Lecture 8: Stochastic Differential Equations

Readings

Recommended:
- Pavliotis [2014] 3.2-3.5
- Oksendal [2005] Ch. 5

Optional:
- Oksendal [2005] 7.1,7.2 (on Markov property)

We’d like to understand solutions to the following type of equation, called a Stochastic Differential Equation:

\[
\frac{dX_t}{dt} = b(X_t,t)dt + \sigma(X_t,t)dW_t.
\]

(1)

Remember that this is short-hand for an integral equation

\[
X_t = \int_0^t b(X_s,s)ds + \sigma(X_s,s)dW_s.
\]

Here are some questions we will ask:

- When do solutions exist? Are they unique?
- What are the solutions’ properties as stochastic processes?
- How can we actually solve them, and extract useful information?

Note that in the physics literature, you will usually see (1) written as

\[
\frac{dx}{dt} = b(x,t) + \sigma(x,t)\eta,
\]

where \(\eta = \frac{dW}{dt}\) is defined to be a “white noise”: a Gaussian process with mean 0 and covariance function 
\(\mathbb{E}\eta(s)\eta(t) = \delta(t-s)\).

8.1 Existence and uniqueness

**Theorem.** Given equation (1), suppose \(b \in \mathbb{R}^n\), \(\sigma \in \mathbb{R}^{n \times m}\) satisfy global Lipschitz and linear growth conditions:

\[
|b(x,t) - b(y,t)| + |\sigma(x,t) - \sigma(y,t)| \leq K|x - y|
\]

\[
|b(x,t)| + |\sigma(x,t)| \leq K(1 + |x|)
\]

for all \(x, y \in \mathbb{R}^n\), \(t \in [0, T]\), where \(K > 0\) is a constant. Assume the initial value \(X_0 = \xi\) is a random variable with \(\mathbb{E}\xi^2 < \infty\) and which is independent of \(\mathcal{F}_0 = \bigcup_{t \geq 0}\mathcal{F}_t\).

Then (1) is solved by a unique \(X_t \in \mathcal{V}([0,T])\), and \(X_t\) can be chosen to be continuous in \(t\) almost surely.
Notes

- The conditions in the theorem are similar to those for ODEs.
- Uniqueness requires only Lipschitz continuity.

**Example.** Consider
\[ dx_t = 3x_t^{1/3} dt + 3x_t^{2/3} dW_t, \quad x_0 = 0. \]
This has (at least) two solutions: \( x_t = 0, \) \( x_t = W_t^3. \) But the coefficients of the SDE are not Lipschitz continuous.

- Global existence requires the coefficients to satisfy linear growth conditions. Local existence does not.

**Example.** Consider
\[ dx_t = x_t^2 dt, \quad x_0 = x_0. \]
The solution is \( x_t = \frac{1}{\sqrt{1-x_0 t^3}}. \) This blows up at \( t = \frac{1}{x_0}. \)

Proof (Uniqueness, 1d). (from E et al. [2014], 7.3) Uniqueness means: if \( X_1, X_2 \) are two solutions, then \( P(\omega(t) = X_2(t, \omega) \text{ for all } t) = 1. \)

Let’s show this, when \( x \in R. \) Let \( X_t, \hat{X}_t \in \mathcal{F}([0, T]) \) be two solutions with the same initial value \( X_0. \) Then
\[
X_t - \hat{X}_t = \int_0^t (b(X_s, s) - b(\hat{X}_s, s)) \, ds + \int_0^t (\sigma(X_s, s) - \sigma(\hat{X}_s, s)) \, dW_s
\]
\[
\Rightarrow \quad E|X_t - \hat{X}_t|^2 \leq 2E \left( \int_0^t (b(X_s, s) - b(\hat{X}_s, s)) \, ds \right)^2 + 2E \left( \int_0^t (\sigma(X_s, s) - \sigma(\hat{X}_s, s)) \, dW_s \right)^2
\]
\[
\leq 2K^2 t \int_0^t E|X_s - \hat{X}_s|^2 \, ds + 2K^2 \int_0^t E|X_s|^2 \, ds
\]
\[
= 2K^2 (t + 1) \int_0^t E|X_s - \hat{X}_s|^2 \, ds.
\]
In the third line, the first term comes from Cauchy Schwarz\(^1\) and the Lipschitz continuity of \( b, \) and the second from the Itô isometry and the Lipschitz continuity of \( \sigma. \)

Therefore \( E|X_t - \hat{X}_t|^2 = 0 \) for all \( t \geq 0, \) by Gronwall’s inequality\(^2\).

Therefore \( X_t(\omega) = \hat{X}_t(\omega) \) a.s., for each fixed \( t. \) To show this holds for all \( t \) simultaneously (i.e. the whole trajectory is equal, except for \( \omega \) in a set of measure 0), note that we can extend the equality to a countable set of \( t \)-values, because at each step the “bad” \( \omega \)-values form a measure-0 set, and a countable union of measure-0 sets is measure 0. Therefore \( P(X_t = \hat{X}_t \forall t \in Q \cap [0, T]) = 1. \) Because a solution to an SDE can be chosen to depend continuously on \( t \) with probability 1, it follows that \( P(X_t = \hat{X}_t \forall t \in [0, T]) = 1. \)

Proof (Existence). (based on Oksendal [2005], E et al. [2014])

We only sketch the arguments here, for the full proof references.

\(^1\)Cauchy-Schwartz: \( \left( \int_0^t (b(X_s, s) - b(\hat{X}_s, s)) \, ds \right)^2 \leq \left( \int_0^t 1 \, ds \right) \left( \int_0^t (b(X_s, s) - b(\hat{X}_s, s))^2 \, ds \right). \)

\(^2\)Gronwall’s inequality: suppose \( f(t) \leq a + b \int_0^t f(s) \, ds \), and \( b \geq 0. \) Then \( f(t) \leq ae^{bt}. \)
This is based on Picard iteration, as for the typical ODE existence proof.

Let \( X_0^{(0)} = \xi \), and let
\[
X_t^{(k+1)} = X_0 + \int_0^t b(X_s^{(k)}, s)ds + \int_0^t \sigma(X_s^{(k)}, s)dW_s.
\]

Similar calculations as in the proof of uniqueness show that
\[
\mathbb{E}|X_t^{(k+1)} - X_t^{(k)}|^2 \leq 2K^2(t + 1) \int_0^t \mathbb{E}|X_s^{(k)} - X_s^{(k-1)}|^2 ds.
\]

For \( k = 0 \), this gives \( \mathbb{E}|X_t^{(1)} - X_t^{(0)}|^2 \leq Ct \), where \( C \) depends on \( K, T, \mathbb{E}\xi^2 \). By induction (integrate \( k \) times), this gives \( \mathbb{E}|X_t^{(k+1)} - X_t^{(k)}|^2 \leq \frac{(Lt)^{k+1}}{(k+1)!} \), where \( L \) depends on \( K, T, \mathbb{E}\xi^2 \).

Putting these together shows that
\[
\|X_t^{(m)} - X_t^{(n)}\|_{L^2([0, T] \times \Omega)}^2 \leq \sum_{k=n}^{m-1} \|X_t^{(k+1)} - X_t^{(k)}\|^2 = \sum_{k=n}^{m-1} \mathbb{E}\int_0^T |X_t^{(k+1)} - X_t^{(k)}| dt
\]
\[
\leq \sum_{k=n}^{m} \int_0^T (Lt)^{k+1} (k+1)! dt = \sum_{k=n}^{m} \frac{(Lt)^{k+2}}{(k+2)!} \to 0.
\]

Therefore \( \{X_t^{(k)}\} \) is Cauchy in \( L^2([0, T] \times \Omega) \), so it converges in \( L^2([0, T] \times \Omega) \). Let the limit be \( X_t \). This is \( \mathcal{F}_t \)-measurable for all \( t \), since each \( X_t^{(k)} \) is. Now let’s show it satisfies the SDE. We do this by showing each of the integrals converges.

The \( dt \)-term: we have \( \int_0^t b(X_s, s)ds \to \int_0^t b(X_s, s)ds \) in \( L^2(\Omega) \), by Hölder inequality.\(^3\)

The \( dW \)-term: \( \int_0^t \sigma(X_s^{(n)}, s)dW_s \to \int_0^t \sigma(X_s, s)dW_s \) in \( L^2(\Omega) \), by Itô isometry.

Therefore \( X_t \) solves the integral version of the SDE: \( X_t = X_0 + \int_0^t b(X_s, s)ds + \int_0^t \sigma(X_s, s)dW_s \). There is a continuous version \( \tilde{X}_t \) of the RHS of the SDE. By uniqueness, \( X_t = \tilde{X}_t \) a.s..

\[
8.2 \quad \text{Properties of solutions}
\]

Suppose \( X_t \) solves (1) a.s., where the coefficients satisfy the conditions given in the existence and uniqueness theorem. Such a process is called a diffusion process. To be a diffusion, it is important that the coefficients of (1) depend only on \( X_t, t \) – not be general non-anticipating functions \( f(\omega, t) \).

It can be shown that such an \( X_t \) has the following properties:

(i) It has continuous paths, almost surely.

(ii) It is a Markov process (specifically, the family of processes \( X_t^{(x)} \) with initial condition \( x \) forms a Markov family.)

For details on the Markov property, see Koralov and Sinai [2010] section 21.4.

\(^3\)Let \( p, q \in [1, \infty] \) with \( 1/p + 1/q = 1 \). Then \( \|f\|_1 \leq \|f\|_p \|x\|_q \).
8.3 Examples of SDEs

Example (Ornstein-Uhlenbeck process). Consider the equation

\[ dX_t = -aX_t dt + \sigma dW_t, \quad X_0 = \xi. \tag{2} \]

The first term is an exponential decay to 0, while the second represents fluctuations about this.

Here are some examples of simulated trajectories, with \( \alpha = \sigma = 0 \) and initial condition \( \xi = 0 \):

Here are some examples with initial condition \( \xi = 5 \):

Let’s solve explicitly for the solution. Multiply both sides by \( e^{at} \) and integrate:

\[
\int d(e^{at}X_t) = e^{at}X_t - X_0 = \int_0^t \sigma e^{as} dW_s
\]

Therefore the solution is

\[
X_t = e^{-at}X_0 + \sigma \int_0^t e^{-a(t-s)} dW_s
\]

Note the following properties of the solution:

- Term (a) shows the initial condition is “forgotten” exponentially quickly.
- Term (b) is almost the convolution of an exponential kernel with Brownian motion: \( e^{-at} * W_t \) – except it only looks at values in the past.
- \( X_t \) is Gaussian if \( X_0 \) is Gaussian, since term (b) is Gaussian.
- Calculate the moments:
  - Mean: \( \mathbb{E}X_t = e^{-at} \mathbb{E}X_0 \). The mean exponentially decays to 0.
Covariance (assuming independence of $X_0, W_t$):

$$
E[X_s X_t] - E[X_s] E[X_t] = e^{-as} e^{-at} (E[X_0^2] - (E[X_0])^2) + \sigma \int_0^a \int_0^t e^{-a(t-u)} e^{-a(s-u)} \delta(u-v) du dv
$$

$$
= e^{-a(s+t)} \text{Var}(X_0) + \sigma^2 \int_0^{s/t} e^{-a(s-u)} e^{-a(t-u)} du
$$

$$
= e^{-a(s+t)} \text{Var}(X_0) + \frac{\sigma^2}{2a} e^{-a(s+t)} \left( e^{2a(s/t)} - 1 \right)
$$

$$
= e^{-a(s+t)} \text{Var}(X_0) + \frac{\sigma^2}{2a} \left( e^{-a|s-t|} - e^{-a(s+t)} \right)
$$

- This shows that as $t \to \infty$, $\text{Var}(X_t) \to \frac{\sigma^2}{2a}$.

- Consider taking initial condition $X_0 \sim N(0, \frac{\sigma^2}{2a})$. Then the mean is $E[X_t] = 0$, and the covariance is $\text{Cov}(X_s, X_t) = \frac{\sigma^2}{2a} e^{-a|s-t|}$.

This only depends on the time difference $s-t$, so $X_t$ is weakly stationary. Since it is Gaussian, it is also strongly stationary.

- This also implies that $N(0, \frac{\sigma^2}{2a})$ is the stationary distribution for the process. Here we found it by a clever guess – later, we will see a systematic way to find stationary distributions.

- It turns out this process is the only stationary, Markov, Gaussian random process with continuous paths. It comes up in a LOT of models.

**Example** (Geometric Brownian Motion). Consider the equation

$$
dN_t = rN_t dt + \alpha N_t dW_t
$$

(3)

This is a model for population growth, with $r =$ growth rate, $\alpha =$ fluctuations. It’s also a model for the stock market, with $N_t =$ price of asset, $r =$ interest rate, $\alpha =$ volatility. The difference from the OU process is in the diffusion term – now the diffusion rate depends on the process itself. This is called *multiplicative* noise, and is fundamentally different from constant noise.

To solve: first, divide by $N_t$:

$$
\frac{dN_t}{N_t} = r dt + \alpha dW_t
$$

We have:

$$
d(\log N_t) = \frac{1}{N_t} dN_t - \frac{(dN_t)^2}{2N_t^2} = \frac{1}{N_t} dN_t - \frac{\alpha^2}{2} dt
$$

$$
\Rightarrow \quad d(\log N_t) = (r - \frac{\alpha^2}{2}) dt + \alpha dW_t
$$

The solution is

$$
N_t = N_0 e^{(r - \frac{\alpha^2}{2})t + \alpha W_t}
$$

Let’s look at properties of this solution.
• Mean: to find this, we’ll need to find the mean of $Y_t \equiv e^{\alpha W_t}$. Let’s express this as an Itô integral, so we can use the non-anticipating property.

\[
\begin{align*}
  dY_t &= \alpha e^{\alpha W_t} dW_t + \frac{\alpha^2}{2} e^{\alpha W_t} dt \\
  \Rightarrow Y_t &= Y_0 + \alpha \int_0^t e^{\alpha W_s} dW_s + \frac{\alpha^2}{2} \int_0^t e^{\alpha W_s} ds \\
  \Rightarrow \quad \mathbb{E}Y_t &= \mathbb{E}Y_0 + \frac{\alpha^2}{2} \int_0^t \mathbb{E}Y_s ds \\
  \Rightarrow \quad \frac{d}{dt} \mathbb{E}Y_t &= \alpha^2 \mathbb{E}Y_t, \quad \mathbb{E}Y_0 = 1 \\
  \Rightarrow \quad \mathbb{E}Y_t &= e^{\frac{\alpha^2 t}{2}}
\end{align*}
\]

Therefore $\mathbb{E}N_t = (\mathbb{E}N_0)e^{rt}$ The expected value grows with the average rate.

• Note that if we had used the Stratonovich interpretation of (3), the solution would be $N_t = N_0e^{rt+\alpha W_t}$ (see section 8.4), so $\mathbb{E}N_t = (\mathbb{E}N_0)e^{(r+\frac{\alpha^2}{2})t}$. The two results are different, as is always the case for multiplicative noise.

• What happens to the trajectories? Do they also increase a.s. with the average rate?

By the Law of the Iterated Logarithm (see Lecture 5), the supremum of $e^{\alpha W_t}$ grows as $e^{\alpha \sqrt{2 \log \log t}}$ as $t \to \infty$, which is slower than linear in the exponential. Therefore the trajectory behaviour depends on $r - \frac{\alpha^2}{2}$:

- If $r > \frac{\alpha^2}{2}$, then $N_t \to \infty$ a.s. as $t \to \infty$.
- If $r < \frac{\alpha^2}{2}$, then $N_t \to 0$ a.s..
- If $r = \frac{\alpha^2}{2}$, then $N_t$ will fluctuate between arbitrarily large and small values, a.s..

Therefore, if $0 < r < \frac{\alpha^2}{2}$, then $\mathbb{E}N_t \to \infty$ while $N_t \to 0$ a.s. If $r = 0$ (no growth, just multiplicative noise), then $\mathbb{E}N_t = \text{constant}$, while $N_t \to 0$ a.s.. This apparent paradox arises because increasingly large (but rare) fluctuations will dominate the expectation.

Food for thought: would you expect this process to be ergodic? (In the sense that time average = probability average)

**Example** (Stochastically forced harmonic oscillator). Consider the following equation, written in physicists’ notation:

\[
\frac{d^2 X}{dt^2} + k^2 X + \gamma \frac{dX}{dt} = \sigma \eta(t). \quad (4)
\]

Here $k^2$ is the spring constant of the oscillator, $\gamma$ is the damping rate, and $\sigma$ is the strength of the noise. All constants have been scaled by mass.

We know that when a deterministic oscillator is forced periodically, and the frequency of the forcing does not equal the resonant frequency of the oscillator, then the oscillations will be bounded, and usually quite small – if you pump your legs on a swing too quickly or too slowly, the swing doesn’t move very much.
However, when the frequency of the forcing exactly equals the resonant frequency, the oscillations will grow without bound – you could make the swing go all the way around, in principle.

What happens when the forcing is stochastic? If you pump your legs completely at random, will you swing? We will answer this by solving explicitly for a solution.

We can write this as as SDE by letting $V = \frac{dX}{dt}$:

$$
\begin{align*}
    dX_t &= V_t dt \\
    dV_t &= (-k^2 X_t - \gamma V_t) dt + \sigma dW_2
\end{align*}
\implies d\begin{pmatrix} X_t \\ V_t \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -k^2 & -\gamma \end{pmatrix} \begin{pmatrix} X_t \\ V_t \end{pmatrix} dt + \begin{pmatrix} 0 & 0 \\ 0 & \sigma \end{pmatrix} dW_t.
$$

This has the form

$$
dU_t = -AU_t dt + BdW_t,
$$

where $U = \begin{pmatrix} X_t \\ V_t \end{pmatrix}$, $W_t = \begin{pmatrix} W_1 \\ W_2 \end{pmatrix}$, and $A, B$ are constant matrices. This is a 2-dimensional Ornstein-Uhlenbeck process. We can solve it in the same way as the first example:

$$
d(e^{At}U_t) = e^{At}dU_t + Ae^{At}U_t
= e^{At}(-AU_t dt + BdW_t) + Ae^{At}U_t
= e^{At}BdW_t
\implies U_t = e^{-At}U_0 + \int_0^t e^{-A(t-s)}BdW_s.
$$

Consider cases.

- Case $\gamma = 0$ (no damping). Then $A = \begin{pmatrix} 0 & -1 \\ k^2 & 0 \end{pmatrix}$, which has eigenvalues $\lambda = \pm ik$ and corresponding eigenvectors $u_1 = (1, -ik)^T, u_2 = (1, ik)^T$. Then $e^{At} = U e^{Dt U^{-1}}$, with $U = (u_1 \ u_2), D = \begin{pmatrix} ik & 0 \\ 0 & -ik \end{pmatrix}$.

Therefore $e^{At} = \begin{pmatrix} \cos kt & -\frac{1}{k} \sin kt \\ -k \sin kt & \cos kt \end{pmatrix}$.

Consider the solution with initial condition $X_0 = 0, X_0' = 0$:

$$
X_t = \frac{\sigma}{k} \int_0^t \sin k(s-t) dW_2.
$$

This represents oscillations that grow without bound. To see this, calculate

$$
\begin{align*}
    \mathbb{E}X_t^2 &= \frac{\sigma^2}{k^2} \int_0^t (\sin k(s-t))^2 ds \\
    &= \frac{\sigma^2}{k^2} \left( \frac{t}{2} - \frac{\sin 2kt}{4k} \right) \\
    &\sim \frac{\sigma^2}{k^2} \frac{t}{2} \rightarrow \infty \quad \text{as } t \rightarrow \infty.
\end{align*}
$$
This means that a stochastically forced swingset will swing! So if you pump your legs completely at random, you can make the swing go as high as you want, in principle. It is somewhat remarkable that although you are forcing all frequencies equally, and almost all of these are non-resonant, you still have enough forcing near the resonant frequency to make the oscillations grow.

- Case $\gamma \neq 0$ (with damping.) In this case the eigenvalues of $A$ have a real and imaginary part. The real part is negative and represents exponential damping. The imaginary part represents oscillations, with a frequency that is slightly shifted from $k$. It is not hard to write down the solution explicitly (ELFS.) To understand the frequency components we can look for a stationary solution $X$. If such a solution exists, then it has a spectral representation so we can look for the spectral measure $dZ_X(\lambda)$. Formally we can look at \ref{eq:4} in spectral space, or take its “Fourier transform,” to find

$$-\lambda^2 \hat{X}_\lambda + k^2 \hat{X}_\lambda + i\lambda \gamma \hat{X}_\lambda = \sigma \hat{\eta},$$

where $\hat{X}_\lambda$ stands for $dZ_X(\lambda)$, and $\hat{\eta}$ is shorthand for the spectral measure of $\eta$. Using $C_\eta(t) = \mathbb{E} \eta(t) \eta(0) = \delta(t)$, we find (formally) $|\hat{\eta}_\lambda|^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\lambda t} \delta(t) dt = \frac{1}{2\pi}$.

Solving gives

$$\hat{X}_\lambda = \frac{\sigma \hat{\eta}}{-\lambda^2 + k^2 + i\gamma} \quad \Rightarrow \quad \mathbb{E} |\hat{X}_\lambda|^2 = \frac{\sigma^2}{2\pi (\lambda^2 - k^2)^2 + \lambda^2 \gamma^2} S(\lambda).$$

Therefore if a stationary solution exists, its spectral density is $S(\lambda)$ given above. Note this is only normalizable when $\gamma \neq 0$, so we only expect a stationary solution in this case.

The main frequency components is at $\sqrt{k^2 - \frac{\gamma^2}{2}}$, which is close to (but slightly less than) the resonant frequency $k$. The swing swings, but it’s bounded.

### 8.4 Stratonovich Integral

Suppose we want to work with the Stratonovich integral. Can this be related to the Itô integral? Yes, by the following formula.

\footnote{Note that we haven’t shown it does so almost surely, only that the variance increases without bound – you can try to think about why the oscillations will grow without bound a.s., using the properties of Brownian motion.}

\footnote{To make the calculations slightly less formal: consider writing each of the functions in \ref{eq:4} in spectral form, as in $X_t = \int e^{i\lambda t} dZ_X(\lambda)$, $X'(t) = \int i\lambda e^{i\lambda t} dZ(\lambda)$, etc. Then the equation becomes $\int (-\lambda^2 + k^2 + i\lambda \gamma) e^{i\lambda t} dZ_X(\lambda) = \int \sigma e^{i\lambda t} dZ_\eta(\lambda)$, where $dZ_\eta(\lambda)$ is the spectral measure for $\eta$. Take the inverse transform and solve to show that $dZ_X(\lambda) = \frac{\sigma}{-\lambda^2 + k^2 + i\gamma} dZ_\eta(\lambda)$. Now use that $\mathbb{E} dZ_X(\lambda) dZ_X(\lambda') = S(\lambda) \delta(\lambda - \lambda') d\lambda d\lambda'$.}
**Theorem.** Suppose $X_t$ solves the Stratonovich equation

$$dX_t = b(t, X_t) dt + \sigma(t, X_t) \circ dW_t.$$  

Then $X_t$ also solves the Itô equation

$$dX_t = \left( b(t, X_t) + \frac{1}{2} \sigma_x(t, X_t) \right) dt + \sigma(t, X_t) dW_t.$$  

**Proof.** (Outline, from E et al., 2014) Write $\sigma(t, X_t) \circ dW_t$ as a limit of sums:

$$\sigma(t, X_t) \circ dW_t = \lim_{\Delta \to 0} \sum_j \left[ \frac{\sigma(t_j, X_{t_j}) + \sigma(t_{j+1}, X_{t_{j+1}})}{2} \right] (W_{t_{j+1}} - W_j)$$

Suppose that $X_t$ solves an Itô SDE of the form

$$dX_t = \alpha(t, X_t) dt + \beta(t, X_t) dW_t,$$  

for some functions $\alpha, \beta$. Let’s substitute the approximation $X_{t_{j+1}} \approx X_{t_j} + \alpha(t_j, X_{t_j}) \Delta t + \beta(t_j, X_{t_j}) \Delta W_{t_j}$ into the limit above. We have

$$\sum_j \sigma(t_{j+1}, X_{t_{j+1}}) \Delta W_{t_j} \approx \sum_j \left( \sigma(t_j, X_{t_j}) \Delta W_{t_j} + \partial_t \sigma(t_j, X_{t_j}) \Delta t + \partial_x \sigma(t_j, X_{t_j}) \Delta W_{t_j} + \frac{1}{2} \sigma_{xx} \beta(t_j, X_{t_j}) (\Delta W_{t_j})^2 \right)$$

$$\xrightarrow{m \to \infty} \int_0^t \sigma(s, X_s) dW_s + \int_0^t \partial_t \sigma \beta(s, X_s) ds.$$  

Therefore $X_t$ satisfies

$$dX_t = \left( b(t, X_t) + \frac{1}{2} \partial_x \sigma \beta(t, X_t) \right) dt + \sigma(t, X_t) dW_t.$$  

In order to match (5), we need to choose $\beta = \sigma$, $\alpha = b + \frac{1}{2} \partial_t \sigma \sigma$.

**Note.** To make the proof rigorous, one needs to rigorously control the error terms. This is left to the reader to fill in.

In multiple dimensions the conversion is given by

**Theorem.** Given $X_t, W_t \in \mathbb{R}^n$, $\sigma \in \mathbb{R}^{n \times n}$. The Stratonovich equation

$$dX_t = b(t, X_t) dt + \sigma(t, X_t) dW_t$$

is equivalent to the Itô equation

$$dX_t = \left( b(t, X_t) + \frac{1}{2} \nabla_x \sigma : \sigma(t, X_t) \right) dt + \sigma(t, X_t) dW_t.$$  

Here $(\nabla_x \sigma : \sigma)_i = \sum_{j,k} (\partial_k \sigma_{ik}) \alpha_{kj}$.

**Notes**

- Why use the Stratonovich integral? There are several advantages:
The regular chain rule holds, i.e. \( df(X_t) = f'(X_t) \circ dX_t \). (ELFS on HW, or see Pavliotis [2014].)

If you start with non-white, smooth noise in your ODE, and take a limit to make the noise white, you get a Stratonovich integral of the noise. (We’ll see this in the lecture on asymptotics.)

If you restrict your process to lie on a submanifold of \( \mathbb{R}^n \), the most natural way to do this is through the Stratonovich integral. For example, “Brownian motion” on the surface of a sphere is the solution to

\[
\frac{dX_t}{dt} = P(X_t) \circ dB_t,
\]

where \( B_t \in \mathbb{R}^3 \) is a 3-dimensional BM and \( P(X_t) \) is the orthogonal projection matrix onto the tangent space to the surface of the unit sphere.

- What are the disadvantages?
  - The Itô Isometry is lost.
  - The property \( \mathbb{E} \int_0^t \sigma(t, \omega) \circ dW_t = 0 \) no longer holds, i.e. the Stratonovich integral “looks into the future.”

- Does it matter which integral you use? Not really – you can always convert from one to the other.

**References**


