Stochastic Calculus Notes, Lecture 7

Last modified December 3, 2004

1 The Ito integral with respect to Brownian motion

1.1. Introduction: Stochastic calculus is about systems driven by noise. The Ito calculus is about systems driven by *white noise*, which is the derivative of Brownian motion. To find the response of the system, we integrate the forcing, which leads to the *Ito integral*, of a function against the derivative of Brownian motion.

The Ito integral, like the Riemann integral, has a definition as a certain limit. The fundamental theorem of calculus allows us to evaluate Riemann integrals without returning to its original definition. *Ito's lemma* plays that role for Ito integration. Ito's lemma has an extra term not present in the fundamental theorem that is due to the non smoothness of Brownian motion paths. We will explain the formal rule: $dW^2 = dt$, and its meaning.

In this section, standard one dimensional Brownian motion is $W(t)$ ($W(0) = 0$, $E[\Delta W^2] = \Delta t$). The change in Brownian motion in time dt is formally called $dW(t)$. The independent increments property implies that $dW(t)$ is independent of $dW(t')$ when $t \neq t'$. Therefore, the $dW(t)$ are a model of driving noise impulses acting on a system that are independent from one time to another. We want a rule to add up the cumulative effects of these impulses. In the first instance, this is the integral

$$Y(T) = \int_0^T F(t)dW(t) \; . \tag{1}$$

Our plan is to lay out the principle ideas first then address the mathematical foundations for them later. There will be many points in the beginning paragraphs where we appeal to intuition rather than to mathematical analysis in making a point. To justify this approach, I (mis)quote a snippet of a poem I memorized in grade school: "So you have built castles in the sky. That is where they should be. Now put the foundations under them." (Author unknown by me).

1.2. The Ito integral: Let $\mathcal{F}_t$ be the filtration generated by Brownian motion up to time t, and let $F(t) \in \mathcal{F}_t$ be an adapted stochastic process. Corresponding to the Riemann sum approximation to the Riemann integral we define the following approximations to the Ito integral

$$Y_{\Delta t}(t) = \sum_{t_k < t} F(t_k)\Delta W_k \; , \tag{2}$$

with the usual notions $t_k = k\Delta t$, and $\Delta W_k = W(t_{k+1}) - W(t_k)$. If the limit exists, the Ito integral is

$$Y(t) = \lim_{\Delta t \to 0} Y_{\Delta t}(t) \; . \tag{3}$$

There is some flexibility in this definition, though far less than with the Riemann integral. It is absolutely essential that we use the *forward difference* rather than, say, the backward difference ((*wrong*) $\Delta W_k = W(t_k) - W(t_{k-1})$), so that

$$E\left[F(t_k)\Delta W_k \;\middle|\; \mathcal{F}_{t_k}\right] = 0 \; . \tag{4}$$

Each of the terms in the sum (2) is measurable measurable in $\mathcal{F}_t$, therefore $Y_n(t)$ is also. If we evaluate at the discrete times t_n, $Y_{\Delta t}$ is a martingale:

$$E[Y_{\Delta t}(t_{n+1}) \;\big|\; \mathcal{F}_{t_n} = Y\Delta t(t_n) \; .$$

In the limit $\Delta t \to 0$ this should make $Y(t)$ also a martingale measurable in $\mathcal{F}_t$.

1.3. Famous example: The simplest interesting integral with an F_t that is random is

$$Y(T) = \int_0^T W(t)dW(t) \; .$$

If $W(t)$ were differentiable with derivative $\dot{W}$, we could calculate the limit of (2) using $dW(t) = \dot{W}(t)dt$ as

$$(\textit{wrong}) \qquad \int_0^T W(t)\dot{W}(t)dt = \tfrac{1}{2}\int_0^T \partial_t\left(W(t)^2\right)dt = \tfrac{1}{2}W(t)^2 \; . \quad (\textit{wrong}) \tag{5}$$

But this is not what we get from the definition (2) with actual rough path Brownian motion. Instead we write

$$W(t_k) = \tfrac{1}{2}\left(W(t_{k+1}) + W(t_k)\right) - \tfrac{1}{2}\left(W(t_{k+1}) - W(t_k)\right) \; ,$$

and get

$$\begin{aligned}
Y_{\Delta t}(t_n) &= \sum_{k<n} W(t_k)\left(W(t_{k+1}) - W(t_k)\right) \\
&= \sum_{k<n} \tfrac{1}{2}\left(W(t_{k+1}) + W(t_k)\right)\left(W(t_{k+1}) - W(t_k)\right) \\
&\quad - \sum_{k<n} \tfrac{1}{2}\left(W(t_{k+1}) - W(t_k)\right)\left(W(t_{k+1}) - W(t_k)\right) \\
&= \sum_{k<n} \tfrac{1}{2}\left(W(t_{k+1})^2 - W(t_k)^2\right) - \sum_{k<n} \tfrac{1}{2}\left(W(t_{k+1}) - W(t_k)\right)^2 \; .
\end{aligned}$$

The first on the bottom right is (since $W(0) = 0$)

$$\tfrac{1}{2}W(t_n)^2$$

The second term is a sum of n independent random variables, each with expected value $\Delta t/2$ and variance $\Delta t^2/2$. As a result, the sum is a random variable with mean $n\Delta t/2 = t_n/2$ and variance $n\Delta t^2/2 = t_n \Delta t/2$. This implies that

$$\tfrac{1}{2} \sum_{t_k<T} \left(W(t_{k+1}) - W(t_k)\right)^2 \to T/2 \quad \text{as} \quad \Delta t \to 0 \,. \tag{6}$$

Together, these results give the correct Ito answer

$$\int_0^T W(t) dW(t) = \tfrac{1}{2}\left(W(t)^2 - T\right) \,. \tag{7}$$

The difference between the right answer (7) and the wrong answer (5) is the $T/2$ coming from (6). This is a quantitative consequence of the roughness of Brownian motion paths. If $W(t)$ were a differentiable function of t, that term would have the approximate value

$$\Delta t \int_0^T \left(\frac{dW}{dt}\right)^2 dt \to 0 \quad \text{as} \quad \Delta t \to 0 \,.$$

1.4. Backward differencing, etc: If we use the backward difference $\Delta W_k = W(t_k) - W(t_{k-1})$, then the martingale property (4) does not hold. For example, if $F(t) = W(t)$ as above, then the right side changes from zero to $(W(t_n) - W(t_{n-1})W(t_n)$ (all quantities measurable in $\mathcal{F}_{t_n}$), which has expected value[1] Δt. In fact, if we use the backward difference and follow the argument used to get (7), we get instead $\frac{1}{2}(W(T)^2 + T)$. In addition to the Ito integral there is a *Stratonovich* integral, which is used the central difference $\Delta W_k = \frac{1}{2}(W(t_{k+1}) - W(t_{k-1}))$. The Stratonovich definition makes the stochastic integral act more like a Riemann integral. In particular, the reader can check that the Stratonovich integral of WdW is $\frac{1}{2}W(T)^2$.

1.5. Martingales: The Ito integral is a martingale. It was defined for that purpose. Often one can compute an Ito integral by starting with the ordinary calculus guess (such as $\frac{1}{2}W(T)^2$) and asking what needs to change to make the answer a martingale. In this case, the balancing term $-T/2$ does the trick.

1.6. The Ito differential: Ito's lemma is a formula for the Ito differential, which, in turn, is defined in using the Ito integral. Let $F(t)$ be a stochastic process. We say $dF = a(t)dW(t) + b(t)dt$ (the *Ito differential*) if

$$F(T) - F(0) = \int_0^T a(t) dW(t) + \int_0^T b(t) dt \,. \tag{8}$$

The first integral on the right is an Ito integral and the second is a Riemann integral. Both $a(t)$ and $b(t)$ may be stochastic processes (random functions of

[1] $E[(W(t_n) - W(t_{n-1}))W(t_{n-1})] = 0$, so $E[(W(t_n) - W(t_{n-1}))W(t_n)] = E[(W(t_n) - W(t_{n-1}))(W(t_n) - W(t_{n-1}))] = \Delta t$

time). For example, the Ito differential of $W(t)^2$ is

$$dW(t)^2 = 2W(t)dW(t) + dt \; ,$$

which we verify by checking that

$$W(T)^2 = 2\int_0^T W(t)dW(t) + \int_0^T dt \; .$$

This is a restatement of (7).

1.7. Ito's lemma: The simplest version of *Ito's lemma* involves a function $f(w,t)$. The "lemma" is the formula (which must have been stated as a lemma in one of his papers):

$$df(W(t),t) = \partial_w f(W(t),t)dW(t) + \tfrac{1}{2}\partial_w^2 f(W(t),t)dt + \partial_t f(W(t),t)dt \; . \quad (9)$$

According to the definition of the Ito differential, this means that

$$f(W(T),T) - f(W(0),0) \quad (10)$$
$$= \int_0^T \partial_w f(W(t),t)dW(t) + \int_0^T \left(\partial_w^2 f(W(t),t) + \partial_t f(W(t),t)\right) dt \quad (11)$$

1.8. Using Ito's lemma to evaluate an Ito integral: Like the fundamental theorem of calculus, Ito's lemma can be used to evaluate integrals. For example, consider

$$Y(T) = \int_0^T W(t)^2 dW(t) \; .$$

A naive guess might be $\frac{1}{3}W(T)^3$, which would be the answer for a differentiable function. To check this, we calculate (using (9), $\partial_w \frac{1}{3}w^3 = w^2$, and $\frac{1}{2}\partial_w^2 \frac{1}{3}w^3 = w$)

$$d\tfrac{1}{3}W(t)^3 = W^2(t)dW(t) + W(t)dt \; .$$

This implies that

$$\tfrac{1}{3}W(t)^3 = \int_0^T d\tfrac{1}{3}W(t)^3 = \int_0^T W(t)^2 dW(t) + \int_0^T W(t)dt \; ,$$

which in turn gives

$$\int_0^T W(t)^2 dW(t) = \tfrac{1}{3}W(t)^3 - \int_0^T W(t)dt \; .$$

This seems to be the end. There is no way to "integrate" $Z(T) = \int_0^T W(t)dt$ to get a function of $W(T)$ alone. This is to say that $Z(T)$ is not measurable in $\mathcal{G}_T$, the algebra generated by $W(T)$ alone. In fact, $Z(T)$ depends equally on all

$W(t)$ values for $0 \leq t \leq T$. A more technical version of this remark is coming after the discussion of the Brownian bridge.

1.9. To tell a martingale: Suppose $F(t)$ is an adapted stochastic process with $dF(t) = a(t)dW(t) + b(t)dt$. Then F is a martingale if and only if $b(t) = 0$. We call $a(t)dW(t)$ the *martingale part* and $b(t)dt$ *drift term*. If $b(t)$ is at all continuous, then it can be identified through (because $E[\int a(s)dW(s) \mid \mathcal{F}_t] = 0$)

$$\begin{aligned} E\left[F(t+\Delta t) - F(t) \mid \mathcal{F}_t\right] &= E\left[\int_t^{t+\Delta t} b(s)ds \mid \mathcal{F}_t\right] \\ &= b(t)\Delta t + o(\Delta t) \ . \end{aligned} \tag{12}$$

We give one and a half of the two parts of the proof of this theorem. If $b = 0$ for all t (and all, or almost all $\omega \in \Omega$), then $F(T)$ is an Ito integral and hence a martingale. If $b(t)$ is a continuous function of t, then we may find a t^* and $\epsilon > 0$ and $\delta > 0$ so that, say, $b(t) > \delta > 0$ when $|t - t^*| < \epsilon$. Then $E[F(t^* + \epsilon) - F(t^* - \epsilon)] > 2\delta\epsilon > 0$, so F is not a martingale[2].

1.10. Deriving a backward equation: Ito's lemma gives a quick derivations of backward equations. For example, take

$$f(W(t), t) = E\left[V(W(T)) \mid \mathcal{F}_t\right] \ .$$

The tower property tells us that $F(t) = F(W(t), t)$ is a martingale. But Ito's lemma, together with the previous paragraph, implies that $F(W(t), t)$ is a martingale of and only if $\partial_t F + \frac{1}{2} = 0$, which is the backward equation for this case. In fact, the proof of Ito's lemma (below) is much like the proof of this backward equation.

1.11. A backward equation with drift: The derivation of the backward equation for

$$f(w, t) = E_{w,t}\left[\int_t^T V(W(s), s)ds\right]$$

uses the above, plus (12). Again using

$$F(t) = E\left[\int_t^T V(W(s), s)ds \mid \mathcal{F}_t\right] \ ,$$

with $F(t) = f(W(t), t)$, we calculate

$$\begin{aligned} E\left[F(t+\Delta t) - F(t) \mid \mathcal{F}_t\right] &= -E\left[\int_t^{t+\Delta t} V(W(s), s)ds \mid \mathcal{F}_t\right] \\ &= -V(W(t), t)\Delta t + o(\Delta t) \ . \end{aligned}$$

[2]This is a somewhat incorrect version of the proof because ϵ, δ, and t^* probably are random. There is a real proof something like this.

This says that $dF(t) = a(t)dW(t) + b(t)dt$ where

$$b(t) = -V(W(t), t) \; .$$

But also, $b(t) = \partial_t f + \frac{1}{2}\partial_w^2 f$. Equating these gives the backward equation from Lecture 6:

$$\partial_t f + \tfrac{1}{2}\partial_w^2 f + V(w,t) = 0 \; .$$

1.12. Proof of Ito's lemma: We want to show that

$$\begin{aligned} f(W(T),T) - f(W(0),0) &= \int_0^T f_w(W(t),t)dW(t) + \int_0^T f_t(W(t),t)dt \\ &\quad + \tfrac{1}{2}\int_0^T f_{ww}(W(t),t)dt \; . \end{aligned} \tag{13}$$

Define $\Delta t = T/n$, $t_k = k\Delta t$, $W_k = W(t_k)$, $\Delta W_k = W(t_{k+1}) - W(t_k)$, and $f_k = f(W_k, t_k)$, and write

$$f_n - f_0 = \sum_{k=0}^{n-1} \left(f_{k+1} - f_k \right) \; . \tag{14}$$

Taylor series expansion of the terms on the right of (14) will produce terms that converge to the three integrals on the right of (13) plus error terms that converge to zero. In our pre-Ito derivations of backward equations, we used the relation $E[(\Delta W)^2] = \Delta t$. Here we argue that with many independent ΔW_k, we may replace $(\Delta W_k)^2$ with Δt (its mean value).

The Taylor series expansion is

$$f_{k+1} - f_k = \partial_w f_k \Delta W_k + \tfrac{1}{2}\partial_w^2 f_k \left(\Delta W\right)^2 + \partial_t f_k \Delta t + R_k \; , \tag{15}$$

where $\partial_w f_k$ means $\partial_w f(W(t_k), t_k)$, etc. The remainder has the bound[3]

$$|R_k| \leq C\left(\Delta t^2 + \Delta t\left|\Delta W_k\right| + \left|\Delta W_k^3\right|\right) \; .$$

Finally, we separate the mean value of ΔW_k^2 from the deviation from the mean:

$$\frac{1}{2}\partial_w^2 f_k \Delta W_k^2 = \frac{1}{2}\partial_w^2 f_k \Delta t + \frac{1}{2}\partial_w^2 f_k (\Delta W_k^2 - \Delta t) \; .$$

The individual summands on the right side all have order of magnitude Δt. However, the mean zero terms (the second sum) add up to much less than the first sum, as we will see. With this, (14) takes the form

$$\begin{aligned} f_n - f_0 &= \sum_{k=0}^{n-1} \partial_w f_k \Delta W_k + \sum_{k=0}^{n-1} \partial_t f_k \Delta t + \tfrac{1}{2}\sum_{k=0}^{n-1} \partial_w^2 f_k \Delta t \\ &\quad + \tfrac{1}{2}\sum_{k=0}^{n-1} \partial_w^2 f_k \left(\Delta W^2 - \Delta t\right) + \sum_{k=0}^{n-1} R_k \; . \end{aligned} \tag{16}$$

[3] We assume that $f(w,t)$ is thrice differentiable with bounded third derivatives. The error in a finite Taylor approximation is bounded by the sized of the largest terms not used. Here, that is Δt^2 (for omitted term $\partial_t^2 f$), $\Delta t(\Delta W)^2$ (for $\partial_t \partial_w$), and ΔW^3 (for ∂_w^3).

The first three sums on the right converge respectively to the corresponding integrals on the right side of (13). A technical digression will show that the last two converge to zero as $n \to \infty$ in a suitable way.

1.13. Like Borel Cantelli: As much as the formulas, the proofs in stochastic calculus rely on calculating expected values of things. Here, S_m is a sequence of random numbers and we want to show that $S_m \to 0$ as $m \to \infty$ (almost surely). We use two observations. First, if s_m is a sequence of numbers with $\sum_{m=1}^{\infty} |s_m| < \infty$, then $s_m \to 0$ as $m \to \infty$. Second, if $B > 0$ is a random variable with $E[B] < \infty$, then $B < \infty$ almost surely (if the event $\{B = \infty\}$ has positive probability, then $E[B] = \infty$). We take $B = \sum_{m=1}^{\infty} |S_m|$. If $B < \infty$ then $\sum_{m=1}^{\infty} |S_m| < \infty$ so $S_m \to 0$ as $m \to \infty$. What this shows is:

$$\left(\sum_{m=1}^{\infty} E\left[|S_m|\right] < \infty \right) \Longrightarrow \left(S_m \to 0 \text{ as } m \to \infty \text{ (a.s.)} \right) \tag{17}$$

This observarion is a variant of the *Borel Cantelli lemma*, which often is used in such arguments.

1.14. One of the error terms: To apply the Borel Cantelli lemma we must find bounds for the error terms, bounds whose sum is finite. We start with the last error term in (16). Choose $n = 2^m$ and define $S_m = \sum_{k=0}^{n-1} R_k$, with

$$|R_k| \leq C\big(\Delta t^2 + \Delta t\,|\Delta W_k| + |\Delta W_k|^3\big) .$$

Since $E[|\Delta W_k|] \leq C\sqrt{\Delta t}$ and $E[|\Delta W_k|^3] \leq C\Delta t^{3/2}$ (you do the math – the integrals), this gives (with $n\Delta t = T$)

$$\begin{aligned} E\left[|S_m|\right] &\leq Cn\left(\Delta t^2 + \Delta t^{3/2}\right) \\ &\leq CT\sqrt{\Delta t} \,. \end{aligned}$$

Expressed in terms of m, we have $\Delta t = T/2^m$ and $\sqrt{\Delta t} = \sqrt{T}2^{-m/2} = \sqrt{T}\left(\sqrt{2}\right)^{-m}$. Therefore $E\left[|S_m|\right] \leq C(T)\left(\sqrt{2}\right)^{-m}$. Now, if z is any number greater than one, then $\sum_{m=1}^{\infty} z^{-m} = 1/(1+1/z)) < \infty$. This implies that $\sum_{m=1}^{\infty} E\left[|S_m|\right] < \infty$ and (using Borel Cantelli) that $S_m \to 0$ as $m \to \infty$ (almost surely).

This argument would not have worked this way had we taken $n = m$ instead of $n = 2^m$. The error bounds of order $1/\sqrt{n}$ would not have had a finite sum. If both error terms in the bottom line of (16) go to zero as $m \to \infty$ with $n = 2^m$, this will prove Ito's lemma. We will return to this point when we discuss the difference between *almost sure convergence*, which we are using here, and *convergence in probability*, which we are not.

1.15. The other sum: The other error sum in (16) is small not because of the smallness of its terms, but because of *cancellation*. The positive and negative terms roughly balance, leaving a sum smaller than the sizes of the terms would suggest. This cancellation is of the same sort appearing in the central limit theorem, where $\sum_{k=0}^{n-1} X_k = U_n$ is of order $\sqrt{n}$ rather than n when the X_k are i.i.d. with finite variance. In fact, using a trick we used before we show that U_n^2 is of order n rather than n^2:

$$E\left[U_n^2\right] = \sum_{jk} E\left[X_j X_k\right] = nE\left[X_k^2\right] = cn \; .$$

Our sum is

$$U_n = \sum \tfrac{1}{2}\partial_w^2 f(W_k,t_k)\left(\Delta W_k^2 - \Delta t_k\right) \; .$$

The above argument applies, though the terms are not independent. Suppose $j \neq k$ and, say, $k > j$. The cross term involving ΔW_j and ΔW_k still vanishes because

$$E\left[\Delta W_k - \Delta t \mid \mathcal{F}_{t_k}\right] = 0 \; ,$$

and the rest is in $\mathcal{F}_{t_k}$. Also (as we have used before)

$$E\left[(\Delta W_k - \Delta t)^2 \mid \mathcal{F}_{t_k}\right] = 2\Delta t^2 \; .$$

Therefore

$$E\left[U_n^2\right] = \tfrac{1}{4}\sum_{k=0}^{n-1}\left(\partial_w^2 f(W_k,t_k)\right)^2 \Delta t^2 \leq C(T)\Delta t \; .$$

As before, we take $n = 2^m$ and sum to find that $U_{2^m}^2 \to 0$ as $m \to \infty$, which of course implies that $U_{2^m} \to 0$ as $m \to \infty$ (almost surely).

1.16. Convergence of Ito sums: Choose Δt and define $t_k = k\Delta t$ and $W_k = W(t_k)$. To approximate the Ito integral

$$Y(T) = \int_0^T F(t)dW(t) \; ,$$

we have the the Ito sums

$$Y_m(T) = \sum_{t_k<T} F(t_k)\left(W_{k+1} - W_k\right) \; , \tag{18}$$

where $\Delta t = 2^{-m}$. In proving convergence of Riemann sums to the Riemann integral, we assume that the integrand is continuous. Here, we will prove that $\lim_{m\to\infty} Y_m(T)$ exists under the hypothesis

$$E\Big[\left(F(t+\Delta t) - F(t)\right)^2\Big] \leq C\Delta t \; . \tag{19}$$

This is natural in that it represents the smoothness of Brownian motion paths. We will discuss what can be done for integrands more rough than (19).

The trick is to compare Y_m with Y_{m+1}, which is to compare the Δt approximation to the $\Delta t/2$ approximation. For that purpose, define $t_{k+1/2} = (k+\frac{1}{2})\Delta t$,

$W_{k+1/2} = W(t_{k+1/2})$, etc. The t_k term in the Y_m sum corresponds to the time interval (t_k, t_{k+1}). The Y_{m+1} sum divides this interval into two subintervals of length $\Delta t/2$. Therefore, for each term in the Y_m sum there are two corresponding terms in the Y_{m+1} sum (assuming T is a multiple of Δt), and:

$$\begin{aligned} Y_{m+1}(T) - Y_m(T) &= \sum_{t_k<T} \big[F(t_k)(W_{k+1/2} - W_k) + F(t_{k+1/2})(W_{k+1} - W_{k+1/2}) \\ &\quad - F(t_k)(W_{k+1} - W_k)\big] \\ &= \sum_{t_k<T} (W_{k+1} - W_{k+1/2})(F(t_{k+1/2}) - F(t_k)) \\ &= \sum_{t_k<T} R_k \ , \end{aligned}$$

where

$$R_k = (W_{k+1} - W_{k+1/2})(F(t_{k+1/2}) - F(t_k)) \ .$$

We compute $E[(Y_{m+1}(T) - Y_m(T))^2] = \sum_{jk} E[R_j R_k]$. As before,[4] $E[R_j R_k] = 0$ unless $j = k$. Also, the independent increments property and (19) imply that[5]

$$\begin{aligned} E[R_k^2] &= E\left[\left(W_{k+1} - W_{k+1/2}\right)^2\right] \cdot E\left[\left(F(t_{k+1/1}) - F(t_k)\right)^2\right] \\ &\leq \frac{\Delta t}{2} \cdot C\frac{\Delta t}{2} = C\Delta t^2 \ . \end{aligned}$$

This gives

$$E\big[(Y_{m+1}(T) - Y_m(T))^2\big] \leq C2^{-m} \ . \tag{20}$$

The convergence of the Ito sums follows from (**??**) using our Borel Cantelli type lemma. Let $S_m = Y_{m+1} - Y_m$. From (20), we have[6] $E\,|S_m|] \leq C\sqrt{2}^{-m}$. Thus

$$\lim_{m\to\infty} Y_m(T) = Y_1(T) + \sum_{m\geq 1} Y_{m+1}(T) - Y_m(T)$$

exists and is finite. This shows that the limit defining the Ito integral exists, at least in the case of an integrand that satisfies (19), which includes most of the cases we use.

1.17. Ito isometry formula: This is the formula

$$E\left[\left(\int_{T_1}^{T_2} a(t)dW(t)\right)^2\right] = \int_{T_1}^{T_2} E[a(t)^2]dt \ . \tag{21}$$

[4]If $j > k$ then $E[W_{j+1} - W_{j+1/2} \mid \mathcal{F}_{t_{j+1/2}}] = 0$, so $E[R_j R_k \mid \mathcal{F}_{t_{j+1/2}}] = 0$, and $E[R_j R_k] = 0$

[5]Mathematicians often use the same letter C to represent different constants in the same formula. For example, $C\Delta t + C^2\Delta t \leq C\Delta t$ really means: "Let $C = C_1 + C_2^2$, if $u \leq C_1\Delta t$ and $v \leq C_2\sqrt{\Delta t}$, then $u + v^2 \leq C\Delta t$. Instead, we don't bother to distinguish between the various constants.

[6]The Cauchy Schwartz inequality gives $E[|S_m|] = E[|S_m| \cdot 1] \leq (E[S_m^2]E[1^2])^{1/2} = E[S_m^2]^{1/2}$.

The derivation uses what we just have done. We approximate the Ito integral by the sum

$$\sum_{T_1 \leq t_k < T_2} a(t_k)\Delta W_k \ .$$

Because the different ΔW_k are independent, and because of the independent increments property, the expected square of this is

$$\sum_{T_1 \leq t_k < T_2} a(t_k)^2 \Delta t \ .$$

The formula (21) follows from this.

An application of this is to understand the roughness of $Y(T) = \int_0^T a(t)dW(t)$. If $E[a(t)^2] < C$ for all $t \leq T$, then $E[(Y(T_2) - Y(T_1))^2] \leq C\Delta t$. This is the same roughness as Brownian motion itself.

1.18. White noise: *White noise* is a *generalized function*,[7] $\xi(t)$, which is thought of as homogeneous and Gaussian with $\xi(t_1)$ independent of $\xi(t_2)$ for $t_1 \neq t_2$. More precisely, if $t_0 < t_1 < \cdots < t_n$ and $Y_k = \int_{t_k}^{t_{k+1}} \xi(t)dt$, then the Y_k are independent and normal with zero mean and $\mathrm{var}(Y_k) = t_{k+1} - t_k$. You can convince yourself that $\xi(t)$ is not a true function by showing that it would have to have $\int_a^b \xi(t)^2 dt = \infty$ for any $a < b$. Brownian motion can be thought of as the motion of a particle pushed by white noise, i.e. $W(t) = \int_0^t \xi(s)ds$. The Y_k defined above then are the increments of Brownian and have the appropriate statistical properties (independent, mean zero, normal, variance $t_{k+1} - t_k$).

These properties may be summarized by saying that $\xi(t)$ has mean zero and

$$\mathrm{cov}(\xi(t), \xi(t') = E[\xi(t)\xi(t') = \delta(t - t') \ . \tag{22}$$

For example, if $f(t)$ and $g(t)$ are deterministic functions and $Y_f = \int f(t)d\xi(t)$ and $Y_g = \int g(t)d\xi(t)$, then, (22) implies that

$$\begin{aligned} E[Y_f Y_g] &= \int_t \int_{t'} f(t)g(t')E[\xi(t)\xi(t')]dtdt' \\ &= \int_t \int_{t'} f(t)g(t')\delta(t - t')dtdt' \\ &= \int_t f(t)g(t)dt \end{aligned}$$

It is tempting to take $dW(t) = \xi(t)dt$ in the Ito integral and use (22) to derive the Ito isometry formula. However, this must by done carefully because the existence of the Ito integral, and the isometry formula, depend on the causality structure that makes $dW(t)$ independent of $a(t)$.

[7] A generalized function is not an actual function, but has properties defined as though it were an actual function through integration. The δ function for example, is defined by the formula $\int f(t)\delta(t)dt = f(0)$. No actual function can do this. Generalized functions also are called *distributions*.

2 Stochastic Differential Equations

2.1. Introduction: The theory of stochastic differential equations (SDE) is a framework for expressing dynamical models that include both random and non random forces. The theory is based on the Ito integral. Like the Ito integral, approximations based on finite differences that do not respect the martingale structure of the equation can converge to different answers. Solutions to Ito SDEs are Markov processes in that the future depends on the past only through the present. For this reason, the solutions can be studied using backward and forward equations, which turn out to be linear *parabolic* partial differential equations of *diffusion type*.

2.2. A Stochastic Differential Equation: An Ito stochastic differential equation takes the form

$$dX(t) = a(X(t),t)dt + \sigma(X(t),t)dW(t) \ . \tag{23}$$

A solution is an adapted process that satisfies (23) in the sense that

$$X(T) - X(0) = \int_0^T a(X(t),t)dt + \int_0^T \sigma(X(t),t)dW(t) \ , \tag{24}$$

where the first integral on the right is a Riemann integral and the second is an Ito integral. We often specify *initial conditions* $X(0) \sim u_0(x)$, where $u_0(x)$ is the given probability density for $X(0)$. Specifying $X(0) = x_0$ is the same as saying $u_0(x) = \delta(x - x_0)$. As in the general Ito differential, $a(X(t),t)dt$ is the *drift term*, and $\sigma(X(t),t)dW(t)$ is the *martingale term*. We often call $\sigma(x,t)$ the *volatility*. However, this is a different use of the letter σ from Black Scholes, where the martingale term is $x\sigma$ for a constant σ (also called volatility).

2.3. Geometric Brownian motion: The SDE

$$dX(t) = \mu X(t)dt + \sigma X(t)dW(t) \ , \tag{25}$$

with initial data $X(0) = 1$, defines *geometric Brownian motion*. In the general formulation above, (25) has drift coefficient $a(x,t) = \mu x$, and volatility $\sigma(x,t) = \sigma x$ (with the conflict of terminology noted above). If $W(t)$ were a differentiable function of t, the solution would be

$$\text{(wrong)} \qquad X(t) = e^{\mu t + \sigma W(t)} \ . \tag{26}$$

To check this, define the function $x(w,t) = e^{\mu t | \sigma w}$ with $\partial_w x = x_w = \sigma x$, $x_t = \mu x$ and $x_{ww} = \sigma^2 x$. so that the Ito differential of the trial function (26) is

$$dX(W(t),t) = \mu X dt + \sigma X dW(t) + \frac{\sigma^2}{2} X dt \ .$$

We can remove the unwanted final term by multiplying by $e^{-\sigma^2 t/2}$, which suggests that the formula

$$X(t) = e^{\mu t - \sigma^2 t/2 + \sigma W(t)} \tag{27}$$

satisfies (25). A quick Ito differentiation verifies that it does.

2.4. Properties of geometric Brownian motion: Let us focus on the simple case $\mu = 0$ $\sigma = 1$, so that

$$dX(t) = X(t)dW(t) \; . \tag{28}$$

The solution, with initial data $X(0) = 1$, is the simple geometric Brownian motion

$$X(t) = \exp(W(t) - t/2) \; . \tag{29}$$

We discuss (29) in relation to the martingale property ($X(t)$ is a martingale because the drift term in (23) is zero in (28)). A simple calculation based on

$$X(t+t') = \exp\bigl(W(t) - t/2 + (W(t') - W(t) - (t'-t)/2)\bigr)$$

and integrals of Gaussians shows that $E[X(t+t' \mid \mathcal{F}_t] = X(t)$.

However, $W(t)$ has the order of magnitude $\sqrt{t}$. For large t, this means that the exponent in (29) is roughly equal to $-t/2$, which suggests that

$$X(t) = \exp(W(t) - t/2) \approx e^{-t/2} \to 0 \quad \text{as} \quad t \to \infty \;\; \text{(a.s.)}$$

Therefore, the expected value $E[X(t)] = 1$, for large t, is not produced by typical paths, but by very exceptional ones. To be quantitative, there is an exponentially small probability that $X(t)$ is as large as its expected value:

$$P(X(t) \geq 1) < e^{-t/4} \quad \text{for large } t.$$

2.5. Dominated convergence theorem: The *dominated convergence theorem* is about expected values of limits of random variables. Suppose $X(t,\omega)$ is a family of random variables and that $\lim_{t\to\infty} X(t,\omega) = Y(\omega)$ for almost every ω. The random variable $U(\omega)$ *dominates* the $X(t,\omega)$ if $|X(t,\omega)| \leq U(\omega)$ for almost every ω and for every $t > 0$. We often write this simply as $X(t) \to Y$ as $t \to \infty$ a.s., and $|X(t)| \leq U$ a.s. The theorem states that if $E[U] < \infty$ then $E[X(t)] \to E[Y]$ as $t \to \infty$. It is fairly easy to prove the theorem from the definition of abstract integration. The simplicity of the theorem is one of the ways abstract integration is simpler than Riemann integration.

The reason for mentioning this theorem here is that geometric Brownian motion (29) is an example showing what can go wrong without a dominating function. Although $X(t) \to 0$ as $t \to \infty$ a.s., the expected value of $X(t)$ does not go to zero, as it would do if the conditions of the dominated convergence theorem were met. The reader is invited to study the maximal function, which is the random variable $M = \max_{t>0}(W(t) - t/2)$, in enough detail to show that $E[e^M] = \infty$.

2.6. Strong and weak solutions: A *strong solution* is an adapted function $X(W,t)$, where the Brownian motion path W again plays the role of the abstract random variable, ω. As in the discrete case, $X(t)$ (i.e. $X(W,t)$) being measurable in $\mathcal{F}_t$ means that $X(t)$ is a function of the values of $W(s)$ for $0 \leq s \leq t$. The two examples we have, geometric Brownian motion (27), and the Ornstein Uhlenbeck process[8]

$$X(t) = \sigma \int_0^t e^{-\gamma(t-s)} dW(s) \; , \tag{30}$$

both have this property. Note that (27) depends only on $W(t)$, while (30) depends on the whole pate up to time t.

A *weak solution* is a stochastic process, $X(t)$, defined perhaps on a different probability space and filtration $(\Omega, \mathcal{F}_t)$ that has the statistical properties called for by (23). These are (using $\Delta X = X(t+\Delta t) - X(t)$) roughly[9]

$$E[\Delta X \mid \mathcal{F}_t] = a(X(t),t)\Delta t + o(\Delta t) \; , \tag{31}$$

and

$$E[\Delta X^2 \mid \mathcal{F}_t] = \sigma^2(X(t),t)\Delta t + o(\Delta t) \; . \tag{32}$$

We will see that a strong solution satisfies (31) and (32), so a strong solution is a weak solution. It makes no sense to ask whether a weak solution is a strong solution since we have no information on how, or even whether, the weak solution depends on W.

The formulas (31) and (32) are helpful in deriving SDE descriptions of physical or financial systems. We calculate the left sides to identify the $a(x,t)$ and $\sigma(x,t)$ in (23). Brownian motion paths and Ito integration are merely a tool for constructing the desired process $X(t)$. We saw in the example of geomertic Brownian motion that expressing the solution in terms of $W(t)$ can be very convenient for understanding its properties. For example, it is not particularly easy to show that $X(t) \to 0$ as $t \to \infty$ from (31) and (32) with $a = \mu X$ and[10] $\sigma(x,t) = \sigma x$.

2.7. Strong is weak: We just verify that the strong solution to (23) that satisfies (24) also satisfies the weak form requirements (31) and (32). This is an important motivation for using the Ito definition of dW rather than, say, the Stratonovich definition.

A slightly more general fact is simpler to explain. Define R and I by

$$R = \int_t^{t+\Delta t} a(t)dt \; , \quad I = \int_t^{t+\Delta t} \sigma(t)dW(t) \; ,$$

[8]This process satisfies the SDE $dX = -\gamma X dt + \sigma dW$, with $X(0) = 0$.

[9]The *little o* notation $f(t) = g(t) + o(t)$ informally means that the difference between f and g is a mathematicians' order of magnitude smaller than t for small t. Formally, it means that $(f(t) - g(t))/t \to 0$ as $t \to 0$.

[10]This conflict of notation is common in discussing geometric Brownian motion. On the left is the coefficient of $dW(t)$. On the right is the *financial* volatility coefficient.

where $a(t)$ and $\sigma(t)$ are continuous adapted stochastic processes. We want to see that

$$E\left[R+I \;\middle|\; \mathcal{F}_t\right] = a(t) + o(\Delta t) \; , \tag{33}$$

and

$$E\left[(R+I)^2 \;\middle|\; \mathcal{F}_t\right] = \sigma^2(t) + o(\Delta t) \; . \tag{34}$$

We may leave I out of (33) because $E[I] = 0$ always. We may leave R out of (34) because $|I| >> |R|$. (If a is bounded then $R = O(\Delta t)$ so $E[R^2 \mid \mathcal{F}_t] = O(\Delta t^2)$. The Ito isometry formula suggests that $E[I^2 \mid \mathcal{F}_t] = O(\Delta t)$. Cauchy Schwartz then gives $E[RI \mid \mathcal{F}_t] = O(\Delta t^{3/2})$. Altogether, $E[(R+I)^2 \mid \mathcal{F}_t] = E[I^2 \mid \mathcal{F}_t] + O(\Delta t^{3/2})$.)

To verify (33) without I, we assume that $a(t)$ is a continuous function of t in the sense that for $s > t$,

$$E\left[a(s) - a(t) \;\middle|\; \mathcal{F}_t\right] \to 0 \quad \text{as} \quad s \to t \; .$$

This implies that

$$\frac{1}{\Delta t}\int_t^{t+\Delta t} E\left[a(s) - a(t) \mid \mathcal{F}_t\right] \to 0 \;\text{ as } \Delta t \to 0,$$

so that

$$\begin{aligned} E\left[R \;\middle|\; \mathcal{F}_t\right] &= \int_t^{t+\Delta t} E\left[a(s) \mid \mathcal{F}_t\right] \\ &= \int_t^{t+\Delta t} E\left[a(t) \mid \mathcal{F}_t\right] + \int_t^{t+\Delta t} E\left[a(s) - a(t) \mid \mathcal{F}_t\right] \\ &= \Delta t a(t) + o(\Delta t) \; . \end{aligned}$$

This verifies (33). The Ito isometry formula gives

$$E\left[I^2 \;\middle|\; \mathcal{F}_t\right] = \int_t^{t+\Delta t} \sigma(s)^2 ds \; ,$$

so (34) follows in the same way.

2.8. Markov diffusions: Roughly speaking, [11] a diffusion process is a continuous stochastic process that satisfies (31) and (32). If the process is Markov, the a of (31) and the σ^2 of (32) must be functions of $X(t)$ and t. If $a(x,t)$ and $\sigma(x,t)$ are Lipschitz ($|a(x,t) - a(y,t)| \leq C|x-y|$, etc.) functions of x and t, then it is possible to find it is possible to express $X(t)$ as a strong solution of an Ito SDE (23).

This is the way equations (23) are often derived in practice. We start off wanting to model a process with an SDE. It could be a random walk on a lattice with the lattice size converging to zero or some other process that we hope will

[11] More detailed treatment are the books by Steele, Chung and Williams, Karatsas and Shreve, and Oksendal.

have a limit as a diffusion. The main step in proving the limit exists is *tightness*, which we hint at a lecture to follow. We identify a and σ by calculations. Then we use the representation theorem to say that the process may be represented as the strong solution to (23).

2.9. Backward equation: The simplest backward equation is the PDE satisfied by $f(x,t) = E_{x,t}[V(X(T))]$. We derive it using the weak form conditions (31) and(32) and the tower property. As with Brownian motion, the tower property gives

$$f(x,t) = E_{x,t}[V(X(T))] = E_{x,t}[F(t+\Delta t)] \ ,$$

where $F(s) = E[V(X(T)) \mid \mathcal{F}_s]$. The Markov property implies that $F(s)$ is a function of $X(s)$ alone, so $F(s) = f(X(s), s)$. This gives

$$f(x,t) = E_{x,t}\left[f(X(t+\Delta t), t+\Delta t)\right] \ . \tag{35}$$

If we assume that f is a smooth function of x and t, we may expand in Taylor series, keeping only terms that contribute $O(\Delta t)$ or more.[12] We use $\Delta X = X(t+\Delta t) - x$ and write f for $f(x,t)$, f_t for $f_t(x,t)$, etc.

$$f(X(t+\Delta t), t+\Delta t) = f + f_t \Delta t + f_x \Delta X + \tfrac{1}{2} f_{xx} \Delta X^2 + \textit{smaller terms.}$$

Therefore (31) and (32) give:

$$\begin{aligned}
f(x,t) &= E_{x,t}\left[f(X(t+\Delta t), t+\Delta t)\right] \\
&= f(x,t) f_t \Delta t + f_x E_{x,t}[\Delta X] + \tfrac{1}{2} f_{xx} E_{x,t}[\Delta X^2] + o(\Delta t) \\
&= f(x,t) + f_t \Delta t + f_x a(x,t) \Delta t + \tfrac{1}{2} f_{xx} \sigma^2(x,t) \Delta t + o(\Delta t) \ .
\end{aligned}$$

We now just cancel the $f(x,t)$ from both sides, let $\Delta t \to 0$ and drop the $o(\Delta t)$ terms to get the backward equation

$$\partial_t f(x,t) + a(x,t) \partial_x f(x,t) + \frac{\sigma^2(x,t)}{2} \partial_x^2 f(x,t) = 0 \ . \tag{36}$$

2.10. Forward equation: The forward equation follows from the backward equation by duality. Let $u(x,t)$ be the probability density for $X(t)$. Since $f(x,t) = E_{x,t}[V(X(T))]$, we may write

$$E[V(X(T))] = \int_{-\infty}^{\infty} u(x,t) f(x,t) dx \ ,$$

which is independent of t. Differentiating with respect to t and using the backward equation (36) for f_t, we get

$$\begin{aligned}
0 &= \int u(x,t) f_t(x,t) dx + \int u_t(x,t) f(x,t) dx \\
&= -\int u(x,t) a(x,t) \partial_x f(x,t) - \tfrac{1}{2} \int u(x,t) \sigma^2(x,t) \partial_x^2 f(x,t) + \int u_t(x,t) f(x,t) \ .
\end{aligned}$$

[12] The homework has more on the terms left out.

We integrate by parts to put the x derivatives on u. We may ignore boundary terms if u decaus fast enough as $|x| \to \infty$ and if f does not grow too fast. The result is

$$\int \left(\partial_x \left(a(x,t)u(x,t) \right) - \tfrac{1}{2}\partial_x^2 \left(\sigma^2(x,t)u(x,t) \right) + \partial_t u(x,t) \right) f(x,t)dx = 0 \ .$$

Since this should be true for every function $f(x,t)$, the integrand must vanish identically, which implies that

$$\partial_t u(x,t) = -\partial_x \left(a(x,t)u(x,t) \right) + \tfrac{1}{2}\partial_x^2 \left(\sigma^2(x,t)u(x,t) \right) \ . \tag{37}$$

This is the forward equation for the Markov process that satisfies (31) and (32).

2.11. Transition probabilities: The transition probability density is the probability density for $X(s)$ given that $X(t) = y$ and $s > t$. We write it as $G(y,s,t,s)$, the probabiltiy density to go from y to x as time goes from t to s. If the drift and diffusion coefficients do not depend on t, then G is a function of $t - s$. Because G is a probability density in the x and s variables, it satisfies the forward equation

$$\partial_s G(y,x,t,s) = -\partial_x \left(a(x,s)G(y,x,t,s) \right) + \tfrac{1}{2}\partial_x^2 \left(\sigma^2(x,s)G(y,x,t,s) \right) \ . \tag{38}$$

In this equation, t and y are merely parameters, but s may not be smaller than t. The initial condition that represents the requirement that $X(t) = y$ is

$$G(y,x,t,t) = \delta(x-y) \ . \tag{39}$$

The transition density is the *Green's function* for the forward equation, which means that the general solution may be written in terms of G as

$$u(x,s) = \int_{-\infty}^{\infty} u(y,t)G(y,x,t,s)dy \ . \tag{40}$$

This formula is a continuous time version of the law of total probability: the probability density to be at x at time s is the sum (integral) of the probability density to be at x at time s conditional on being at y at time t (which is $G(y,x,t,s)$) multiplied by the probability density to be at y at time s (which is $u(y,t)$).

2.12. Green's function for the backward equation: We can also express the solution of the backward equation in terms the transition probabilities G. For $s > t$,

$$f(y,t) = E_{y,t}\left[f(X(s),s) \right] \ ,$$

which is an expression of the tower property. The expected value on the right may be evaluated using the transition probability density for $X(s)$. The result is

$$f(y,t) = \int_{-\infty}^{\infty} G(y,x,t,s)f(x,s)dx \ . \tag{41}$$

For this to hold, G must satisfy the backward equation as a function of y and t (which were parameters in (38). To show this, we apply the backward equation "operator" (see below for terminology) $\partial_t + a(y,t)\partial_y + \frac{1}{2}\sigma^2(y,t)\partial_y^2$ to both sides. The left side gives zero because f satisfies the backward equation. Therefore we find that

$$0 = \int \left(\partial_t + a(y,t)\partial_y + \tfrac{1}{2}\sigma^2(y,t)\partial_y^2 \right) G(y,x,t,s) f(x,s) dx$$

for any $f(x,s)$. Therefore, we conclude that

$$\partial_t G(y,x,t,s) + a(y,t)\partial_y G(y,x,t,s) + \tfrac{1}{2}\sigma^2(y,t)\partial_y^2 G(y,x,t,s) = 0 \; . \tag{42}$$

Here x and s are parameters. The final condition for (42) is the same as (39). The equality $s = t$ represents the initial time for s and the final time for t because G is defined for all $t \le s$.

2.13. The generator: The *generator* of an Ito process is the *operator* containing the spatial part of the backward equation[13]

$$L(t) = a(x,t)\partial_x + \tfrac{1}{2}\sigma^2(x,t)\partial_x^2 \; .$$

The backward equation is $\partial_t f(x,t) + L(t)f(x,t) = 0$. We write just L when a and σ do not depend on t. For a general continuous time Markov process, the generator is defined by the requirement that

$$\frac{d}{dt} E[g(X(t),t)] = E\left[(L(t)g)(X(t),t) + g_t(X(t),t) \right] \; , \tag{43}$$

for a sufficiently rich (dense) family of functions g. This applies not only to Ito processes (diffusions), but also to jump diffusions, continuous time birth/death processes, continuous time Markov chains, etc. Part of the requirement is that the limit defining the derivative on the left side should exist. Proving (43) for an Ito process is more or less what we did when we derived the backward equation. On the other hand, if we know (43) we can derive the backward equation by requiring that $\frac{d}{dt}E[f(X(t),t)] = 0$.

2.14. Adjoint: The adjoint of L is another operator that we call L^*. It is defined in terms of the inner product

$$\langle u, f \rangle = \int_{-\infty}^{\infty} u(x) f(x) dx \; .$$

We leave out the t variable for the moment. If u and f are complex, we take the complex conjugate of u above. The adjoint is defined by the requirement that for general u and f,

$$\langle u, Lf \rangle = \langle L^* u, f \rangle \; .$$

[13]Some people include the time derivative in the definition of the generator. Watch for this.

In practice, this boils down to the same integration by parts we used to derive the forward equation from the backward equation:

$$\begin{aligned}\langle u, Lf\rangle &= \int_{-\infty}^{\infty} u(x)\big(a(x)\partial_x f(x) + \tfrac{1}{2}\sigma^2(x)\partial_x^2 f(x)\big)dx \\ &= \int_{-\infty}^{\infty} \big(-\partial_x(a(x)u(x)) + \tfrac{1}{2}\partial_x^2(\sigma^2(x)u(x))\big) f(x)dx \ .\end{aligned}$$

Putting the t dependence back in, we find the "action" of L^* on u to be

$$(L(t)^* u)(x,t) = -\partial_x(a(x,t)u(x,t)) + \tfrac{1}{2}\partial_x^2(\sigma^2(x,t)u(x,t)) \ .$$

The forward equation (37) then may be written

$$\partial_t u = L(t)^* u \ .$$

All we have done here is define notation (L^*) and show how our previous derivation of the forward equation is expressed in terms of it.

2.15. Adjoints and the Green's function: Let us summarize and record what we have said about the transition probability density $G(y,x,t,s)$. It is defined for $s \geq t$ and has $G(x,y,t,t) = \delta(x-y)$. It moves probabilities forward by integrating over y (38) and moves expected values backward by integrating over x **??**). As a function of x and s it satisfies the forward equation

$$\partial_s G(y,x,t,s) = (L_x^*(t)G)(y,x,t,s) \ .$$

We write L_x^* to indicate that the derivatives in L^* are with respect to the x variable:

$$(L_x^*(t)G)(y,x,t,s) = -\partial_x(a(x,t)G(y,x,t,s)) + \tfrac{1}{2}\partial_x^2(\sigma^2(x,t)G(y,x,t,s)) \ .$$

As a function of y and t it satisfies the backward equation

$$\partial_t G(y,x,t,s) + (L_y(t)G)(y,x,t,s) = 0 \ .$$

3 Properties of the solutions

3.1. Introduction: The next few paragraphs describe some properties of solutions of backward and forward equations. For Brownian motion, f and u have every property because the forward and backward equations are essentially the same. Here f has some and u has others.

3.2. Backward equation maximum principle: The backward equation has a *maximum principle*

$$\max_x f(x,t) \leq \max_y f(y,s) \quad \text{for } s > t. \tag{44}$$

This follows immediately from the representation

$$f(x,t) = E_{x,t}[f(X(s),s)] \ .$$

The expected value of $f(X(s),s)$ cannot be larger than its maximum value. Since this holds for every x, it holds in particular for the maximizer.

There is a more complicated proof of the maximum principle that uses the backward equation. I give a slightly naive explination to avoid taking too long with it. Let $m(t) = \max_x f(x,t)$. We are trying to show that $m(t)$ never increases. If, on the contrary, $m(t)$ does increase as t decreases, there must be a t_* with $\frac{dm}{dt}(t_*) = \alpha < 0$. Choose x_* so that $f(x_*,t_*) = \max_x f(x,t_*)$. Then $f_x(x_*,t_*) = 0$ and $f_{xx}(x_*,t_*) \leq 0$. The backward equation then implies that $f_t(x_*,t_*) \geq 0$ (because $\sigma^2 \geq 0$), which contradicts $f_t(x_*,t_*) \leq \alpha < 0$.

The PDE proof of the maximum principle shows that the coefficients a and σ^2 have to be outside the derivatives in the backward equation. Our argument that $Lf \leq 0$ at a maximum where $f_x = 0$ and $f_{xx} \leq 0$ would be wrong if we had, say, $\partial_x(a(x)f(x,t))$ rather than $a(x)\partial_x f(x,t)$. We could get a non zero value because of variation in $a(x)$ even when f was constant. The forward equation does not have a maximum principle for this reason. Both the Ornstein Uhlenbeck and geometric Brownian motion problems have cases where $\max_x u(x,t)$ increases in forward time or backward time.

3.3. Conservation of probability: The probability density has $\int_{-\infty}^{\infty} u(x,t)dx = 1$. We can see that

$$\frac{d}{dt}\int_{-\infty}^{\infty} u(x,t)dx = 0$$

also from the forward equation (37). We simply differentiate under the integral, substitute from the equation, and integrate the resulting x derivatives. For this it is crucial that the coefficients a and σ^2 be inside the derivatives. Almost any example with $a(x,t)$ or $\sigma(x,t)$ not independent of x will show that

$$\frac{d}{dt}\int_{-\infty}^{\infty} f(x,t)dx \neq 0 \ .$$

3.4. Martingale property: If there is no drift, $a(x,t) = 0$, then $X(t)$ is a martingale. In particular, $E[X(t)]$ is independent of t. This too follows from the forward equation (37). There will be no boundary contributions in the integrations by parts.

$$\begin{aligned}
\frac{d}{dt}E[X(t)] &= \frac{d}{dt}\int_{-\infty}^{\infty} xu(x,t)dx \\
&= \int_{-\infty}^{\infty} xu_t(x,t) \\
&= \int_{-\infty}^{\infty} x\tfrac{1}{2}\partial_x^2(\sigma^2(x,t)u(x,t))dx
\end{aligned}$$

$$\begin{aligned} &= -\int_{-\infty}^{\infty} \tfrac{1}{2}\partial_x(\sigma^2(x,t)u(x,t))dx \\ &= 0\;. \end{aligned}$$

This would not be true for the backward equation form $\frac{1}{2}\sigma^2(x,t)\partial_x^2 f(x,t)$ or even for the mixed form we get from the Stratonovich calculus $\frac{1}{2}\partial_x(\sigma^2(x,t)\partial_x f(x,t))$. The mixed Stratonovich form conserves probability but not expected value.

3.5. Drift and advection: If there is no drift then the SDE (23) becomes the ordinary differential equation (ODE)

$$\frac{dx}{dt} = a(x,t)\;. \tag{45}$$

If $x(t)$ is a solution, then clearly the expected payout should satisfy $f(x(t),t) = f(x(s),s)$, if nothing is random then the expected value is *the* value. It is easy to check using the backward equation that $f(x(t),t)$ is independent of t if $x(t)$ satisfies (45) and $\sigma = 0$:

$$\frac{d}{dt}f(x(t),t) = f_t(x(t),t) + \frac{dx}{dt}f_x(x(t),t) = f_t(x(t),t) + a(x(t),t)f_x(x(t),t) = 0\;.$$

Advection is the process of being carried by the wind. If there is no diffusion, then the values of f are simply advected by the drift. The term *drift* implies that this advection process is slow and gentle. If σ is small but not zero, then f may be essentially advected with a little *spreading* or *smearing* induced by diffusion. Computing drift dominated solutions can be more challenging than computing diffusion dominated ones.

The probability density does not have $u(x(t),t)$ a constant (try it in the forward equation). There is a conservation of probability correction to this that you can find if you are interested.