Stochastic Calculus Notes, Lecture 7
Last modified November 7, 2002

1 Ito Stochastic Differential Equations

1.1. Notation: We switch back to the notation W_t for Brownian motion. We use X_t to denote the solution of the stochastic differential equation (SDE). When we write forward and backward equations for X_t, the independent variable will still be x. Often we work in more than one dimension. In this case, W_t may be a vector of independent Brownian motion paths. As far as possible, we will use the same notation for the one dimensional (scalar) and multidimensional cases. The solution of the Ito differential equation will be X_t. We sometimes call these "diffusions".

1.2. The SDE: A stochastic differential equation is written

$$dX_t = a(X_t, t)dt + \sigma(X_t, t)dW_t \ . \qquad (1)$$

A solution to (1) is a process $X_t(W)$ that is an adapted function of W ($X_t \in \mathcal{F}_t$, where $\mathcal{F}_t$ is generated by the values W_s for $s \leq t$), so that

$$X_T = X_0 + \int_0^T a(X_t, t)dt + \int_0^T \sigma(X_t, t)dW_t \ . \qquad (2)$$

Because X_t is adapted, the Ito integral on the right of (2) makes sense. The term $a(X_t, t)dt$ is called the "drift" term. If $a \equiv 0$, X_t will be a martingale; any change in $E[X_t]$ is due to the drift term. The term $\sigma(X_t, t)dW_t$ is the "noise" term. The coefficient σ may be called the "diffusion" coefficient, or the "volitility" coefficient, though both of these are slight misnomers. The volitility coefficient determines the size of the small scale random motions that characterize diffusion processes. The form (1) is really just a shorthand for (2). It is traditionally written in differential notation (dX_t, dt, dW_t) as a reminder that Ito differentials are mor subtle than ordinary differentials from calculus with differentiable functions.

What separates diffusion processes from simple Brownian motions is that in diffusions the drift and volatility coefficients may depend on X and t. It might be, for example, that when X is large, its fluctuation rate is also large. This would be modelled by having $\sigma(x, t)$ being an increasing function of x.

In the multidimensional case, we might have $X_t \in R^n$. Clearly, this calls for $a(x, t) \in R^n$ also. This might be called the "drift vector" or "velocity field" or "drift field". The volitility coefficient becomes an $n \times m$ matrix, with $W_t \in R^m$ being m independent sources of noise. The case $m < n$ is called "degenerate diffusion" and arises often in applications. The case $n = m$ and σ non singular is called "nondegenerate diffusion". The mathematical character of the forward and backward equations is far more subtle for degenerate diffusions than for nondegenerate diffusions. The case $m > n$ arises in practice only by mistake.

1.3. Existence and uniqueness of Ito solutions: Just as the Ito value of the stochastic integral is one of several possible values depending on details of the definition, we might expect the solution of (1) to be ambiguous. We will now see that this is not so as long as we use the Ito definition of the stochastic integral in (2). The main technical fact in the existence/uniqueness theory is a "short time contraction estimate": the mapping defined by (2) is a contraction for if t is small enough. Both the existence and uniqueness theorems follow quickly from this.

Suppose X_t and Y_t are two adapted stochastic processes with $X_0 = Y_0$. We define $\tilde{X}_t$ from X_t using (2) by

$$\tilde{X}_T = \int_{t=0}^{T} a(X_t, t)dt + \int_{t=0}^{T} \sigma(X_t, t)dW_t \ .$$

In the same way, $\tilde{Y}$ is defined from Y. We assume that a and σ are Lipshchitz continuous in the x arguments: $|a(x,t) - a(y,t)| \leq M\,|x-y|$, $|a(x,t) - a(y,t)| \leq M\,|x-y|$. The best possible constants in these inequalities are called the "Lipschitz constants" for a and σ. The mapping $X \mapsto \tilde{X}$ is a" contraction" if

$$\left\| \tilde{X} - \tilde{Y} \right\| \leq \alpha \left\| X - Y \right\| \ ,$$

for some $\alpha < 1$, that is, if the mapping shortens distances between objects by a definite ratio less than one. Of course, whether a mapping is a contraction might depend on the sense of distance, the norm $\|\cdot\|$. Because our tool is the Ito isometry formula, we use

$$\|X - Y\|_T^2 = \max_{0 \leq t \leq T} E\left[\left(X_t - Y_t\right)^2 \right] ; .$$

The contraction lemma is:
Lemma: If a and σ are Lipshcitz with Lipschitz constant M, then

$$\left\| \tilde{X} - \tilde{Y} \right\|_T^2 \leq 4M^2 T \left\| X - Y \right\|_T^2 \ . \tag{3}$$

For the proof, we first write

$$\tilde{X}_T - \tilde{Y}_T = \int_{t=0}^{T} \big(a(X_t,t) - a(Y_t,t)\big)dt + \int_{t=0}^{T} \big(\sigma(X_t,t) - \sigma(Y_t,t)\big)dW_t \ .$$

We have $E[(\tilde{X}_T - \tilde{Y}_T)^2] \leq 2A + 2B$ where

$$A = E\left[\left(\int_{t=0}^{T} \big(a(X_t,t) - a(Y_t,t)\big)dt \right)^2 \right] \ .$$

and

$$B = E\left[\left(\int_{t=0}^{T} \big(\sigma(X_t,t) - \sigma(Y_t,t)\big)dW_t \right)^2 \right] \ .$$

Bounding the B term is an application of the Ito isometry formula. Indeed,

$$B \leq \int_{t=0}^{T} E\left[\left(\sigma(X_t,t) - \sigma(Y_t,t)\right)^2\right] dt \ ,$$

Using the Lipschitz continuity of σ then gives

$$B \leq M^2 T \max_{0\leq t\leq T} E[(X_t - Y_t)^2] \ ,$$

which is the sort of bound we need.

The A term is an application of the Cauchy Schwartz inequality

$$\left(\int_{t=0}^{T} \left(a(X_t,t) - a(Y_t,t)\right)dt\right)^2 \leq \int_{t=0}^{T} \left(a(X_t,t) - a(Y_t,t)\right)^2 dt \cdot \int_{t=0}^{T} 1dt$$

If we now use the Lipschitz continuity of a and take expectations of both sides, we get

$$A \leq M^2 T \max_{0\leq t\leq T} E[(X_t - Y_t)^2] \ ,$$

These two inequalities prove the contraction lemma estimate (3).

1.4. Uniqueness: The contraction inequality gives a quick proof of the uniqueness theorem. We will see that if X_0 is a random variable, then the solution up to some time T is unique. Of course, then X_T is a random variable and may be thought of as initial data for the next T time period. This give uniqueness up to time $2T$, and so on. Suppose X_t and Y_t were two solutions of (2). We want to argue that $E[(X_t - Y_t)^2] < \alpha E[(X_t - Y_t)^2]$ for $\alpha < 1$. This is impossible unless $[(X_t - Y_t)^2] = 0$, that is, unless $X_t = Y_t$. From (3), we will have $\alpha < 1$ if $T < 1/4M^2$.

The contraction lemma does not say $E[(\tilde{X}_t - \tilde{Y}_t)^2] < \alpha E[(X_t - Y_t)^2]$. To work with the information it actually gives, define $m_T = \max_{t<T} E[(X_t - Y_t)^2]$, and $\tilde{m}_T = \max_{t<T} E[(\tilde{X}_t - \tilde{Y}_t)^2]$. From the definitions, it is clear that m_t is an increasing function of t, so that (3) implies that $E[(\tilde{X}_t - \tilde{Y}_t)^2] \leq \alpha m_t \leq \alpha m_T$ if $T > t$. That is, (3) implies that $\tilde{m}_T \leq \alpha m_T$. This gives a contradiction as before: Since X and Y are solutions, we have $\tilde{m} = m$, so $\tilde{m}_T \leq \alpha m_T$ is impossible unless $m_T = 0$.

1.5. Existence of solutions via Picard iteration: The contraction inequality (3) allows us also to show that there is an X_T satisfying (2), at least for $T < 1/4M^2$. You might remember this construction, Picard iteration, from a class in ordinary differential equations. The first "iterate" does not come close to satisfying the equations but just gets the ball rolling: $X_t^{(0)} = X_0$ for all $t \leq T$. This $X_t^{(0)}$ does not depend on W_t, but it will still be random if X_0 is random. For $k > 0$, the iterates are defined by

$$X_t^{(k)} = \int_{s=0}^{t} a(X_s^{(k-1)},t)dt + \int_{s=0}^{t} \sigma(X_s^{(k-1)},t)dW_t \ . \tag{4}$$

The contraction inequality implies that the Picard iterates, $X^{(k)}$, converge as $k \to \infty$. In (3), take X to be $X^{(k-1)}$, and $Y = X^{(k)}$. Then $\tilde{X} = X^{(k)}$ and $\tilde{Y} = X^{(k+1)}$. If we define

$$m_T^{(k)} = \max_{0 \le t \le T} E\left[\left(X_t^{(k)} - X_t^{(k)} \right)^2 \right] \ ,$$

and use the ideas of the previous paragraph, (3) gives

$$m_T^{(k+1)} \le \alpha m_T^{(k)} \ .$$

This implies that, for any $t \le T$, the iterates $X_t^{(k)}$ have $E[(X_t^{(k+1)} - X_t^{(k)})^2] \le m_T^{(0)}$, which (as we saw in the previous lecture) implies that $\lim_k \to \infty X_t^{(k)}$ exists. The contraction inequality also shows (reader: think this through) that this limit, X_t satisfies (2) and therefore is what we are looking for.

1.6. Diffusions as martingales: If the drift coefficient in (1) vanishes, $a(x,t) \equiv 0$, then the process X_t is a martingale. Indeed, any process $X_t = \int_0^t F_s dW_s$, with a nonanticipating F_s is a martingale. There is a very general converse to this statement. More or less (leaving out the technical details, obviously), any adapted process, X_t, with continuous sample paths,

$$P(\text{"}X_t \text{ is a continuous function of } t'') = 1 \ ,$$

has a representation in the form (1), except that in general, we must take σ to be a general adapted function of t, not necessarily a function of X_t only. In discrete time, a martingale, X_k, may be written as a sum of martingale differences, $Y_k = X_k - X_{k-1}$, in that $X_k = X_0 + \sum_{j=0}^{k-1} Y_j$. The Ito integral representation of the continuous time martingale X_t is a continuous time version of the representation of a discrete time martingale as a sum of martingale differences. What makes the continuous time version really different (rather than just technically different) is the unique role of Brownian motion. The proof has to construct the Brownian motion path related to X.

1.7. The structure of correlated gaussians: In the multidimensional case, σ will be a matrix. We think of σdW_t as the source of noise. The several components of σdW_t may be correlated, modeling the fact that the noise terms driving the several components of X_t are correlated. The matrix σ tells us how to make correlated noises of varying strengths from uncorrelated noises of constant strength, the components of W_t. The role of σ is to correlate the noise sources and to modulate their strengths.

One often hears people referring, for example, to tow correlated Brownian motion paths, with correlation coefficient ρ. A simpler special case of this would be standard normal random variables, Z_1 and Z_2, with correlation ρ. If we suppose that (Z_1, Z_2) form a multivariate (bivariate) normal, the covariance matrix has entries $C_{11} = \text{var}(Z_1) = 1$, $C_{22} = \text{var}(Z_2) = 1$, and $C_{12} = \text{cov}(Z_1, Z_2) = \rho$. The correlation coefficient and the covariance are the same here because the variances are both one. We can make such correlated normals from uncorrelated normals in the following way. Let U_1 and U_2 be independent standard normals. Take $Z_1 = U_1$ and $Z_2 = \rho U_1 + \sqrt{1-\rho^2}U_2$. The term ρU_1 in the Z_2 formula gives the correct correlation with U_1, provided the rest of Z_2 is independent of Z_1. The term $\sqrt{1-\rho^2}U_2$ gives enough independent noise so that $\mathrm{var}(Z_2) = 1$. In matrix form, this is

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix} \cdot \begin{pmatrix} U_1 \\ U_2 \end{pmatrix} \ .$$

The point is that you can make correlated standard normals from independent ones, but you need a matrix, σ.

And the σ you need is not unique. Suppose σ is an $n \times M$ matrix, and $\sigma_1 = \sigma Q$, where Q is an $m \times m$ orthogonal matrix. If U is an m vector of independent standard normals, then $Z = \sigma U$, and $Z^{(1)} = \sigma_1 U$ are each multivariate normals with the same probability distribution. That is, Z and $Z^{(1)}$ are indistinguishable if you do now know U. Applied to SDEs, this says that Z and $Z^{(1)}$ produce paths X and $X^{(1)}$ that have are indistinguishable if you do not know W. In particular, the "QR" factorization of σ^*, w(i.e. the "LQ" factorization of σ) says that we may take σ to be lower triangular. If σ is lower triangular, the components of W beyond the n^{th} all have coefficient zero. This is why it is a mistake if you have more sources of noise than components of X.

2 Ito's Lemma

We want to work out the first few Picard iterates in an example. This leads to a large number of stochastic integrals. We could calculate any of them in an hour or so, but we would soon long for something like the Fundamental Theorem of calculus to make the calculations mechanical. That result is called "Ito's lemma". Not only is it helpful in working with stochastic integrals and SDE's, it is also a common interview question for young potential quants. Here is the answer.

2.1. The Fundamental Theorem of calculus: The following derivation of the Fundamental Theorem of ordinary calculus provides a template for the derivation of Ito's lemma. Let $V(t)$ be a differentiable function of t with $\partial_t V$ being Lipschitz continuous. The Fundamental Theorem states that (writing $\partial_t V$ for dV/dt although V depends only on t):

$$V(T) - V(0) = \int_0^T dV = \int_0^T \partial_t V(t) dt \ .$$

This exact formula follows from two approximate short time approximations, the first of which is

$$V(t + \Delta t) - V(t) = \partial_t V(t) \Delta t + O(\Delta t^2) \ .$$

The second approximation is (writing $\partial_t V(s)$ for $V'(s)$):

$$\int_t^{t+\Delta t} \partial_t V(s) ds = \partial_t V(t)\Delta t + O(\Delta t^2) \ .$$

Using our habitual notation ($\Delta t = T/n = T/2^L$, $t_k = k\Delta t$, $V_k = V(t_k)$), we have, using both approximations above,

$$\begin{aligned} V_n - V_0 &= \sum_{k=0}^{n-1} \big(V_{k+1} - V_k \big) \\ &= \sum_{k=0}^{n-1} \big(\partial_t V(t_k)\Delta t + O(\Delta t^2) \big) \\ &= \sum_{k=0}^{n-1} \big(\int_{t_k}^{t_{k+1}} \partial_t V(t_k) dt + O(\Delta t^2) \big) \\ &= \int_0^T \partial_t V(t) dt + nO(\Delta t^2) \ . \end{aligned}$$

Because $n\Delta t = T$, $n\Delta t^2 = TO(\Delta t) \to 0$ as $n \to \infty$.

2.2. The Ito dV_t: The Fundamental Theorem may be stated $dV = \partial_t V dt$. This definition makes

$$\int_0^T dV_t = V(T) - V(0) \ . \tag{5}$$

We want to extend this to functions V_t that depend on W as well as t. For any adapted function, we define dV_t so that (5) holds. For example, if U_t is an adapted process and $V_T = \int_0^T U_t dW_t$, then $dV_t = U_t dW_t$ because that makes (5) hold. Ito's lemma is a statement of what makes (5) hold for specific adapted functions V_t.

2.3. First version: Our first version of Ito's lemma is a calculation of dV_t when $V_t = V(W_t, t)$ and V and W are one dimensional. The result is

$$dV_t = \partial_W V(W_t, t) dW_t + \frac{1}{2}\partial_W^2 V(W_t, t) dt + \partial_t V dt \ . \tag{6}$$

What's particular to stochastic calculus is the "Ito term" $\frac{1}{2}\partial_W^2 V(W_t, t)dt$. Even if we can't guess the precise form of the term, we know something has to be there. In the special case $V_t = V(W_t)$, the $\partial_t V dt$ term is missing. The guess $dV = \partial_W V dW_t$ would give (see (5)) $V(t) - V(0) = \int_0^T \partial_W V dW_t$. We know this cannot be correct: the right side is a martingale while the left side is not (see assignment 5, question 1). To make the martingale integral into the non martingale answer, we have to add a dt integral, which is why some term like $\frac{1}{2}\partial_W^2 V(W_t, t)dt$ is needed. A motivation for the specific form of the Ito term is the observation that it should vanish when V is a linear function of W.

2.4. Derivation, short time approximations: The defivation of Ito's lemma starts with the stochastic versions of the two short time approximations behind the Fundamental Theorem. For convenience, we drop all t subscripts and write ΔW for $W_{t+\Delta t} - W_t$. We have

$$\begin{aligned} & V(W_{t+\Delta t}, t+\Delta t) - V(W,t) = \\ & \qquad \partial_W V(W,t)\Delta W + \frac{1}{2}\partial_W^2 V(W,t)\Delta W^2 + \partial_t V(W,t)\Delta t + O(\Delta t^{3/2}) \; . \end{aligned}$$

The other short time approximation is provided by assignment 7, question 3, applied to $\partial_W V$:

$$\int_t^{t+\Delta t} \partial_W V(W_s,s)dW_s = \partial_W V(W_t,t)\Delta W + \frac{1}{2}\partial_W^2 V(W,t)\big(\Delta W^2 - \Delta t\big) + O(\Delta t^{3/2}) \; .$$

For dt integrals, the result is simply

$$\int_t^{t+\Delta t} U(W_s,s)ds = U(W,t)\Delta t + O(\Delta t^{3/2} \; .$$

The error term is $O(\Delta t^{3/3})$ rather than $O(\Delta t^2)$ because W_t is not a Lipshcitz continuous function of t. We conbine these approximations with a little algebra $(\Delta W^2 = \Delta t + (\Delta W^2 - \Delta t)$, which might be considered the main idea of this section) gives

$$\begin{aligned} \Delta V \quad = \quad & \int_t^{t+\Delta t} \partial_W V(W_s,t)dW_s \\ & + \int_t^{t+\Delta t} \left(\frac{1}{2}\partial_W^2 V(W_s,s) + \partial_t V(W_s,s)\right) ds \\ & + \partial_W^2 V(W,t)\big(\Delta W^2 - \Delta t\big) + O(\Delta t^{3/2}) \; . \end{aligned}$$

As with the Fundamental Theorem, we apply this with $t = t_k$ (in the habitual notation) and sum over k, giving:

$$\begin{aligned} V(T) - V(0) \quad = \quad & \int_0^T \partial_W V(W_t,t)dW_t \\ & + \int_0^T \left(\frac{1}{2}\partial_W^2 V(W_t,t) + \partial_t V(W_t,t)\right) dt \\ & + \sum_{k=0}^{n-1} \partial_W^2 V(W_k,t_k)\big(\Delta W_k^2 - \Delta t\big) + O(T\sqrt{\Delta t}) \; . \end{aligned}$$

2.5. The non Newtonian step: The final step in deriving Ito's lemma has no anologue in the proof of Newton's Fundamental Theorem of calculus. We study the term

$$A = \sum_{k=0}^{n-1} \partial_W^2 V(W_k,t_k)\big(\Delta W_k^2 - \Delta t\big)$$

and show that $A \to 0$ as $\Delta t \to 0$ (actually, as $L \to \infty$ with $\Delta t = T/2^L$) almost surely. Previous experience might lead us to calculate $E[A_L^2]$. This follows a well worn path. We have the double sum expression:

$$E[A_L^2] = \frac{1}{4}\sum_{j,k} E\left[(\cdot)_j(\cdot)_k\right] .$$

The $j \neq k$ terms have expected value zero because (if $k > j$) $E[\Delta W_k^2 - \Delta t \mid \mathcal{F}_{t_k}] = 0$. We get a bound for the $j = k$ terms using $E[(\Delta W_k^2 - \Delta t)^2 \mid \mathcal{F}_{t_k}] = 2\Delta t^2$:

$$E\left[\partial_W^2 V(W_k, .t_k)^2 \left(\Delta W_k^2 - \Delta t\right)^2 \mid \mathcal{F}_{t_k}\right] \leq C \cdot \Delta t^2 .$$

Altogether, we get $E[A_L^2] \leq C\Delta t = C2^{-L}$, which implies that $A_L \to 0$ as $L \to \infty$, almost surely (see the next paragraph). This completes our proof of the first form of Ito's lemma, (6).

2.6. A Technical Detail: Here is a proof that uses the inequalities $E[A_L^2] \leq Ce^{-\beta L}$ for some $\beta > 0$ and proves that $A_L \to 0$ as $L \to \infty$ almost surely. The proof is an easier version of an argument used in the previous lecture. As in that lecture, we start with a observation, this time that $|A_L| \to 0$ as $L \to \infty$ if $\sum_{L=1}^{\infty} |A_L| < \infty$. Also, the sum is finite almost surely if it's expected value is finite. That is, if $\sum_{L=1}^{\infty} E[|A_L|] < \infty$. Finally, the Cauchy Schwartz inequality gives $E[|A_l] \leq Ce^{-\beta L/2}$. Since this has a finite sum (over L), we get almost sure convergence $A_L \to 0$ as $L \to \infty$.

2.7. Integration by parts: In ordinary Newtonian (and Leibnitzian) calculus, the integration by parts identity is a consequence of the Fundamental Theorem and facts about differentiation (the Leibnitz rule). So let it be for Ito. For instance, integration by parts might lead to

$$\int_0^T t dW_t = TW_T - \int_0^T W_t dt . \tag{7}$$

We can check whether this actually is true by taking the Ito differential of tW_t:

$$\begin{aligned} d(tW_t) &= \partial_W(tW_t)dW_t + \frac{1}{2}\partial_W^2(tW_t)dt + \partial_t(tW_t)dt \\ &= tdW_t + W_t dt . \end{aligned}$$

This implies that

$$TW_T = \int_0^T t dW_t + \int_0^T W_t dt ,$$

which is a confirmation of (7). We can get a more general version of the same thing if we apply the Ito differential to $f(t)g(W_t)$ (reader: do this).

2.8. Doing $\int W_t dW_t$ the easy way: If Ito's lemma is to play the role of the Fundamental Theorem of calculus, it should help us calculate stochastic integrals. In ordinary calculus we calculate integrals by differentiating guesses to see which guess works. After a while, we become more systematic guessers. To compute a stochastic integral, we need to guess a function F_t so that dF_t is the integrand. A first example of this is

$$Y_T = \int_0^T W_t dW_t \; . \tag{8}$$

Using ordinary calculus as a clue, we might try $F_t = \frac{1}{2}W_t^2$. We calculate, using (6),

$$dF = \partial_W F dW + \frac{1}{2}\partial_W^2 F dt + \partial_t F = W dW + dt + 0 \; .$$

We see that we did not get the desired answer, dF is not the integrand WdW. However, it is almost right, missing by dt. To correct for this, try the more sophisticated guess $F = \frac{1}{2}W_t^2 - t$. Repeating the differentiation, we see that indeed

$$d(\frac{1}{2}W_t^2 - t) = W_t dW_t \; .$$

as desired. Ito's lemma than tells us that

$$\frac{1}{2}W_T^2 - T - (\frac{1}{2}W_T^2 - T) = \int_0^T W_t dW_t \; .$$

2.9. $\int W_t^2 dX_t$ the easy way: To calculate

$$\int_0^T W_t^2 dW_t$$

we again start with the calculus guess, which this time is $F = \frac{1}{3}W_t^3$. The Ito differential of this is

$$d\frac{1}{3}W_t^3 = W_t^2 dW_t + \frac{1}{2}2W_t dt \; .$$

This differs from our integrand ($W_t^2 dW_t$) by the term $W_t dt$. We can get $W_t dt$ by differentiating $\int_0^t W_s ds$. Therefore,

$$\int_0^T W_t^2 dW_t = \frac{1}{3}W_T^3 - \int_0^T W_t dt \; .$$

If you still consider this to be a guess, you can check it by taking the differential of both sides. The left side gives $W_T^2 dW_T$. The right side gives $W_T^2 dW_T + W_T dt - W_T dt$, which is the same thing.

2.10. Solving an SDE: Here is one way to solve the SDE:

$$dX_t = X_t dW_t \;\; , \quad X_0 = 1 \; . \tag{9}$$

The ordinary calculus result would be $X_T = e^{W_T}$. To see whether this satisfies (9), we calculate the Ito differential:

$$de_T^W = \partial_W e_T^W dW_T + \frac{1}{2}\partial_W^2 e^{W_T} dt + \partial_t e^{W_T} dt = e_T^W dW_T + \frac{1}{2} e^{W_T} dt \; .$$

The first term on the left is indeed $X_T dW_T$, so we need somehow to get rid of the second term. After some false starts, we hit on the idea to try a solution of the form $X_t = A(t)e^{W_t}$. Now the differential is

$$\begin{aligned} d\big(A(t)e^{W_t}\big) &= A(t)\partial_W e_t^W dW_t + A(t)\frac{1}{2}\partial_W^2 e^{W_t} dt + \partial_t A(t) e^{W_t} dt \\ &= A(t)e_t^W dW_t + A(t)\frac{1}{2} e^{W_t} dt + \dot{A}(t) e^{W_t} dt \; . \end{aligned}$$

The first term on the right is the desired answer $X_t dW_t$. The second and third terms will cancel if $\frac{1}{2}A + \dot{A} = 0$, i.e. if $A(t) = e^{-t/2}$. Our new guess, then, is $X_t = e^{W_t - t/2}$. We can check this with the calculation $de^{W_t - t/2} = e^{W_t - t/2} dW_t$ (because, by design, the dt terms cancel).

A consistency check is that X_t should be a martingale, because $X_T = \int_0^T X_t dW_t$, and the Ito integral always gives a martingale. We can check, for example, that $E[X_t] = 1$ for any t.

2.11. Differentials of functions of X_t: The formal formulation (1) of an Ito SDE is in fact a relation among Ito differentials, which is precisely what (2) says. We can also compute $dV(X_t)$ (or even $dV(X_t, t)$, which is more complicated but not harder) using the reasoning in paragraphs 2.4 and 2.5 above. I will breeze through the argument, commenting only on the differences. Some of the details are left to assignment 8. We can calculate

$$\Delta V(X) = \partial_X V(x_t)\Delta X_t + \frac{1}{2}\partial_X^2 V(X)\Delta X^2 + O(\Delta t^{3/2}) \; .$$

Also

$$\int_t^{t+\Delta t} \partial_X V(X_s) dX_s = \partial_X V(X_t)\Delta X_t + \frac{1}{2}\partial_X^2 V(X_t)\big(\Delta X^2 - \sigma(X_t)^2 \Delta t\big) + O(\Delta t^{3/2}) \; .$$

The new feature is that $E[\Delta X^2] = \sigma(X_t)^2 \Delta t + O(\Delta t^{3/2})$. After this, the derivation proceeds as before, eventually giving

$$dV(X_t) = \partial_X V(X_t) dX_t + \frac{1}{2}\partial_X^2 V(X_t)\sigma(X_t)^2 dt \; . \tag{10}$$

2.12. The "Ito rule" $dW^2 = dt$: The first version of Ito's lemma can be summarized as using Taylor series calculations and neglecting all terms of higher than first order except for dW_t^2, which we replace by dt. You might think this is based on the approximation $\Delta W^2 \approx \Delta t$ for small Δt. The real story is a little more involved. The relative accuracy of the approximation $\Delta W^2 \approx \Delta t$ does not improve as $\Delta t \to 0$. Both sides go to zero, and at the same rate, but they do not get closer to each other in relative terms. In fact, the expected error, $E[|\Delta W^2 - \Delta t|]$, is also of order Δt. If $\Delta t = .1$ then ΔW^2 is just as likely to be .2 as .1, not really a useful approximation. The origin of Ito's rule is that ΔW^2 and Δt have the same expected value. For that reason, if we add up m ΔW^2 values, we are likely to get a number close to $m\Delta t$ if m is large. We might say $\int_a^b dW^2 = \int_a^b dt$, thinking that each side is make up of a large number (an infinite number) of tiny ΔW^2 or Δt values. Remember that for any Q, the Ito dQ is what you have to integrate to get Q. Integrating dW^2 gives the same result as integrating dt.

2.13. Quadratic variation: The informal ideas of the preceding paragraph may be fleshed out using the "quadratic variation" of a process. We already discussed the quadratic variation of Brownian motion. For a general stochastic process, X_t, the quadratic variation is

$$\langle X \rangle_t = \lim_{\Delta t \to 0} \sum_{k=0}^{n} \left(X_{k+1} - X_k \right)^2 . \tag{11}$$

If we apply the approximation from assignment 8, question 2b, we get

$$\sum_{k=0}^{n} \left(X_{k+1} - X_k \right)^2 = \left(\sum_{k=0}^{n} \sigma(X_k, t_k)^2 \Delta X_k^2 \right) + O(n\Delta t^3/2) .$$

Our usual trick is to use $\Delta X_k^2 = \Delta t + (\Delta X^2 - \Delta t)$ to write the last sum as an approximation of a dt integral plus something with mean zero that does not add up to much. The result is

$$\langle X \rangle_t = \int_0^t \sigma^2(X_s, t) ds .$$

In particular,

$$d \langle X \rangle_t = \sigma^2(X_t) dt .$$

Ito's lemma for X_t satisfying the SDE (1) may be written

$$dV(X_t) = \partial_X V(X_t) dX_t + \frac{1}{2} \partial_X^2 V(X_t) d \langle X \rangle_t . \tag{12}$$

2.14. Geometric Brownian motion again: Here is another way to find the solution of $dX_t = X_t dt$. Since we expect X_t to be an exponential, we calculate the SDE satisfied by $Y_t = \log(X_t)$. Ito's lemma in the form (12) allows us to calculate

$$dY_t \quad = \quad \partial_X \log(X_t) dX_t + \frac{1}{2} \partial_X^2 \log(X_t) X_t^2 dt$$

$$
\begin{aligned}
&= \frac{1}{X_t} X_t dW_t + \frac{1}{2}\left(-\frac{1}{X_t^2}\right) X_t^2 dt \\
dY_t &= dW_t - \frac{1}{2} dt \ .
\end{aligned}
$$

This gives $Y_t = Y_0 + W_t - \frac{t}{2}$. Since $X_t = e^{Y_t}$, we get $X_t = X_0 e^{W_t - t/2}$, as before.

2.15. Remarks on the solution: The solution $X_t = X_0 e^{W_t - t/2}$ provides some insight into how martingales can behave and the importance of rare events. We know that Brownian motion paths W_t are on the order of $\sqrt{T}$. Therefore for large t, the exponent is $W_t - \frac{t}{2} \approx -t/2$. That is, nearly all (not almost all) geometric Brownian motion paths are exponentially small for any particular large t. Nevertheless, since X_t is a martingale, $E[X_t] = 1$. Those rare paths with $W_t > t/2$ are just big enough and just likely enough to save $E[X_t]$ from being exponentially small. For the record, $P(W_t > t/2) < e^{-t/8}$, is very small (about 1/1000 for $t = 100$). This means that if you simulate, say, 500 paths, there is a pretty good chance that none of them is as big as the mean. Monte carlo simulation is very unreliable in such cases.