D(t)^2\,dt \quad \textit{almost surely.}$$

REMARK 4.3. For the above to work, it is *crucial* that D is adapted, and is sampled at the left endpoint of the time intervals. That is, the terms in the sum are $D(t_i)(W(t_{i+1}) - W(t_i))$, and not $D(t_{i+1})(W(t_{i+1}) - W(t_i))$ or $\frac{1}{2}(D(t_i) + D(t_{i+1}))(W(t_{i+1}) - W(t_i))$, or something else.

Usually if the process is not adapted, there is no meaningful way to make sense of the limit. However, if you sample at different points, it still works out (usually) but what you get is *different* from the Itô integral (one example is the Stratonovich integral).

REMARK 4.4. The variable t used in (4.4) is a "dummy" integration variable. Namely one can write

$$\int_0^T D(t)\,dW(t) = \int_0^T D(s)\,dW(s) = \int_0^T D(r)\,dW(r)\,,$$

or any other variable of your choice.

COROLLARY 4.5 (Itô Isometry). *If* (4.5) *holds then*

$$\boldsymbol{E}\Big(\int_0^T D(t)\,dW(t)\Big)^2 = \boldsymbol{E}\int_0^T D(t)^2\,dt\,.$$

PROPOSITION 4.6 (Linearity). *If D_1 and D_2 are two adapted processes, and $\alpha \in \mathbb{R}$, then*

$$\int_0^T (D_1(t) + \alpha D_2(t))\,dW(t) = \int_0^T D_1(t)\,dW(t) + \alpha\int_0^T D_2(t)\,dW(t)\,.$$

REMARK 4.7. Positivity, however, is not preserved by Itô integrals. Namely if $D_1 \leqslant D_2$, there is no reason to expect $\int_0^T D_1(t)\,dW(t) \leqslant \int_0^T D_2(t)\,dW(t)$. Indeed choosing $D_1 = 0$ and $D_2 = 1$ we see that we can not possibly have $0 = \int_0^T D_1(t)\,dW(t)$ to be almost surely smaller than $W(T) = \int_0^T D_2(t)\,dW(t)$.

Recall, our starting point in these notes was modelling stock prices as *geometric Brownian motions*, given by the equation

$$dS(t) = \alpha S(t)\,dt + \sigma S(t)\,dW(t)\,.$$

After constructing Itô integrals, we are now in a position to describe what this means. The above is simply shorthand for saying S is a process that satisfies

$$S(T) - S(0) = \int_0^T \alpha S(t)\,dt + \int_0^T \sigma S(t)\,dW(t)\,.$$

The first integral on the right is a standard Riemann integral. The second integral, representing the noisy fluctuations, is the Itô integral we just constructed.

Note that the above is a little more complicated than the Itô integrals we will study first, since the process S (that we're trying to define) also appears as an integrand on the right hand side. In general, such equations are called *Stochastic differential equations*, and are extremely useful in many contexts.

5. The Itô formula

Using the abstract "limit" definition of the Itô integral, it is hard to compute examples. For instance, what is

$$\int_0^T W(s)\,dW(s)\,?$$

This, as we will shortly, can be computed easily using the Itô formula (also called the Itô-Doeblin formula).

Suppose b and σ are adapted processes. (In particular, they could but need not, be random). Consider a process X defined by

$$X(T) = X(0) + \int_0^T b(t)\,dt + \int_0^T \sigma(t)\,dW(t)\,. \tag{5.1}$$

Note the first integral $\int_0^T b(t)\,dt$ is a regular Riemann integral that can be done directly. The second integral the Itô integral we constructed in the previous section.

DEFINITION 5.1. The process X is called an Itô process if $X(0)$ is deterministic (not random) and for all $T \geqslant 0$,

$$\boldsymbol{E}\int_0^T \sigma(t)^2\,dt < \infty \qquad\text{and}\qquad \int_0^T b(t)\,dt < \infty\,. \tag{5.2}$$

REMARK 5.2. The shorthand notation for (5.1) is to write

$$dX(t) = b(t)\,dt + \sigma(t)\,dW(t)\,. \tag{5.1$'$}$$

PROPOSITION 5.3. *The quadratic variation of X is*

$$[X,X](T) = \int_0^T \sigma(t)^2\,dt \quad \textit{almost surely.} \tag{5.3}$$

PROOF. Define B and M by

$$B(T) = \int_0^T b(t)\,dt \qquad\text{and}\qquad M(T) = \int_0^T \sigma(t)\,dW(t)\,,$$

and let $P = \{0 = t_0 < t_1 < \cdots < t_n = T\}$ be a partition of $[0,T]$, and $\|P\| = \max_i t_{i+1} - t_i$. Observe

$$\sum_{i=0}^{n-1}(\Delta_i X)^2 = \sum_{i=0}^{n-1}(\Delta_i M)^2 + \sum_{i=0}^{n-1}(\Delta_i B)^2 + 2\sum_{i=0}^{n-1}(\Delta_i B)(\Delta_i M)\,.$$

The first sum on the right converges (as $\|P\| \to 0$) to $[M,M](T)$, which we know is exactly $\int_0^T \sigma(t)^2\,dt$. For the second sum, observe

$$(\Delta_i B)^2 = \Big(\int_{t_i}^{t_{i+1}} b(s)\,ds\Big)^2 \leqslant \big(\max|b|^2\big)(t_{i+1} - t_i)^2 \leqslant \big(\max|b|^2\big)\|P\|(t_{i+1} - t_i)\,.$$

Hence

$$\Big|\sum_{i=0}^{n-1}(\Delta_i B)^2\Big| \leqslant \|P\|\big(\max|b|^2\big)T \xrightarrow{\|P\|\to 0} 0\,.$$

For the third term, one uses the *Cauchy-Schwartz* inequality to observe

$$\Big|\sum_{i=0}^{n-1}(\Delta_i B)(\Delta_i M)\Big| \leqslant \Big(\sum_{i=0}^{n-1}(\Delta_i B)^2\Big)^{1/2}\Big(\sum_{i=0}^{n-1}(\Delta_i M)^2\Big)^{1/2} \xrightarrow{\|P\|\to 0} 0\cdot[M,M](T)=0\,.$$

□

REMARK 5.4. It's common to decompose $X = B + M$ where M is a martingale and B has finite first variation. Processes that can be decomposed in this form are called *semi-martingales*, and the decomposition is unique. The process M is called the martingale part of X, and B is called the *bounded variation* part of X.

PROPOSITION 5.5. *The semi-martingale decomposition of X is unique. That is, if $X = B_1 + M_1 = B_2 + M_2$ where B_1, B_2 are continuous adapted processes with finite first variation, and M_1, M_2 are continuous (square integrable) martingales, then $B_1 = B_2$ and $M_1 = M_2$.*

PROOF. Set $M = M_1 - M_2$ and $B = B_1 - B_2$, and note that $M = -B$. Consequently, M has finite first variation and hence 0 quadratic variation. This implies $\boldsymbol{E}M(t)^2 = \boldsymbol{E}[M,M](t) = 0$ and hence $M = 0$ identically, which in turn implies $B = 0$, $B_1 = B_2$ and $M_1 = M_2$. □

As an immediate consequence, we see that the sum of an Itô integral and Riemann integral is a martingale, if and only if the Riemann integral is identically 0.

PROPOSITION 5.6. *Suppose b, σ are two adapted processes satisfying (5.2), and let X be as in equation (5.1). Then X is a martingale if and only if b is identically 0.*

PROOF 1. If b is identically 0, then we already know that X is a martingale. To prove the converse, suppose X is a martingale. Then the process $B(T) = \int_0^T b(t)\,dt = X(T) - X(0) - \int_0^T \sigma(s)\,ds$ is the difference of two martingales, and so must itself be a martingale. Now the Itô decomposition of the process B can be expressed in two different ways: First $B = B + 0$, where B has bounded variation, and 0 is a martingale. Second $B = 0 + B$, where 0 has bounded variation, and B is a martingale. By Proposition 5.5, the bounded variation parts and martingale parts must be equal, showing $B = 0$ identically. □

PROOF 2. Here's an alternate, direct proof of Proposition 5.6 without relying on Proposition 5.5. Suppose X is a martingale. Then as above, the process $B(T) = \int_0^T b(t)\,dt$ must also be a martingale, and so we must have $\boldsymbol{E}(B(T+h) \,|\, \mathcal{F}_T) = B(T)$ for every $h > 0$. Thus

$$\begin{aligned}\int_0^T b(t)\,dt = B(T) = \boldsymbol{E}(B(T+h)\,|\,\mathcal{F}_T) &= \boldsymbol{E}\Big(\int_0^{T+h} b(t)\,dt \,\Big|\, F_T\Big)\\ &= \int_0^T b(t)\,dt + \boldsymbol{E}\Big(\int_T^{T+h} b(t)\,dt \,\Big|\, F_T\Big)\,.\end{aligned}$$

This implies

$$\boldsymbol{E}\Big(\int_T^{T+h} b(t)\,dt \,\Big|\, F_T\Big) = 0\,,$$

for every $h > 0$. Dividing both sides by h, and taking the limit as $h \to 0$ shows

$$\begin{aligned}0 = \lim_{h\to 0}\boldsymbol{E}\Big(\frac{1}{h}\int_T^{T+h} b(t)\,dt \,\Big|\, F_T\Big) &= \boldsymbol{E}\Big(\lim_{h\to 0}\frac{1}{h}\int_T^{T+h} b(t)\,dt \,\Big|\, F_T\Big)\\ &= \boldsymbol{E}\Big(b(T) \,\Big|\, F_T\Big) = b(T)\,.\end{aligned}$$

Thus $b(T) = 0$ for every $T \geqslant 0$. This forces the process B to be identically 0, concluding the proof. □

Given an adapted process D, interpret X as the price of an asset, and D as our position on it. (We could either be long, or short on the asset so D could be positive or negative.)

DEFINITION 5.7. We define the integral of D with respect to X by

$$\int_0^T D(t)\,dX(t) \stackrel{\text{def}}{=} \int_0^T D(t)b(t)\,dt + \int_0^T D(t)\sigma(t)\,dW(t)\,.$$

As before, $\int_0^T D\,dX$ represents the winnings or profit obtained using the trading strategy D.

REMARK 5.8. Note that the first integral on the right $\int_0^T D(t)b(t)\,dt$ is a regular Riemann integral, and the second one is an Itô integral. Recall that Itô integrals with respect to Brownian motion (i.e. integrals of the form $\int_0^t D(s)\,dW(s)$ are martingales). Integrals with respect to a general process X are only guaranteed to be martingales if X itself is a martingale (i.e. $b = 0$), or if the integrand is 0.

REMARK 5.9. If we define I_P by

$$I_P(T) = \sum_{i=0}^{n-1} D(t_i)(\Delta_i X) + D(t_n)(X(T) - X(t_n)) \quad \text{if } T \in [t_n, t_{n+1})\,,$$

then I_P converges to the integral $\int_0^T D(t)\,dX(t)$ defined above. This works in the same way as Theorem 4.2.

Suppose now $f(t,x)$ is some function. If X is differentiable as a function of t *(which it most certainly is not)*, then the chain rule gives

$$\begin{aligned}f(T, X(T)) - f(0, X(0)) &= \int_0^T \partial_t\Big(f(t, X(t))\Big)\,dt\\ &= \int_0^T \partial_t f(t, X(t))\,dt + \int_0^T \partial_x f(t, X(t))\,\partial_t X(t)\,dt\\ &= \int_0^T \partial_t f(t, X(t))\,dt + \int_0^T \partial_x f(t, X(t))\,dX(t)\,.\end{aligned}$$

Itô process are *almost never* differentiable as a function of time, and so the above has no chance of working. It turns out, however, that for Itô process you can make the above work by adding an *Itô correction* term. This is the celebrated Itô formula (more correctly the Itô-Doeblin[1] formula).

[1]W. Doeblin was a French-German mathematician who was drafted for military service during the second world war. During the war he wrote down his mathematical work and sent it in a

THEOREM 5.10 (Itô formula, aka Itô-Doeblin formula). *If $f = f(t,x)$ is $C^{1,2}$ function*[2] *then*

$$f(T,X(T)) - f(0,X(0)) = \int_0^T \partial_t f(t,X(t))\,dt + \int_0^T \partial_x f(t,X(t))\,dX(t) + \frac{1}{2}\int_0^T \partial_x^2 f(t,X(t)\,d[X,X](t)\,. \tag{5.4}$$

REMARK 5.11. To clarify notation, $\partial_t f(t,X(t))$ means the following: differentiate $f(t,x)$ with respect to t (treating x as a constant), and then substitute $x = X(t)$. Similarly $\partial_x f(t,X(t))$ means differentiate $f(t,x)$ with respect to x, and then substitute $x = X(t)$. Finally $\partial_x^2 f(t,X(t))$ means take the second derivative of the function $f(t,x)$ with respect to x, and the substitute $x = X(t)$.

REMARK 5.12. In short hand differential form, this is written as

$$df(t,X(t)) = \partial_t f(t,X(t))\,dt + \partial_x f(t,X(t))\,dX(t) + \frac{1}{2}\partial_x^2 f(t,X(t))\,d[X,X](t)\,. \tag{5.4$'$}$$

The term $\frac{1}{2}\partial_x^2 f\,d[X,X](t)$ is an "extra" term, and is often referred to as the *Itô correction term.* The Itô formula is simply a version of the *chain rule* for stochastic processes.

REMARK 5.13. Substituting what we know about X from (5.1) and (5.3) we see that (5.4) becomes

$$f(T,X(T)) - f(0,X(0)) = \int_0^T \bigl(\partial_t f(t,X(t)) + \partial_x f(t,X(t))b(t)\bigr)\,dt + \int_0^T \partial_x f(t,X(t))\sigma(t)\,dW(t) + \frac{1}{2}\int_0^T \partial_x^2 f(t,X(t))\,\sigma(t)^2\,dt\,.$$

The second integral on the right is an Itô integral (and hence a martingale). The other integrals are regular Riemann integrals which yield processes of finite variation.

While a complete rigorous proof of the Itô formula is technical, and beyond the scope of this course, we provide a quick heuristic argument that illustrates the main idea clearly.

INTUITION BEHIND THE ITÔ FORMULA. Suppose that the function f is only a function of x and doesn't depend on t, and X is a standard Brownian motion (i.e. $b = 0$ and $\sigma = 1$). In this case proving Itô's formula reduces to proving

$$f(W(T)) - f(W(0)) = \int_0^T f'(W(t))\,dW(t) + \frac{1}{2}\int_0^T f''(W(t))\,dt\,.$$

Let $P = \{0 = t_0 < t_1 < \cdots < t_n = T\}$ be a partition of $[0,T]$. Taylor expanding f to second order gives

$$\begin{aligned} f(W(T)) - f(W(0)) &= \sum_{i=0}^{n-1} f(W(t_{i+1})) - f(W(t_i)) \\ &= \sum_{i=0}^{n-1} f'(W(t_i))(\Delta_i W) + \frac{1}{2}\sum_{i=0}^{n-1} f''(W(t_i))(\Delta_i W)^2 + \frac{1}{2}\sum_{i=0}^{n-1} o\bigl((\Delta_i W)^2\bigr)\,, \end{aligned} \tag{5.5}$$

where the last sum on the right is the remainder from the Taylor expansion.

Note the first sum on the right of (5.5) converges to the Itô integral

$$\int_0^T f'(W(t))\,dW(t)\,.$$

For the second sum on the right of (5.5), note

$$f''(W(t_i))(\Delta_i W)^2 = f''(W(t_i))(t_{i+1} - t_i) + f''(W(t_i))\bigl[(\Delta_i W)^2 - (t_{i+1} - t_i)\bigr]\,.$$

After summing over i, first term on the right converges to the Riemann integral $\int_0^T f''(W(t))\,dt$. The second term on the right is similar to what we had when computing the quadratic variation of W. The variance of $\xi_i \stackrel{\text{def}}{=} (\Delta_i W)^2 - (t_{i+1} - t_i)$ is of order $(t_{i+1} - t_i)^2$. Thus we expect that the second term above, when summed over i, converges to 0.

Finally each summand in the remainder term (the last term on the right of (5.5)) is smaller than $(\Delta_i W)^2$. (If, for instance, f is three times continuously differentiable in x, then each summand in the remainder term is of order $(\Delta_i W)^3$.) Consequently, when summed over i this should converge to 0 □

6. A few examples using Itô's formula

Technically, as soon as you know Itô's formula you can "jump right in" and derive the Black-Scholes equation. However, because of the importance of Itô's formula, we work out a few simpler examples first.

EXAMPLE 6.1. Compute the quadratic variation of $W(t)^2$.

SOLUTION. Let $f(t,x) = x^2$. Then, by Itô's formula,

$$\begin{aligned} d\bigl(W(t)^2\bigr) &= df(t,W(t)) \\ &= \partial_t f(t,W(t))\,dt + \partial_x f(t,W(t))\,dW(t) + \frac{1}{2}\partial_x^2 f(t,W(t))\,dt \\ &= 2W(t)\,dW(t) + dt\,. \end{aligned}$$

Or, in integral form,

$$W(T)^2 - W(0)^2 = W(T)^2 = 2\int_0^T W(t)\,dW(t) + T\,.$$

---

sealed envelope to the French Academy of Sciences, because he did not want it to "fall into the wrong hands". When he was about to be captured by the Germans he burnt his mathematical notes and killed himself.

The sealed envelope was opened in 2000 which revealed that he had a treatment of stochastic Calculus that was essentially equivalent to Itô's. In posthumous recognition, Itô's formula is now referred to as the Itô-Doeblin formula by many authors.

[2] Recall a function $f = f(t,x)$ is said to be $C^{1,2}$ if it is C^1 in t (i.e. differentiable with respect to t and $\partial_t f$ is continuous), and C^2 in x (i.e. twice differentiable with respect to x and $\partial_x f$, $\partial_x^2 f$ are both continuous).

Now the second term on the right has finite first variation, and won't affect our computations for quadratic variation. The first term is an martingale whose quadratic variation is $\int_0^T W(t)^2\,dt$, and so

$$[W^2, W^2](T) = 4\int_0^T W(t)^2\,dt\,. \qquad \square$$

REMARK 6.2. Note the above also tells you

$$2\int_0^T W(t)\,dW(t) = W(T)^2 - T\,.$$

EXAMPLE 6.3. Let $M(t) = W(t)$ and $N(t) = W(t)^2 - t$. We know M and N are martingales. Is MN a martingale?

SOLUTION. Note $M(t)N(t) = W(t)^3 - tW(t)$. By Itô's formula,

$$d(MN) = -W(t)\,dt + (3W(t)^2 - t)\,dW(t) + 3W(t)\,dt\,.$$

Or in integral form

$$M(t)N(t) = \int_0^t 2W(s)\,ds + \int_0^t (3W(s)^2 - s)\,dW(s)\,.$$

Now the second integral on the right is a martingale, but the first integral most certainly is not. So MN can not be a martingale. $\square$

REMARK 6.4. Note, above we changed the integration variable from t to s. This is perfectly legal – the variable with which you integrate with respect to is a dummy variable (just line regular Riemann integrals) and you can replace it what your favourite (unused!) symbol.

REMARK 6.5. It's worth pointing out that the Itô integral $\int_0^t \Delta(s)\,dW(s)$ is always a martingale (under the finiteness condition (4.5)). However, the Riemann integral $\int_0^t b(s)\,ds$ is only a martingale if $b = 0$ identically.

PROPOSITION 6.6. *If $f = f(t,x)$ is $C_b^{1,2}$ then the process*

$$M(t) \stackrel{\text{def}}{=} f(t, W(t)) - \int_0^t \Big(\partial_t f(s, W(s)) + \frac{1}{2}\partial_x^2 f(s, W(s))\Big)\,ds$$

is a martingale.

REMARK 6.7. We'd seen this earlier, and the proof involved computing the conditional expectations directly and checking an algebraic identity involving the density of the normal distribution. With Itô's formula, the proof is "immediate".

PROOF. By Itô's formula (in integral form)

$$\begin{aligned} &f(t, W(t)) - f(0, W(0)) \\ &\quad= \int_0^t \partial_t f(s, W(s))\,ds + \int_0^t \partial_x f(s, W(s))\,dW(s) + \frac{1}{2}\int_0^t \partial_x^2 f(s, W(s))\,ds \\ &\quad= \int_0^t \Big(\partial_t f(s, W(s)) + \frac{1}{2}\partial_x^2 f(s, W(s))\Big)\,ds + \int_0^t \partial_x f(s, W(s))\,dW(s)\,. \end{aligned}$$

Substituting this we see

$$M(t) = f(0, W(0)) + \int_0^t \partial_x f(s, W(s))\,dW(s)\,,$$

which is a martingale. $\square$

REMARK 6.8. Note we said $f \in C_b^{1,2}$ to "cover our bases". Recall for Itô integrals to be martingales, we need the finiteness condition (4.5) to hold. This will certainly be the case if $\partial_x f$ is bounded, which is why we made this assumption. The result above is of course true under much more general assumptions.

EXAMPLE 6.9. Let $X(t) = t\sin(W(t))$. Is $X^2 - [X, X]$ a martingale?

SOLUTION. Let $f(t,x) = t\sin(x)$. Observe $X(t) = f(t, W(t))$, $\partial_t f = \sin x$, $\partial_x f = t\cos x$, and $\partial_x^2 f = -t\sin x$. Thus by Itô's formula,

$$\begin{aligned} dX(t) &= \partial_t f(t, W(t))\,dt + \partial_x f(t, W(t))\,dW(t) + \frac{1}{2}\partial_x^2 f(t, W(t))\,d[W, W](t) \\ &= \sin(W(t))\,dt + t\cos(W(t))\,dW(t) - \frac{1}{2}t\sin(W(t))\,dt\,, \end{aligned}$$

and so

$$d[X, X](t) = t^2\cos^2(W(t))\,dt\,.$$

Now let $g(x) = x^2$ and apply Itô's formula to compute $dg(X(t))$. This gives

$$dX(t)^2 = 2X(t)\,dX(t) + d[X, X](t)$$

and so

$$\begin{aligned} &d(X(t)^2 - [X, X]) = 2X(t)\,dX(t) \\ &\quad= 2t\sin(t)\Big(\sin(W(t)) - \frac{t\sin(W(t))}{2}\Big)\,dt + 2t\sin(t)\big(t\cos(W(t))\big)\,dW(t)\,. \end{aligned}$$

Since the dt term above isn't 0, $X(t)^2 - [X, X]$ can not be a martingale. $\square$

Recall we said earlier (Theorem 3.4) that for any martingale M, $M^2 - [M, M]$ is a martingale. In the above example X is not a martingale, and so there is no contradiction when we show that $X^2 - [X, X]$ is not a martingale. If M is a martingale, Itô's formula can be used to "prove"[3] that $M^2 - [M, M]$ is a martingale.

PROPOSITION 6.10. *Let $M(t) = \int_0^t \sigma(s)\,dW(s)$. Then $M^2 - [M, M]$ is a martingale.*

PROOF. Let $N(t) = M(t)^2 - [M, M](t)$. Observe that by Itô's formula,

$$d(M(t)^2) = 2M(t)\,dM(t) + d[M, M](t)\,.$$

Hence

$$dN = 2M(t)\,dM(t) + d[M, M](t) - d[M, M](t) = 2M(t)\sigma(t)\,dW(t)\,.$$

Since there is no "dt" term and Itô integrals are martingales, N is a martingale. $\square$

[3]We used the fact that $M^2 - [M, M]$ is a martingale crucially in the construction of Itô integrals, and hence in proving Itô's formula. Thus proving $M^2 - [M, M]$ is a martingale using the Itô's formula is circular and not a valid proof. It is however instructive, and helps with building intuition, which is why it is presented here.

7. Review Problems

PROBLEM 7.1. If $0 \leqslant r < s < t$, compute $\boldsymbol{E}\big(W(r)W(s)W(t)\big)$.

PROBLEM 7.2. Define the processes X, Y, Z by

$$X(t) = \int_0^{W(t)} e^{-s^2}\, ds\,, \quad Y(t) = \exp\Big(\int_0^t W(s)\, ds\Big)\,, \quad Z(t) = tX(t)^2$$

Decompose each of these processes as the sum of a martingale and a process of finite first variation. What is the quadratic variation of each of these processes?

PROBLEM 7.3. Define the processes X, Y by

$$X(t) \stackrel{\text{def}}{=} \int_0^t W(s)\, ds\,, \quad Y(t) \stackrel{\text{def}}{=} \int_0^t W(s)\, dW(s)\,.$$

Given $0 \leqslant s < t$, compute the conditional expectations $\boldsymbol{E}(X(t)|\mathcal{F}_s)$, and $\boldsymbol{E}(Y(t)|\mathcal{F}_s)$.

PROBLEM 7.4. Let $M(t) = \int_0^t W(s)\, dW(s)$. Find a function f such that

$$E(t) \stackrel{\text{def}}{=} \exp\Big(M(t) - \int_0^t f(s, W(s))\, ds\Big)$$

is a martingale.

PROBLEM 7.5. Suppose $\sigma = \sigma(t)$ is a deterministic (i.e. non-random) process, and X is the Itô process defined by

$$X(t) = \int_0^t \sigma(u)\, dW(u)\,.$$

(a) Given $\lambda, s, t \in \mathbb{R}$ with $0 \leqslant s < t$ compute $\boldsymbol{E}(e^{\lambda(X(t)-X(s))} \mid \mathcal{F}_s)$.
(b) If $r \leqslant s$ compute $\boldsymbol{E} \exp(\lambda X(r) + \mu(X(t) - X(s)))$.
(c) What is the joint distribution of $(X(r), X(t) - X(s))$?
(d) *(Lévy's criterion)* If $\sigma(u) = \pm 1$ for all u, then show that X is a standard Brownian motion.

PROBLEM 7.6. Define the process X, Y by

$$X = \int_0^t s\, dW(s)\,, \quad Y = \int_0^t W(s)\, ds\,.$$

Find a formula for $\boldsymbol{E}X(t)^n$ and $\boldsymbol{E}Y(t)^n$ for any $n \in \mathbb{N}$.

PROBLEM 7.7. Let $M(t) = \int_0^t W(s)\, dW(s)$. For $s < t$, is $M(t) - M(s)$ independent of $\mathcal{F}_s$? Justify.

PROBLEM 7.8. Determine whether the following identities are true or false, and justify your answer.

(a) $e^{2t}\sin(2W(t)) = 2\int_0^t e^{2s}\cos(2W(s))\, dW(s)$.
(b) $|W(t)| = \int_0^t \operatorname{sign}(W(s))\, dW(s)$. (Recall $\operatorname{sign}(x) = 1$ if $x > 0$, $\operatorname{sign}(x) = -1$ if $x < 0$ and $\operatorname{sign}(x) = 0$ if $x = 0$.)

8. The Black Scholes Merton equation.

The price of an asset with a constant rate of return α is given by

$$dS(t) = \alpha S(t)\, dt\,.$$

To account for noisy fluctuations we model stock prices by adding the term $\sigma S(t)\, dW(t)$ to the above:

$$dS(t) = \alpha S(t)\, dt + \sigma S(t) dW(t)\,. \tag{8.1}$$

The parameter α is called the *mean return rate* or the *percentage drift*, and the parameter σ is called the *volatility* or the *percentage volatility*.

DEFINITION 8.1. A stochastic process S satisfying (8.1) above is called a *geometric Brownian motion*.

The reason S is called a geometric Brownian motion is as follows. Set $Y = \ln S$ and observe

$$dY(t) = \frac{1}{S(t)}\, dS(t) - \frac{1}{2S(t)^2}\, d[S,S](t) = \Big(\alpha - \frac{\sigma^2}{2}\Big)\, dt + \sigma\, dW(t)\,.$$

If $\alpha = \sigma^2/2$ then $Y = \ln S$ is simply a Brownian motion.

We remark, however, that our application of Itô's formula above is not completely justified. Indeed, the function $f(x) = \ln x$ is *not* differentiable at $x = 0$, and Itô's formula requires f to be at least C^2. The reason the application of Itô's formula here is valid is because the process S *never* hits the point $x = 0$, and at all other points the function f is infinitely differentiable.

The above also gives us an explicit formula for S. Indeed,

$$\ln\Big(\frac{S(t)}{S(0)}\Big) = \Big(\alpha - \frac{\sigma^2}{2}\Big)t + \sigma W(t)\,,$$

and so

$$S(t) = S(0)\exp\Big(\Big(\alpha - \frac{\sigma^2}{2}\Big)t + \sigma W(t)\Big)\,. \tag{8.2}$$

Now consider a European call option with strike price K and maturity time T. This is a security that allows you the option (not obligation) to buy S at price K and time T. Clearly the price of this option at time T is $(S(T) - K)^+$. Our aim is to compute the *arbitrage free*[4] price of such an option at time $t < T$.

Black and Scholes realised that the price of this option at time t should only depend on the asset price $S(t)$, and the current time t (or more precisely, the time to maturity $T - t$), and of course the model parameters α, σ. In particular, the option price does not depend on the price history of S.

THEOREM 8.2. *Suppose we have an arbitrage free financial market consisting of a money market account with constant return rate r, and a risky asset whose price is given by S. Consider a European call option with strike price K and maturity T.*

[4]In an arbitrage free market, we say p is the arbitrage free price of a non traded security if given the opportunity to trade the security at price p, the market is still arbitrage free. (Recall a financial market is said to be *arbitrage free* if there doesn't exist a self-financing portfolio X with $X(0) = 0$ such that at some $t > 0$ we have $X(t) \geqslant 0$ and $\boldsymbol{P}(X(t) > 0) > 0$.)

(1) *If $c = c(t, x)$ is a function such that at any time $t \leqslant T$, the arbitrage free price of this option is $c(t, S(t))$, then c satisfies*

$$\partial_t c + rx\partial_x c + \frac{\sigma^2 x^2}{2}\partial_x^2 c - rc = 0 \qquad x > 0,\ t < T\,, \tag{8.3}$$

$$c(t, 0) = 0 \qquad t \leqslant T\,, \tag{8.4}$$

$$c(T, x) = (x - K)^+ \qquad x \geqslant 0\,. \tag{8.5}$$

(2) *Conversely, if c satisfies (8.3)–(8.5) then $c(t, S(t))$ is the arbitrage free price of this option at any time $t \leqslant T$.*

Remark 8.3. Since α, σ and T are fixed, we suppress the explicit dependence of c on these quantities.

Remark 8.4. The above result assumes the following:

(1) The market is *frictionless* (i.e. there are no transaction costs).
(2) The asset is liquid and fractional quantities of it can be traded.
(3) The borrowing and lending rate are both r.

Remark 8.5. Even though the asset price $S(t)$ is random, the function c is a deterministic (non-random) function. The option price, however, is $c(t, S(t))$, which is certainly random.

Remark 8.6. Equation (8.3)–(8.5) are the *Black Scholes Merton PDE*. This is a *partial differential equation*, which is a differential equation involving derivatives with respect to more than one variable. Equation (8.3) governs the evolution of c for $x \in (0, \infty)$ and $t < T$. Equation (8.5) specifies the terminal condition at $t = T$, and equation (8.4) specifies a boundary condition at $x = 0$.

To be completely correct, one also needs to add a boundary condition as $x \to \infty$ to the system (8.3)–(8.5). When x is very large, the call option is deep in the money, and will very likely end in the money. In this case the replicating portfolio should be long one share of the asset and short $e^{-r(T-t)}K$, the discounted strike price, in cash. This means that when x is very large, $c(x, t) \approx x - Ke^{-r(T-t)}$, and hence a boundary condition at $x = \infty$ can be obtained by supplementing (8.4) with

$$\lim_{x\to\infty}\bigl(c(t, x) - (x - Ke^{-r(T-t)})\bigr) = 0\,. \tag{8.4$'$}$$

Remark 8.7. The system (8.3)–(8.5) can be solved explicitly using standard calculus by substituting $y = \ln x$ and converting it into the *heat equation*, for which the solution is explicitly known. This gives the Black-Scholes-Merton formula

$$c(t, x) = xN(d_+(T - t, x)) - Ke^{-r(T-t)}N(d_-(T - t, x)) \tag{8.6}$$

where

$$d_\pm(\tau, x) \stackrel{\text{def}}{=} \frac{1}{\sigma\sqrt{\tau}}\Bigl(\ln\Bigl(\frac{x}{K}\Bigr) + \Bigl(r \pm \frac{\sigma^2}{2}\Bigr)\tau\Bigr)\,, \tag{8.7}$$

and

$$N(x) \stackrel{\text{def}}{=} \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-y^2/2}\,dy\,, \tag{8.8}$$

is the CDF of a standard normal variable.

Even if you're unfamiliar with the techniques involved in arriving at the solution above, you can certainly check that the function c given by (8.6)–(8.7) above satisfies (8.3)–(8.5). Indeed, this is a direct calculation that only involves patience and a careful application of the chain rule. We will, however, derive (8.6)–(8.7) later using risk neutral measures.

We will prove Theorem 8.2 by using a *replicating portfolio*. This is a portfolio (consisting of cash and the risky asset) that has exactly the same cash flow at maturity as the European call option that needs to be priced. Specifically, let $X(t)$ be the value of the replicating portfolio and $\Delta(t)$ be the number of shares of the asset held. The remaining $X(t) - S(t)\Delta(t)$ will be invested in the money market account with return rate r. (It is possible that $\Delta(t)S(t) > X(t)$, in which means we borrow money from the money market account to invest in stock.) For a replicating portfolio, the trading strategy Δ should be chosen in a manner that ensures that we have the same cash flow as the European call option. That is, we must have $X(T) = (S(T) - K)^+ = c(T, S(T))$. Now the arbitrage free price is precisely the value of this portfolio.

Remark 8.8. Through the course of the proof we will see that given the function c, the number of shares of S the replicating portfolio should hold is given by the *delta hedging rule*

$$\Delta(t) = \partial_x c(t, S(t))\,. \tag{8.9}$$

Remark 8.9. Note that there is no α dependence in the system (8.3)–(8.5), and consequently the formula (8.6) does not depend on α. At first sight, this might appear surprising. (In fact, Black and Scholes had a hard time getting the original paper published because the community couldn't believe that the final formula is independent of α.) The fact that (8.6) is independent of α can be heuristically explained by the fact that the replicating portfolio also holds the same asset: thus a high mean return rate will help both an investor holding a call option and an investor holding the replicating portfolio. (Of course this isn't the entire story, as one has to actually write down the dependence and check that an investor holding the call option benefits exactly as much as an investor holding the replicating portfolio. This is done below.)

Proof of Theorem 8.2 part 1. If $c(t, S(t))$ is the arbitrage free price, then, by definition

$$c(t, S(t)) = X(t)\,, \tag{8.10}$$

where $X(t)$ is the value of a replicating portfolio. Since our portfolio holds $\Delta(t)$ shares of S and $X(t) - \Delta(t)S(t)$ in a money market account, the evolution of the value of this portfolio is given by

$$\begin{aligned} dX(t) &= \Delta(t)\,dS(t) + r\bigl(X(t) - \Delta(t)S(t)\bigr)\,dt\\ &= \bigl(rX(t) + (\alpha - r)\Delta(t)S(t)\bigr)\,dt + \sigma\Delta(t)S(t)\,dW(t)\,. \end{aligned}$$

Also, by Itô's formula we compute

$$dc(t, S(t)) = \partial_t c(t, S(t))\,dt + \partial_x c(t, S(t))\,dS(t) + \frac{1}{2}\partial_x^2 c(t, S(t))\,d[S, S](t)\,,$$

$$= \Big(\partial_t c + \alpha S \partial_x c + \frac{1}{2}\sigma^2 S^2 \partial_x^2 c\Big)\, dt + \partial_x c\, \sigma S\, dW(t)$$

where we suppressed the $(t, S(t))$ argument in the last line above for convenience.

Equating $dc(t, S(t)) = dX(t)$ gives

$$\big(rX(t) + (\alpha - r)\Delta(t)S(t)\big)\, dt + \sigma\Delta(t)S(t)\, dW(t)$$
$$= \Big(\partial_t c + \alpha S \partial_x c + \frac{1}{2}\sigma^2 S^2 \partial_x^2 c\Big)\, dt + \partial_x c\, \sigma S\, dW(t)\,.$$

Using uniqueness of the semi-martingale decomposition (Proposition 5.5) we can equate the dW and the dt terms respectively. Equating the dW terms gives the delta hedging rule (8.9). Writing $S(t) = x$ for convenience, equating the dt terms and using (8.10) gives (8.3). Since the payout of the option is $(S(T) - K)^+$ at maturity, equation (8.5) is clearly satisfied.

Finally if $S(t_0) = 0$ at one particular time, then we must have $S(t) = 0$ at all times, otherwise we would have an arbitrage opportunity. (This can be checked directly from the formula (8.2) of course.) Consequently the arbitrage free price of the option when $S = 0$ is 0, giving the *boundary condition* (8.4). Hence (8.3)–(8.5) are all satisfied, finishing the proof. $\square$

Proof of Theorem 8.2 part 2. For the converse, we suppose c satisfies the system (8.3)–(8.5). Choose $\Delta(t)$ by the delta hedging rule (8.9), and let X be a portfolio with initial value $X(0) = c(0, S(0))$ that holds $\Delta(t)$ shares of the asset at time t and the remaining $X(t) - \Delta(t)S(t)$ in cash. We claim that X is a replicating portfolio (i.e. $X(T) = (S(T) - K)^+$ almost surely) and $X(t) = c(t, S(t))$ for all $t \leqslant T$. Once this is established $c(t, S(t))$ is the arbitrage free price as desired.

To show X is a replicating portfolio, first claim that $X(t) = c(t, S(t))$ for all $t < T$. To see this, let $Y(t) = e^{-rt}X(t)$ be the discounted value of X. (That is, $Y(t)$ is the value of $X(t)$ converted to cash at time $t = 0$.) By Itô's formula, we compute

$$dY(t) = -rY(t)\, dt + e^{-rt} dX(t)$$
$$= e^{-rt}(\alpha - r)\Delta(t)S(t)\, dt + e^{-rt}\sigma\Delta(t)S(t)\, dW(t)\,.$$

Similarly, using Itô's formula, we compute

$$d\big(e^{-rt}c(t, S(t))\big) = e^{-rt}\Big(-rc + \partial_t c + \alpha S\partial_x c + \frac{1}{2}\sigma^2 S^2 \partial_x^2 c\Big)\, dt + e^{-rt}\partial_x c\, \sigma S\, dW(t)\,.$$

Using (8.3) this gives

$$d\big(e^{-rt}c(t, S(t))\big) = e^{-rt}(\alpha - r)S\partial_x c\, dt + e^{-rt}\partial_x c\, \sigma S\, dW(t) = dY(t)\,,$$

since $\Delta(t) = \partial_x c(t, S(t))$ by choice. This forces

$$e^{-rt}X(t) = X(0) + \int_0^t dY(s) = X(0) + \int_0^t d\big(e^{-rs}c(s, S(s))\big)$$
$$= X(0) + e^{-rt}c(t, S(t)) - c(0, S(0)) = e^{-rt}c(t, S(t))\,,$$

since we chose $X(0) = c(0, S(0))$. This forces $X(t) = c(t, S(t))$ for all $t < T$, and by continuity also for $t = T$. Since $c(T, S(T)) = (S(T) - K)^+$ we have $X(T) = (S(T) - K)^+$ showing X is a replicating portfolio, concluding the proof. $\square$

Remark 8.10. In order for the application of Itô's formula to be valid above, we need $c \in C^{1,2}$. This is certainly false at time T, since $c(T, x) = (x - K)^+$ which is not even differentiable, let alone twice continuously differentiable. However, if c satisfies the system (8.3)–(8.5), then it turns out that for every $t < T$ the function c will be infinitely differentiable with respect to x. This is why our proof first shows that $c(t, S(t)) = X(t)$ for $t < T$ and not directly that $c(t, S(t)) = X(t)$ for all $t \leqslant T$.

Remark 8.11 (Put Call Parity). The same argument can be used to compute the arbitrage free price of European put options (i.e. the option to sell at the strike price, instead of buying). However, once the price of the price of a call option is computed, the *put call parity* can be used to compute the price of a put.

Explicitly let $p = p(t, x)$ be a function such that at any time $t \leqslant T$, $p(t, S(t))$ is the arbitrage free price of a European put option with strike price K. Consider a portfolio X that is long a call and short a put (i.e. buy one call, and sell one put). The value of this portfolio at time $t < T$ is

$$X(t) = c(t, S(t)) - p(t, S(t))$$

and at maturity we have[5]

$$X(T) = (S(T) - K)^+ - (K - S(T))^+ = S(T) - K\,.$$

This payoff can be replicated using a portfolio that holds one share of the asset and borrows Ke^{-rT} in cash (with return rate r) at time 0. Thus, in an arbitrage free market, we should have

$$c(t, S(t)) - p(t, S(t)) = X(t) = S(t) - Ke^{-r(T-t)}\,.$$

Writing x for $S(t)$ this gives the *put call parity* relation

$$c(t, x) - p(t, x) = x - Ke^{-r(T-t)}\,.$$

Using this the price of a put can be computed from the price of a call.

We now turn to understanding properties of c. The partial derivatives of c with respect to t and x measure the sensitivity of the option price to changes in the time to maturity and spot price of the asset respectively. These are called "the Greeks":

(1) The *delta* is defined to be $\partial_x c$, and is given by

$$\partial_x c = N(d_+) + xN'(d_+)d_+' - Ke^{-r\tau}N'(d_-)d_-'\,.$$

where $\tau = T - t$ is the time to maturity. Recall $d_\pm = d_\pm(\tau, x)$, and we suppressed the (τ, x) argument above for notational convenience. Using the formulae (8.6)–(8.8) one can verify

$$d_+' = d_-' = \frac{1}{x\sigma\sqrt{\tau}} \qquad \text{and} \qquad xN'(d_+) = Ke^{-r\tau}N'(d_-)\,,$$

and hence the delta is given by

$$\partial_x c = N(d_+)\,.$$

[5] A *forward contract* requires the holder to buy the asset at price K at maturity. The value of this contract at maturity is exactly $S(T) - K$, and so a portfolio that is long a call and short a put has exactly the same cash flow as a forward contract.

Recall that the *delta hedging rule* (equation (8.9)) explicitly tells you that the replicating portfolio should hold precisely $\partial_x c(t, S(t))$ shares of the risky asset and the remainder in cash.

(2) The *gamma* is defined to be $\partial_x^2 c$, and is given by

$$\partial_x^2 c = N'(d_+) d'_+ = \frac{1}{x\sigma\sqrt{2\pi\tau}} \exp\Big(\frac{-d_+^2}{2}\Big).$$

(3) Finally the *theta* is defined to be $\partial_t c$, and simplifies to

$$\partial_t c = -rKe^{-r\tau} N(d_-) - \frac{\sigma x}{2\sqrt{\tau}} N'(d_+)$$

PROPOSITION 8.12. *The function $c(t, x)$ is convex and increasing as a function of x, and is* decreasing *as a function of t.*

PROOF. This follows immediately from the fact that $\partial_x c > 0$, $\partial_x^2 c > 0$ and $\partial_t c < 0$. ∎

REMARK 8.13 (Hedging a short call). Suppose you sell a call option valued at $c(t, x)$, and want to create a replicating portfolio. The delta hedging rule calls for $x\partial_x c(t, x)$ of the portfolio to be invested in the asset, and the rest in the money market account. Consequently the value of your money market account is

$$c(t, x) - x\partial_x c = xN(d_+) - Ke^{-r\tau} N(d_-) - xN(d_+) = -Ke^{-r\tau} N(d_-) < 0\,.$$

Thus to properly hedge a short call you will have to borrow from the money market account and invest it in the asset. As $t \to T$ you will end up selling shares of the asset if $x < K$, and buying shares of it if $x > K$, so that at maturity you will hold the asset if $x > K$ and not hold it if $x < K$. To hedge a long call you do the opposite.

REMARK 8.14 (Delta neutral and Long Gamma). Suppose at some time t the price of a stock is x_0. We short $\partial_x c(t, x_0)$ shares of this stock buy the call option valued at $c(t, x_0)$. We invest the balance $M = x_0\partial_x c(t, x_0) - c(t, x_0)$ in the money market account. Now if the stock price changes to x, *and we do not change our position*, then the value of our portfolio will be

$$\begin{aligned} c(t, x) - \partial_x c(t, x_0)x + M &= c(t, x) - x\partial_x c(t, x_0) + x_0\partial_x c(t, x_0) - c(t, x_0) \\ &= c(t, x) - \big(c(t, x_0) + (x - x_0)\partial_x c(t, x_0)\big)\,. \end{aligned}$$

Note that the line $y = c(t, x_0) + (x - x_0)\partial_x c(t, x_0)$ is the equation for the tangent to the curve $y = c(t, x)$ at the point $(x_0, c(t, x_0))$. For this reason the above portfolio is called *delta neutral*.

Note that any convex function lies entirely above its tangent. Thus, under instantaneous changes of the stock price (both rises and falls), we will have

$$c(t, x) - \partial_x c(t, x_0)x + M > 0\,, \quad \text{both for } x > x_0 \text{ and } x < x_0.$$

For this reason the above portfolio is called *long gamma*.

Note, even though under instantaneous price changes the value of our portfolio always rises, this is *not* an arbitrage opportunity. The reason for this is that as time increases c decreases since $\partial_t c < 0$. The above instantaneous argument assumed c is constant in time, which it most certainly is not!

9. Multi-dimensional Itô calculus.

Finally we conclude this chapter by studying Itô calculus in higher dimensions. Let X, Y be Itô process. We typically expect X, Y will have finite and non-zero quadratic variation, and hence both the increments $X(t+\delta t) - X(t)$ and $Y(t+\delta t) - Y(t)$ should typically be of size $\sqrt{\delta}$. If we multiply these and sum over some finite interval $[0, T]$, then we would have roughly $T/\delta t$ terms each of size δt, and expect that this converges as $\delta t \to 0$. The limit is called the joint quadratic variation.

DEFINITION 9.1. Let X and Y be two Itô processes. We define the *joint quadratic variation* of X, Y, denoted by $[X, Y]$ by

$$[X, Y](T) = \lim_{\|P\| \to 0} \sum_{i=0}^{n-1} (X(t_{i+1}) - X(t_i))(Y(t_{i+1}) - Y(t_i))\,,$$

where $P = \{0 = t_1 < t_1 \cdots < t_n = T\}$ is a partition of $[0, T]$.

Using the identity

$$4ab = (a + b)^2 - (a - b)^2$$

we quickly see that

$$[X, Y] = \frac{1}{4}\big([X + Y, X + Y] - [X - Y, X - Y]\big)\,. \tag{9.1}$$

Using this and the properties we already know about quadratic variation, we can quickly deduce the following.

PROPOSITION 9.2 (Product rule). *If X and Y are two Itô processes then*

$$d(XY) = X\,dY + Y\,dX + d[X, Y]\,. \tag{9.2}$$

PROOF. By Itô's formula

$$\begin{aligned} d(X + Y)^2 &= 2(X + Y)\,d(X + Y) + d[X + Y, X + Y] \\ &= 2X\,dX + 2Y\,dY + 2X\,dY + 2Y\,dX + d[X + Y, X + Y]\,. \end{aligned}$$

Similarly

$$d(X - Y)^2 = 2X\,dX + 2Y\,dY - 2X\,dY - 2Y\,dX + d[X - Y, X - Y]\,.$$

Since

$$4\,d(XY) = d(X + Y)^2 - d(X - Y)^2\,,$$

we obtain (9.2) as desired. ∎

As with quadratic variation, processes of finite variation do not affect the joint quadratic variation.

PROPOSITION 9.3. *If X is and Itô process, and B is a continuous adapted process with finite variation, then $[X, B] = 0$.*

PROOF. Note $[X \pm B, X \pm B] = [X, X]$ and hence $[X, B] = 0$. ∎

With this, we can state the higher dimensional Itô formula. Like the one dimensional Itô formula, this is a generalization of the chain rule and has an extra correction term that involves the joint quadratic variation.

THEOREM 9.4 (Itô-Doeblin formula). *Let $X_1, \ldots, X_n$ be n Itô processes and set $X = (X_1, \ldots, X_n)$. Let $f \colon [0,\infty) \times \mathbb{R}^n$ be C^1 in the first variable, and C^2 in the remaining variables. Then*

$$f(T, X(T)) = f(0, X(0)) + \int_0^T \partial_t f(t, X(t))\,dt + \sum_{i=1}^N \int_0^T \partial_i f(t, X(t))\,dX_i(t) + \frac{1}{2} \sum_{i,j=1}^N \int_0^T \partial_i \partial_j f(t, X(t))\,d[X_i, X_j](t)\,,$$

REMARK 9.5. Here we think of $f = f(t, x_1, \ldots, x_n)$, often abbreviated as $f(t, x)$. The $\partial_i f$ appearing in the Itô formula above is the partial derivative of f with respect to x_i. As before, the $\partial_t f$ and $\partial_i f$ terms above are from the usual chain rule, and the last term is the extra Itô correction.

REMARK 9.6. In differential form Itô's formula says

$$d\big(f(t, X(t)\big) = \partial_t f(t, X(t))\,dt + \sum_{i=1}^n \partial_i f(t, X(t))\,dX_i(t) + \frac{1}{2} \sum_{i,j=1}^n \partial_i \partial_j f(t, X(t))\,d[X_i, X_j](t)\,.$$

For compactness, we will often omit the $(t, X(t))$ and write the above as

$$d\big(f(t, X(t)\big) = \partial_t f\,dt + \sum_{i=1}^n \partial_i f\,dX_i(t) + \frac{1}{2} \sum_{i,j=1}^n \partial_i \partial_j f\,d[X_i, X_j](t)\,.$$

REMARK 9.7. We will most often use this in two dimensions. In this case, writing X and Y for the two processes, the Itô formula reduces to

$$d\big(f(t, X(t), Y(t))\big) = \partial_t f\,dt + \partial_x f\,dX(t) + \partial_y f\,dY(t) + \frac{1}{2}\big(\partial_x^2 f\,d[X,X](t) + 2\partial_x\partial_y f\,d[X,Y](t) + \partial_y^2 f\,d[Y,Y](t)\big)\,.$$

INTUITION BEHIND THE ITÔ FORMULA. Let's assume we only have two Itô processes X, Y and $f = f(x, y)$ doesn't depend on t. Let $P = \{0 = t_0 < t_1 \cdots < t_m = T\}$ be a partition of the interval $[0, T]$ and write

$$f(X(T), Y(T)) - f(X(0), Y(0)) = \sum_{i=0}^{m-1} f(\xi_{i+1}) - f(\xi_i)\,,$$

where we write $\xi_i = (X(t_i), Y(t_i))$ for compactness. Now by Taylor's theorem,

$$\begin{aligned} f(\xi_{i+1}) - f(\xi_i) &= \partial_x f(\xi_i)\,\Delta_i X + \partial_y f(\xi_i)\,\Delta_i Y \\ &\quad + \frac{1}{2}\Big(\partial_x^2 f(\xi_i)(\Delta_i X)^2 + 2\partial_x\partial_y f(\xi_i)\,\Delta_i X\,\Delta_i Y + \partial_y^2 f(\xi_i)(\Delta_i Y)^2\Big) \\ &\quad + \text{higher order terms.} \end{aligned}$$

Here $\Delta_i X = X(t_{i+1}) - X(t_i)$ and $\Delta_i Y = Y(t_{i+1}) - Y(t_i)$. Summing over i, the first two terms converge to $\int_0^T \partial_x f(t)\,dX(t)$ and $\int_0^T \partial_y f(t)\,dY(t)$ respectively. The terms involving $(\Delta_i X)^2$ should to $\int_0^T \partial_x^2 f\,d[X,X]$ as we had with the one dimensional Itô formula. Similarly, the terms involving $(\Delta_i Y)^2$ should to $\int_0^T \partial_y^2 f\,d[Y,Y]$ as we had with the one dimensional Itô formula. For the cross term, we can use the identity (9.1) and quickly check that it converges to $\int_0^T \partial_x\partial_y f\,d[X,Y]$. The higher order terms are typically of size $(t_{i+1} - t_i)^{3/2}$ and will vanish as $\|P\| \to 0$. ☐

The most common use of the multi-dimensional Itô formula is when the Itô processes are specified as a combination of Itô integrals with respect to different Brownian motions. Thus our next goal is to find an effective way to to compute the joint quadratic variations in this case.

We've seen earlier (Theorems 3.4–3.5) that the quadratic variation of a martingale M is the unique increasing process that make $M^2 - [M, M]$ a martingale. A similar result holds for the joint quadratic variation.

PROPOSITION 9.8. *Suppose M, N are two continuous martingales with respect to a common filtration $\{\mathcal{F}_t\}$ such that $\boldsymbol{E} M(t)^2, \boldsymbol{E} N(t)^2 < \infty$.*

(1) The process $MN - [M, N]$ is also a martingale with respect to the same filtration.
(2) Moreover, if A is any continuous adapted process with finite first variation such that $A(0) = 0$ and $MN - A$ is a martingale with respect to $\{\mathcal{F}_t\}$, then $A = [M, N]$.

PROOF. The first part follows immediately from Theorem 3.4 and the fact that

$$4\big(MN - [M,N]\big) = (M+N)^2 - [M+N, M+N] - \big((M-N)^2 - [M-N, M-N]\big)\,.$$

The second part follows from the first part and uniqueness of the semi-martingale decomposition (Proposition 5.5). ☐

PROPOSITION 9.9 (Bi-linearity). *If X, Y, Z are three Itô processes and $\alpha \in \mathbb{R}$ is a (non-random) constant, then*

$$[X, Y + \alpha Z] = [X, Y] + \alpha [X, Z]\,. \tag{9.3}$$

PROOF. Let L, M and N be the martingale part in the Itô decomposition of X, Y and Z respectively. Clearly

$$L(M + \alpha N) - \big([L, M] + \alpha [L, N]\big) = \big(LM - [L, M]\big) + \alpha\big(LN - [L, N]\big)\,,$$

which is a martingale. Thus, since $[L, M] + \alpha [L, N]$ is also continuous adapted and increasing, by Proposition 9.8 we must have $[L, M + \alpha N] = [L, M] + \alpha [L, N]$. Since the joint quadratic variation of Itô processes can be computed in terms of their martingale parts alone, we obtain (9.3) as desired. ☐

For integrals with respect to Itô processes, we can compute the joint quadratic variation explicitly.

PROPOSITION 9.10. *Let X_1, X_2 be two Itô processes, σ_1, σ_2 be two adapted processes and let I_j be the integral defined by $I_j(t) = \int_0^t \sigma_j(s)\,dX_j(s)$ for $j \in \{1, 2\}$. Then*

$$[I_1, I_2](t) = \int_0^t \sigma_1(s)\sigma_2(s)\,d[X_1, X_2](s)\,.$$

PROOF. Let P be a partition and, as above, let $\Delta_i X = X(t_{i+1}) - X(t_i)$ denote the increment of a process X. Since

$$I_j(T) = \lim_{\|P\|\to 0} \sum_{i=0}^{n-1} \sigma_j(t_i)\Delta_i X_j\,, \quad\text{and}\quad [X_1, X_2] = \lim_{\|P\|\to 0} \sum_{i=0}^{n-1} \Delta_i X_1\, \Delta_i X_2\,,$$

we expect that $\sigma_j(t_i)\Delta_i(X_j)$ is a good approximation for $\Delta_i I_j$, and $\Delta_i X_1\, \Delta_i X_2$ is a good approximation for $\Delta_i [X_1, X_2]$. Consequently, we expect

$$\begin{aligned}[I_i, I_j](T) &= \lim_{\|P\|\to 0} \sum_{i=0}^{n-1} \Delta_i I_1\, \Delta_i I_2 = \lim_{\|P\|\to 0} \sum_{i=0}^{n-1} \sigma_1(t_i)\Delta_i X_1\, \sigma_2(t_i)\Delta_i X_2 \\ &= \lim_{\|P\|\to 0} \sum_{i=0}^{n-1} \sigma_1(t_i)\, \sigma_2(t_i)\Delta_i [X_1, X_2] = \int_0^T \sigma_1(t)\sigma_2(t)\, d[X_1, X_2](t)\,,\end{aligned}$$

as desired. ☐

PROPOSITION 9.11. *Let M, N be two continuous martingales with respect to a common filtration $\{\mathcal{F}_t\}$ such that $\boldsymbol{E}M(t)^2 < \infty$ and $\boldsymbol{E}N(t)^2 < \infty$. If M, N are independent, then $[M, N] = 0$.*

REMARK 9.12. If X and Y are independent, we know $\boldsymbol{E}XY = \boldsymbol{E}X\boldsymbol{E}Y$. However, we *need not* have $\boldsymbol{E}(XY \mid \mathcal{F}) = \boldsymbol{E}(X \mid F)\boldsymbol{E}(Y \mid F)$. So we can not prove the above result by simply saying

$$\boldsymbol{E}\big(M(t)N(t) \bigm| \mathcal{F}_s\big) = \boldsymbol{E}(M(t) \mid \mathcal{F}_s)\boldsymbol{E}(N(t) \mid \mathcal{F}_s) = M(s)N(s) \tag{9.4}$$

because M and N are independent. Thus MN is a martingale, and hence $[M, N] = 0$ by Proposition 9.8.

This reasoning is incorrect, even though the conclusion is correct. If you're not convinced, let me add that there exist martingales that are *not continuous* which are independent and have nonzero joint quadratic variation. The above argument, if correct, would certainly also work for martingales that are not continuous. The error in the argument is that the first equality in (9.4) need not hold even though M and N are independent.

PROOF. Let $P = \{0 = t_0 < t_1 < \cdots < t_n = T\}$ be a partition of $[0, T]$, $\Delta_i M = M(t_{i+1}) - M(t_i)$ and $\Delta_i N = N(t_{i+1}) - N(t_i)$. Observe

$$\begin{aligned}\boldsymbol{E}\Big(\sum_{i=0}^{n-1} \Delta_i M\, \Delta_i N\Big)^2 &= \boldsymbol{E}\sum_{i=0}^{n-1} (\Delta_i M)^2 (\Delta_i N)^2 \\ &\quad + 2\boldsymbol{E}\sum_{j=0}^{n-1}\sum_{i=0}^{j-1} \Delta_i M\, \Delta_i N \Delta_j M\, \Delta_j N\,.\end{aligned} \tag{9.5}$$

We claim the cross term vanishes because of independence of M and N. Indeed,

$$\begin{aligned}\boldsymbol{E}\Delta_i M\, \Delta_i N \Delta_j M\, \Delta_j N &= \boldsymbol{E}\big(\Delta_i M\, \Delta_j M\big)\boldsymbol{E}\big(\Delta_i N\, \Delta_j N\big) \\ &= \boldsymbol{E}\big(\Delta_i M \boldsymbol{E}(\Delta_j M \mid \mathcal{F}_{t_j})\big)\boldsymbol{E}\big(\Delta_i N\, \Delta_j N\big) = 0\,.\end{aligned}$$

Thus from (9.5)

$$\boldsymbol{E}\Big(\sum_{i=0}^{n-1} \Delta_i M\, \Delta_i N\Big)^2 = \boldsymbol{E}\sum_{i=0}^{n-1} (\Delta_i M)^2 (\Delta_i N)^2 \leqslant \boldsymbol{E}\Big(\max_i (\Delta_i M)^2\Big)\Big(\sum_{i=0}^{n-1} (\Delta_i N)^2\Big)$$

As $\|P\| \to 0$, $\max_i \Delta_i M \to 0$ because M is continuous, and $\sum_i (\Delta_i N)^2 \to [N, N](T)$. Thus we expect[6]

$$\boldsymbol{E}[M, N](T)^2 = \lim_{\|P\|\to 0} \boldsymbol{E}\Big(\sum_{i=0}^{n-1} \Delta_i M\, \Delta_i N\Big)^2 = 0\,,$$

finishing the proof. ☐

REMARK 9.13. The converse is false. If $[M, N] = 0$, it does *not* mean that M and N are independent. For example, if

$$M(t) = \int_0^t \mathbf{1}_{\{W(s) < 0\}}\, dW(s)\,, \qquad\text{and}\qquad N(t) = \int_0^t \mathbf{1}_{\{W(s) \geqslant 0\}}\, dW(s)\,,$$

then clearly $[M, N] = 0$. However,

$$M(t) + N(t) = \int_0^t 1\, dW(s) = W(t)\,,$$

and with a little work one can show that M and N are not independent.

DEFINITION 9.14. We say $W = (W_1, W_2, \dots, W_d)$ is a standard d-dimensional Brownian motion if:

(1) Each coordinate W_i is a standard (1-dimensional) Brownian motion.
(2) If $i \neq j$, the processes W_i and W_j are independent.

When working with a multi-dimensional Brownian motion, we usually choose the filtration to be that generated by *all* the coordinates.

DEFINITION 9.15. Let W be a d-dimensional Brownian motion. We define the filtration $\{\mathcal{F}_t^W\}$ by

$$\mathcal{F}_t^W = \sigma\Big(\bigcup_{\substack{s \leqslant t,\\ i \in \{1, \dots, d\}}} \sigma(W_i(s))\Big)$$

With $\{\mathcal{F}_t^W\}$ defined above note that:

(1) Each coordinate W_i is a martingale with respect to $\{\mathcal{F}_t^W\}$.
(2) For every $s < t$, the increment of each coordinate $W_i(t) - W_i(s)$ is independent of $\{\mathcal{F}_s^W\}$.

REMARK 9.16. Since W_i is independent of W_j when $i \neq j$, we know $[W_i, W_j] = 0$ if $i \neq j$. When $i = j$, we know $d[W_i, W_j] = dt$. We often express this concisely as

$$d[W_i, W_j](t) = \mathbf{1}_{\{i = j\}}\, dt\,.$$

[6] For this step we need to use $\lim_{\|P\|\to 0} \boldsymbol{E}(\cdots) = \boldsymbol{E}\lim_{\|P\|\to 0}(\cdots)$. To make this rigorous we need to apply the Lebesgue dominated convergence theorem. This is done by first assuming M and N are bounded, and then choosing a *localizing sequence* of stopping times, and a full discussion goes beyond the scope of these notes.

An extremely important fact about Brownian motion is that the converse of the above is also true.

THEOREM 9.17 (Lévy). *If $M = (M_1, M_2, \dots, M_d)$ is a continuous martingale such that $M(0) = 0$ and*

$$d[M_i, M_j](t) = \mathbf{1}_{\{i=j\}}\, dt\,,$$

then M is a d-dimensional Brownian motion.

PROOF. The main idea behind the proof is to compute the moment generating function (or characteristic function) of M, in the same way as in Problem 7.5. This can be used to show that $M(t) - M(s)$ is independent of $\mathcal{F}_s$ and $M(t) \sim N(0, tI)$, where I is the $d \times d$ identity matrix. □

EXAMPLE 9.18. If W is a 2-dimensional Brownian motion, then show that

$$B = \int_0^t \frac{W_1(s)}{|W(t)|}\, dW_1(s) + \int_0^t \frac{W_2(s)}{|W(t)|}\, dW_2(s)\,,$$

is also a Brownian motion.

PROOF. Since B is the sum of two Itô integrals, it is clearly a continuous martingale. Thus to show that B is a Brownian motion, it suffices to show that $[B, B](t) = t$. For this, define

$$X(t) = \int_0^t \frac{W_1(s)}{|W(t)|}\, dW_1(s) \qquad \text{and} \qquad Y(t) = \int_0^t \frac{W_2(s)}{|W(t)|}\, dW_2(s)\,,$$

and note

$$\begin{aligned} d[B, B](t) &= d[X + Y, X + Y](t) = d[X, X](t) + d[Y, Y](t) + 2d[X, Y](t) \\ &= \Big(\frac{W_1(t)^2}{|W(t)|^2} + \frac{W_2(t)^2}{|W(t)|^2}\Big)\, dt + 0 = dt\,. \end{aligned}$$

So by Lévy's criterion, B is a Brownian motion. □

EXAMPLE 9.19. Let W be a 2-dimensional Brownian motion and define

$$X = \ln(|W|^2) = \ln(W_1^2 + W_2^2)\,.$$

Compute dX. Is X a martingale?

SOLUTION. This is a bit tricky. First, if we set $f(x) = \ln|x|^2 = \ln(x_1^2 + x_2^2)$, then it is easy to check

$$\partial_i f = \frac{2x_i}{|x|^2} \qquad \text{and} \qquad \partial_1^2 f + \partial_2^2 f = 0\,.$$

Consequently,

$$dX(t) = \frac{2W_1(t)}{|W|^2}\, dW_1(t) + \frac{2W_2(t)}{|W|^2}\, dW_2(t)\,.$$

With this one would be tempted to say that since there are no dt terms above, X is a martingale. This, however, is false! Martingales have constant expectation, but

$$\begin{aligned} \boldsymbol{E}X(t) &= \frac{1}{2\pi t} \iint_{\mathbb{R}^2} \ln\big(x_1^2 + x_2^2\big) \exp\Big(-\frac{x_1^2 + x_2^2}{2t}\Big)\, dx_1\, dx_2 \\ &= \frac{1}{2\pi} \iint_{\mathbb{R}^2} \ln\big(t(y_1^2 + y_2^2)\big) \exp\Big(-\frac{y_1^2 + y_2^2}{2}\Big)\, dy_1\, dy_2 \\ &= \ln t + \frac{1}{2\pi} \iint_{\mathbb{R}^2} \ln\big(y_1^2 + y_2^2\big) \exp\Big(-\frac{y_1^2 + y_2^2}{2}\Big)\, dy_1\, dy_2 \xrightarrow{t \to \infty} \infty\,. \end{aligned}$$

Thus $\boldsymbol{E}X(t)$ is not constant in t, and so X can not be a martingale. □

REMARK 9.20. We have repeatedly used the fact that Itô integrals are martingales. The example above obtains X as an Itô integral, but can not be a martingale. The reason this doesn't contradict Theorem 4.2 is that in order for Itô integral $\int_0^t \sigma(s)\, dW(s)$ to be defined, we only need the finiteness condition $\int_0^t \sigma(s)^2\, ds < \infty$ almost surely. However, for an Itô integral to be a martingale, we need the stronger condition $\boldsymbol{E} \int_0^t \sigma(s)^2\, ds < \infty$ (given in (4.5)) to hold. This is precisely what fails in the previous example. The process X above is an example of a *local martingale* that is not a martingale, and we will encounter a similar situation when we study exponential martingales and risk neutral measures.

EXAMPLE 9.21. Let $f = f(t, x_1, \dots, x_d) \in C^{1,2}$ and W be a d-dimensional Brownian motion. Then Itô's formula gives

$$d\big(f(t, W(t))\big) = \Big(\partial_t f(t, W(t)) + \frac{1}{2}\Delta f(t, W(t))\Big)\, dt + \sum_{i=1}^d \partial_i f(t, W(t))\, dW_i(t)\,.$$

Here $\Delta f = \sum_1^d \partial_i^2 f$ is the Laplacian of f.

EXAMPLE 9.22. Consider a d-dimensional Brownian motion W, and n Itô processes $X_1, \dots, X_n$ which we write (in differential form) as

$$dX_i(t) = b_i(t)\, dt + \sum_{k=1}^d \sigma_{i,k}(t)\, dW_k(t)\,,$$

where each b_i and $\sigma_{i,j}$ are adapted processes. For brevity, we will often write b for the vector process $(b_1, \dots, b_n)$, σ for the *matrix* process $(\sigma_{i,j})$ and X for the n-dimensional Itô process $(X_1, \dots, X_n)$.

Now to compute $[X_i, X_j]$ we observe that $d[W_i, W_j] = dt$ if $i = j$ and 0 otherwise. Consequently,

$$d[X_i, X_j](t) = \sum_{k,l=1}^d \sigma_{i,k}\sigma_{j,l} \mathbf{1}_{\{k=l\}}\, dt = \sum_{k=1}^d \sigma_{i,k}(t)\sigma_{j,k}(t)\, dt\,.$$

Thus if f is any $C^{1,2}$ function, Itô formula gives

$$d\Big(f(t, X(t))\Big) = \Big(\partial_t f + \sum_{i=1}^n b_i \partial_i f\Big)\, dt + \sum_{i=1}^n \sigma_i \partial_i f\, dW_i(t) + \frac{1}{2} \sum_{i,j=1}^n a_{i,j} \partial_i \partial_j f\, dt$$

where

$$a_{i,j}(t) = \sum_{k=1}^N \sigma_{i,k}(t)\sigma_{j,k}(t)\,.$$

In matrix notation, the matrix $a = \sigma\sigma^T$, where σ^T is the transpose of the matrix σ.