

## Lecture 8: Stochastic Differential Equations

### Readings

Recommended:

- Pavliotis (2014) 3.2-3.5
- Oksendal (2005) Ch. 5

Optional:

- Gardiner (2009) 4.3-4.5
- Oksendal (2005) 7.1,7.2 (on Markov property)
- Korolov and Sinai (2010) 21.4 (on Markov property)

In this lecture we will study stochastic differential equations (SDEs), which have the form

$$dX_t = b(X_t, t)dt + \sigma(X_t, t)dW_t, \quad X_0 = \xi \quad (1)$$

where  $X_t, b \in \mathbb{R}^n$ ,  $\sigma \in \mathbb{R}^{n \times n}$ , and  $W$  is an  $n$ -dimensional Brownian motion. The initial condition  $\xi$  is assumed independent of  $W$ . We write the solution as  $X = (X_t)_{t \geq 0}$ . An SDE may equivalently be written as

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

where  $\eta(t)$  is a white noise: a stationary Gaussian process with mean 0 and covariance function  $\mathbb{E}\eta(s)\eta(t) = \delta(t-s)$ . Recall that (1) is short-hand for the integral equation

$$X_t = \int_0^t b(X_s, s)ds + \int_0^t \sigma(X_s, s)dW_s + \xi. \quad (2)$$

Each term in (1) has a different interpretation.

- The term  $b(X_t, t)dt$  is called the *drift* term. It describes the deterministic part of the equation. When this is the only term, we obtain an ODE.
- The term  $\sigma(X_t, t)dW_t$  is called the *diffusion* term. It describes random motion proportional to a Brownian motion. Over small times, this term causes the probability to spread out diffusively with a diffusivity locally proportional to  $\sigma^2$ .

Often we will consider time-homogeneous SDEs, where  $b, \sigma$  only depend on  $x$ . Any time-inhomogeneous SDE can be converted to a time-homogeneous one by introducing an additional variable  $Y = t$ .

If the diffusion term is constant in  $x$ , i.e.  $\frac{\partial \sigma}{\partial x} = 0$ , then the noise is said to be *additive*. If the diffusion term depends on  $x$ ,  $\frac{\partial \sigma}{\partial x} \neq 0$ , the noise is said to be *multiplicative*. Equations with multiplicative noise have to be treated more carefully than equations with additive noise.

We learned how to define the integrals in the expressions above last class. In this one we'll look at properties of the solutions themselves. We will ask: when do solutions exist? Are they unique? And how can we actually solve them, to extract useful information?

## 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:

$$\begin{aligned} |b(x,t) - b(y,t)| + |\sigma(x,t) - \sigma(y,t)| &\leq K|x - y| \\ |b(x,t)| + |\sigma(x,t)| &\leq K(1 + |x|) \end{aligned}$$

for all  $x, y \in \mathbb{R}^n$ ,  $t \in [0, T]$ , and some constant  $K > 0$ . Assume the initial value  $X_0 = \xi$  is a random variable with  $\mathbb{E}\xi^2 < \infty$  and which is independent of  $W$ . Then (1) has a unique solution  $X$ , such that  $X$  is continuous with probability 1,  $X$  is adapted<sup>1</sup> to  $W$ , and  $\mathbb{E} \int_0^T |X_t|^2 dt < \infty$ .

“Unique” means that if  $X^{(1)}, X^{(2)}$  are two strong solutions, then  $P(X^{(1)}(t, \omega) = X^{(2)}(t, \omega) \text{ for all } t) = 1$ . That is, the two solutions are equal everywhere with probability 1. This is different from the statement that  $X^{(1)}, X^{(2)}$  are versions of each other – you should think about how.

*Remark.* The global Lipschitz condition and linear growth condition ask for constants  $K$  that are independent of  $t$ . If  $b, \sigma$  are functions of  $x$  only, then the Lipschitz condition implies the linear growth condition. When  $b, \sigma$  are also functions of  $t$ , the Lipschitz condition implies the linear growth condition only with additional assumptions on how  $b, \sigma$  behave with  $t$  – for example, if they are continuous in  $t$ , or bounded in  $t$ .

This theorem bears a lot in common with similar theorems regarding the existence and uniqueness to the solution to an ODE. Counterexamples that show the necessity of each of the conditions of the theorem that apply to ODEs, can also be used for SDEs.

**Example 8.1** To construct an equation whose solution is not unique, we drop the condition of Lipschitz continuity. Consider the ODE

$$dX_t = 3X_t^{2/3} dt, \quad X_0 = 0,$$

which has solutions  $X_t = \begin{cases} 0 & t \leq a, \\ (t-a)^3 & t > a \end{cases}$ , for any  $a > 0$ . This doesn’t violate the theorem because  $b(x) = 3x^{2/3}$  is not Lipschitz continuous at 0. For a similar example involving a Brownian motion, consider

$$dX_t = 3X_t^{1/3} dt + 3X_t^{2/3} dW_t, \quad X_0 = 0.$$

This has (at least) two solutions:  $X_t = 0$ , and  $X_t = W_t^3$ .

**Example 8.2** To construct an equation which has no global solution, we drop the linear growth conditions. Consider

$$dX_t = X_t^2 dt, \quad X_0 = x_0.$$

The solution is  $X_t = \frac{1}{\frac{1}{x_0} - t}$ , which blows up at  $t = \frac{1}{x_0}$ .

*Remark.* In the example above, the drift  $b(x) = x^2$  is not globally Lipschitz continuous. It is locally Lipschitz continuous, however, which is sufficient to show uniqueness (Karatzas and Shreve (1991), Section 5 Theorem 2.5).

<sup>1</sup>Actually we ask for something slightly stronger, namely that  $X$  be progressively measurable with respect to  $\mathcal{F}$ , the filtration generated by  $(W_t)_{t \geq 0}$

*Proof (Uniqueness).* (Evans (2013), Section 5.B.3) Let  $X, \hat{X}$  be two strong solutions to (1). Then for  $0 \leq t \leq T$ ,

$$X_t - \hat{X}_t = \int_0^t (b(X_s, s) - b(\hat{X}_s, s)) ds + \int_0^t (\sigma(X_s, s) - \sigma(\hat{X}_s, s)) dW_s.$$

Square each side, use  $(a + b)^2 \leq 2a^2 + 2b^2$ , and take expectations to get

$$\mathbb{E}|X_t - \hat{X}_t|^2 \leq 2\mathbb{E}\left|\int_0^t (b(X_s, s) - b(\hat{X}_s, s)) ds\right|^2 + 2\mathbb{E}\left|\int_0^t (\sigma(X_s, s) - \sigma(\hat{X}_s, s)) dW_s\right|^2.$$

We estimate the first term on the right-hand side using the Cauchy-Schwarz inequality  $|\int_0^t f ds|^2 \leq t \int_0^t |f|^2 ds$ . We then use the Lipschitz continuity of  $b$ . The result is

$$\mathbb{E}\left|\int_0^t (b(X_s, s) - b(\hat{X}_s, s)) ds\right|^2 \leq T\mathbb{E}\int_0^t |b(X_s, s) - b(\hat{X}_s, s)|^2 ds \leq K^2 T \int_0^t \mathbb{E}|X_s - \hat{X}_s|^2 ds.$$

Now we estimate the second term using the Itô isometry and the Lipschitz continuity of  $\sigma$ :

$$\mathbb{E}\left|\int_0^t (\sigma(X_s, s) - \sigma(\hat{X}_s, s)) dW_s\right|^2 = \int_0^t \mathbb{E}|\sigma(X_s, s) - \sigma(\hat{X}_s, s)|^2 ds \leq K^2 \int_0^t \mathbb{E}|X_s - \hat{X}_s|^2 ds.$$

Putting these estimates together shows that

$$\mathbb{E}|X_t - \hat{X}_t|^2 \leq C \int_0^t \mathbb{E}|X_s - \hat{X}_s|^2 ds$$

for some constant  $C$ , for  $0 \leq t \leq T$ . Now we can use Gronwall's Inequality, which says that if we are given a function  $f$  and nonnegative numbers  $a, b \geq 0$  such that

$$f(t) \leq a + b \int_0^t f(s) ds, \quad \text{then} \quad f(t) \leq ae^{bt}.$$

The proof is given in the appendix. Applying Gronwall's Inequality with  $f(t) = \mathbb{E}|X_t - \hat{X}_t|^2$  and  $a = 0, b = C$  shows that

$$\mathbb{E}|X_t - \hat{X}_t|^2 = 0 \quad \text{for all } 0 \leq t \leq T.$$

Therefore for each fixed  $t \in [0, T]$  we have that  $X_t = \hat{X}_t$  a.s., i.e.  $P(\{\omega \in \Omega : X_t(\omega) \neq \hat{X}_t(\omega)\}) = 0$ . It remains to show that  $X_t = \hat{X}_t$  for all  $t$  simultaneously, except for  $\omega$  in a set of measure 0.

We can argue that  $X_r = \hat{X}_r$  for all rational  $0 \leq r \leq T$ , almost surely. This is because for any countable set of  $t$ -values, such as an enumeration of the rationals  $\{t_1, t_2, \dots\}$ , we have

$$P(\{\omega : X_t(\omega) \neq \hat{X}_t(\omega) \exists t \in \{t_1, t_2, \dots\}\}) = \bigcup_i P(\{\omega : X_{t_i}(\omega) \neq \hat{X}_{t_i}(\omega)\}) = 0.$$

Therefore  $P(\{\omega : X_t(\omega) = \hat{X}_t(\omega) \forall t \in \{t_1, t_2, \dots\}\}) = 1$ . By assumption  $X, \hat{X}$  have continuous sample paths almost surely, so we can extend the equality to all values of  $t$  using the fact that the rationals form a dense set in  $\mathbb{R}$ , and therefore  $P(X_t = \hat{X}_t \forall t \in [0, T]) = 1$ .  $\square$

*Proof (Existence).* (Evans (2013), Section 5.B.3 and Varadhan (2007), Thm 6.1 p.88) The proof is based on Picard iteration, as for the typical ODE existence proof. For simpler notation we consider a 1-dimensional SDE but the proof is almost identical in higher dimensions. Let the 0th iterate be  $X_t^0 = \xi$ , and define the  $(n+1)$ th iterate be

$$X_t^{n+1} = X_0 + \int_0^t b(X_s^n, s) ds + \int_0^t \sigma(X_s^n, s) dW_s, \quad n = 1, 2, \dots, \quad 0 \leq t \leq T.$$

One can verify (by induction) that  $X^{n+1}$  is adapted (and progressively measurable), continuous almost surely, and has  $\sup_{0 \leq t \leq T} \mathbb{E}|X_t^{n+1}|^2 < \infty$ . This implies that  $\mathbb{E} \int_0^T |\sigma(X_s^{n+1}, s)|^2 ds < \infty$  so the next iterate is well-defined. Let the mean-squared difference between successive iterates be

$$D^n(t) = \mathbb{E}|X_t^{n+1} - X_t^n|^2.$$

We claim that

$$D^n(t) \leq \frac{(Mt)^{n+1}}{n!}$$

for some constant  $M$  depending on  $K, T, \xi$ . We prove this by induction. For  $n = 0$ ,

$$D^0(t) = \mathbb{E}|X_t^1 - X_t^0|^2 \leq 2\mathbb{E} \left| \int_0^t K(1 + |\xi|) ds \right|^2 + 2\mathbb{E} \int_0^t K^2(1 + |\xi|^2) ds \leq tM$$

for some  $M$  large enough. Next assume the claim holds for  $n - 1$ , and calculate

$$\begin{aligned} D^n(t) &= \mathbb{E} \left| \int_0^t (b(X_s^n, s) - b(X_s^{n-1}, s)) ds + \int_0^t (\sigma(X_s^n, s) - \sigma(X_s^{n-1}, s)) dW_s \right|^2 \\ &\leq 2TK^2 \mathbb{E} \int_0^t |X_s^n - X_s^{n-1}|^2 ds + 2K^2 \mathbb{E} \int_0^t |X_s^n - X_s^{n-1}|^2 ds \end{aligned}$$

following the same calculations as in the proof of uniqueness. Therefore by the induction hypothesis,

$$D^n(t) \leq 2K^2(1+T) \int_0^t \frac{M^n s^n}{n!} ds \leq \frac{(Mt)^{n+1}}{n!},$$

provided  $M \geq 2K^2(1+T)$ .

Now, using the triangle inequality on norm  $\|f\|_{L_2(P)} = (\mathbb{E}f^2)^{1/2}$  gives that

$$\left\| \sup_{0 \leq t \leq T} |X_t^{n+1} - X_t^n| \right\|_{L_2(P)} \leq \left\| \sup_{0 \leq t \leq T} |Y_t^n| \right\|_{L_2(P)} + \left\| \sup_{0 \leq t \leq T} |Z_t^n| \right\|_{L_2(P)},$$

where

$$Y_t^n = \int_0^t (\sigma(X_s^n, s) - \sigma(X_s^{n-1}, s)) dW_s, \quad Z_t^n = \int_0^t (b(X_s^n, s) - b(X_s^{n-1}, s)) ds.$$

We bound the first term using Doob's maximal inequality (a variant of Doob's martingale inequality), which says that, given a process  $M = (M_t)_{t \geq 0}$  which is a martingale<sup>2</sup> with respect to  $W$ , we have that

<sup>2</sup> $\mathbb{E}|M_t| < \infty$  and  $\mathbb{E}(M_t | (W_u)_{0 \leq u \leq s}) = M_s$  for all  $s \leq t$ .

$\| \sup_{0 \leq s \leq t} M_s \|_{L_2(P)} \leq 2 \| M_t \|_{L_2(P)}$ . Applying this inequality to  $Y_t^n$ , which can be shown to be a martingale, gives

$$\begin{aligned} \| \sup_{0 \leq t \leq T} |Y_t^n| \|_{L_2(P)} &\leq 2 \| Y_T^n \|_{L_2(P)} = 2 \left( \mathbb{E} \int_0^T |\sigma(X_s^n, s) - \sigma(X_s^{n-1}, s)|^2 ds \right)^{1/2} \\ &\leq 2K \left( \int_0^T \mathbb{E} |X_s^n - X_s^{n-1}|^2 ds \right)^{1/2} = 2K \left( \int_0^T D^{n-1}(s) ds \right)^{1/2} \\ &\leq C \sqrt{\frac{(MT)^n}{n!}} \quad \text{by the claim,} \end{aligned}$$

for some constant  $C$ . Using Cauchy-Schwartz and other familiar calculations, one can show that

$$\| \sup_{0 \leq t \leq T} |Z_t^n| \|_{L_2(P)} \leq A \left( \int_0^T D^{n-1}(s) ds \right)^{1/2} \leq C \sqrt{\frac{(MT)^n}{n!}},$$

for some constant  $A$  and then some constant  $C$ .

Therefore,

$$\sum_n \left\| \sup_{0 \leq t \leq T} |X_t^{n+1} - X_t^n| \right\|_{L_2(P)} < \infty,$$

so  $\sum_n X_t^{n+1} - X_t^n$  converges uniformly on every finite interval. Therefore  $X^n$  converges to a limit  $X$ , which is adapted (and progressively measurable), and almost surely continuous. One can check that  $X$  is a solution to (1), by passing to the limit in the integrals defining  $X^n$ . Finally, one can verify that  $\mathbb{E} \int_0^T X(t)^2 dt < \infty$  using induction to obtain  $\mathbb{E} |X^{n+1}(t)|^2 \leq C(1 + \mathbb{E} |\xi^2|) e^{Ct}$ , then the same bound for  $\mathbb{E} |X(t)|^2$  (see Evans (2013)).

□

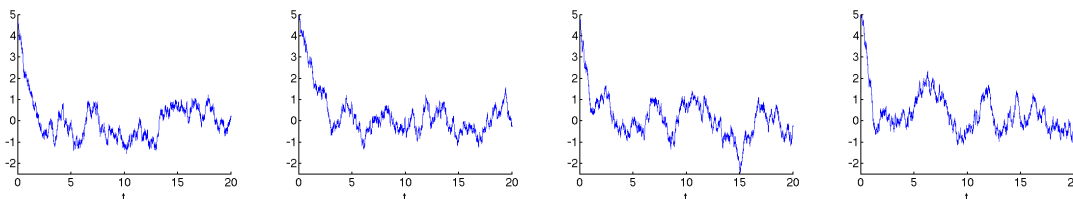
## 8.2 Examples of SDEs and their solutions

**Example 8.3** (Ornstein-Uhlenbeck process) Let  $a, \sigma \in \mathbb{R}$  be constants, and let  $X_t \in \mathbb{R}$ . The Ornstein-Uhlenbeck process (OU process) is the solution  $X$  to

$$dX_t = -aX_t dt + \sigma dW_t, \quad X_0 = \xi, \tag{3}$$

where  $\xi$  is independent of  $(W_t)_{t \geq 0}$ .

Here are some examples of simulated trajectories with  $\alpha = \sigma = 1$  and initial condition  $\xi = 5$ :



The process initially decays exponentially quickly, and when it gets close to zero it fluctuates around zero.

Let's solve explicitly for the solution. Multiply both sides by  $e^{at}$  and integrate from 0 to  $t$ :

$$\underbrace{e^{at} dX_t + ae^{at} X_t dt}_{=d(e^{at} X_t)} = \sigma e^{at} dW_t \quad \Rightarrow \quad e^{at} X_t - X_0 = \int_0^t \sigma e^{as} dW_s.$$

This gives solution

$$X_t = e^{-at} X_0 + \sigma \int_0^t e^{-a(t-s)} dW_s. \quad (4)$$

The first term shows the initial condition is “forgotten” exponentially quickly. The second term represents the stochastic fluctuations, and is similar to the convolution of an exponential kernel with Brownian motion,  $e^{-at} * W_t$ , except the convolution only looks at values in the past. This term is Gaussian, since it is a linear functional applied to a Gaussian process. Hence, if  $X_0$  is Gaussian, then  $X$  is Gaussian.

From (4) we can calculate the moments of  $X_t$ . The mean is

$$\mathbb{E}X_t = e^{-at} \mathbb{E}X_0.$$

The mean decays exponentially to 0,  $\mathbb{E}X_t \rightarrow 0$  as  $t \rightarrow \infty$ . The covariance is

$$\begin{aligned} B(s, t) &= \mathbb{E}X_s X_t - \mathbb{E}X_s \mathbb{E}X_t = e^{-as} e^{-at} (\mathbb{E}X_0^2 - (\mathbb{E}X_0)^2) + \sigma \int_0^s \int_0^t e^{-a(t-v)} e^{-a(s-u)} \underbrace{\mathbb{E}dW_u dW_v}_{=\delta(u-v)dudv} \\ &= e^{-a(s+t)} \text{Var}(X_0) + \sigma^2 \int_0^{s \wedge t} e^{-a(s-u)} e^{-a(t-u)} du \\ &= e^{-a(s+t)} \text{Var}(X_0) + \frac{\sigma^2}{2a} e^{-a(s+t)} \left( e^{2a(s \wedge t)} - 1 \right) \\ &= e^{-a(s+t)} \text{Var}(X_0) + \frac{\sigma^2}{2a} \left( e^{-a|s-t|} - e^{-a(s+t)} \right). \end{aligned}$$

If  $s, t \rightarrow \infty$  with  $s - t$  fixed, the covariance approaches  $B(s, t) \rightarrow \frac{\sigma^2}{2a} e^{-a|s-t|}$ , the covariance function for a weakly stationary process.

Now, consider what happens if we take the distribution of  $X_0$  to be normal, with the long-time mean and variance,  $X_0 \sim N(0, \frac{\sigma^2}{2a})$ . The mean and covariance at all times are

$$\mathbb{E}X_t = 0, \quad B(s, t) = \frac{\sigma^2}{2a} e^{-a|s-t|}.$$

We obtain a weakly stationary process. Since  $X$  is Gaussian, it is also strongly stationary. Since it is strongly stationary, the one-point distributions are constant with time. Therefore  $X_t \sim N(0, \frac{\sigma^2}{2a})$  for all  $t \geq 0$ . Such a distribution, which doesn't change with time, is called a *stationary distribution*. We found it here by a clever guess, but later we will learn a systematic way to find stationary distributions.

It turns out the Ornstein-Uhlenbeck process with initial condition as above is the only process that is *stationary, Markov, Gaussian*, and that has *continuous paths*. It comes up in a lot of models, since it arises from linearizing many SDEs.

**Example 8.4** (Geometric Brownian Motion) Consider the equation one-dimensional SDE

$$dN_t = rN_t dt + \alpha N_t dW_t, \quad N_0 = \xi. \quad (5)$$

This process, called a Geometric Brownian motion (GBM), models for example population growth, where  $N_t$  is the population size at time  $t$ ,  $r$  is the average growth rate, and  $\alpha dW_t$  captures fluctuations in the growth rate. It also models the stock market, where  $N_t$  is the price of an asset,  $r$  is the interest rate, and  $\alpha$  is the volatility. The difference from the OU process is in the diffusion term – now the noise is multiplicative.

To solve (5) divide by  $N_t$ :

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

It would be nice to write the LHS as  $d(\log N_t)$  but we can't, since we need to use Itô's formula. Instead, we calculate

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

Now substitute for  $dN_t$  using the equation to get

$$d(\log N_t) = \left(r - \frac{\alpha^2}{2}\right) dt + \alpha dW_t.$$

Integrate and take the exponential to find the solution as

$$N_t = N_0 e^{(r - \frac{\alpha^2}{2})t + \alpha W_t}. \quad (6)$$

Let's look at properties of this solution. First, let's calculate the mean,  $\mathbb{E}N_t$ . It is possible to calculate this from (6) (see Exercise 8.1) but a simpler calculation starts with (5) directly, writes it in integral form, takes the expectation, and uses the nonanticipating property, to obtain an integral equation,

$$\mathbb{E}N_t = \int_0^t r \mathbb{E}N_s ds + \mathbb{E}N_0.$$

This equation can be solved by taking the time derivative and solving the corresponding ODE, to obtain

$$\mathbb{E}N_t = (\mathbb{E}N_0) e^{rt}. \quad (7)$$

Therefore, the mean grows with the average growth rate.

**Exercise 8.1.** Another way to derive (7) is to find the mean of  $Y_t = e^{\alpha W_t}$ , by calculating  $dY_t$  to express  $Y_t$  as an Itô integral, and then using the non-anticipating property. Do this, to show that  $\mathbb{E}Y_t = e^{\frac{\alpha^2}{2}t}$ .

What happens to the trajectories themselves? Do they also increase with the average rate? Recall the Law of the Iterated Logarithm, which says that

$$\limsup_{t \rightarrow \infty} \frac{W_t}{\sqrt{2t \log \log t}} = 1 \quad \text{a.s.}$$

This implies that the supremum of  $e^{\alpha W_t}$  grows as  $e^{\alpha \sqrt{2t \log \log t}}$  as  $t \rightarrow \infty$ , which is slower than linear in the exponential. Therefore the trajectory behaviour depends on the deterministic growth rate  $r - \frac{\alpha^2}{2}$  in the exponential:

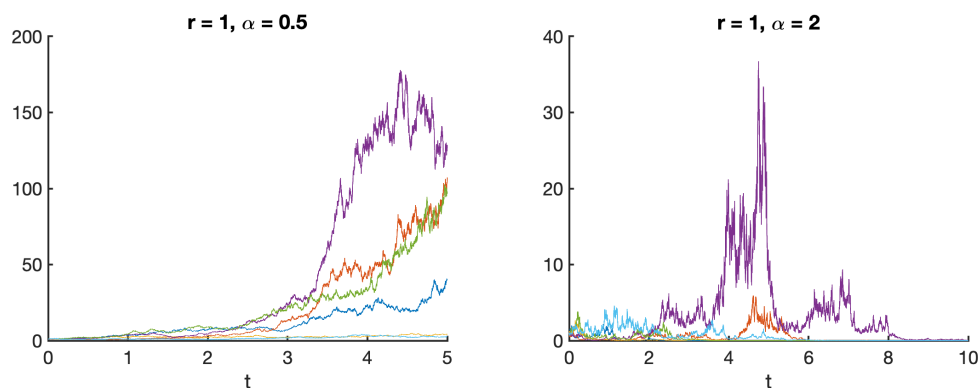


Figure 1: Trajectories of GBM with different parameters. Left:  $r > \alpha^2/2$ , so that  $N_t \rightarrow \infty$  a.s.. Right:  $r < \alpha^2/2$ , so that  $N_t \rightarrow 0$  a.s.. However,  $\mathbb{E}N_t \rightarrow \infty$  almost surely, which is possible because trajectories exhibit rare but excursions to large values. Each plot shows 6 independent trajectories.

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

Notice that the mean and the trajectory do not always behave the same way at  $\infty$ . If  $0 < r < \frac{\alpha^2}{2}$ , then  $\mathbb{E}N_t \rightarrow \infty$  while  $N_t \rightarrow 0$  a.s.! This apparent paradox arises because increasingly large (but rare) fluctuations dominate the expectation (Figure 1). It is worth pausing to think about this.

**Example 8.5** (Stochastically forced harmonic oscillator)

Consider an ODE for a forced, damped harmonic oscillator  $X$ :

$$m \frac{d^2 X_t}{dt^2} + k X_t + \gamma \frac{dX_t}{dt} = f(t). \quad (8)$$

Here  $m$  is the oscillator's mass,  $k$  is its spring constant,  $\gamma$  is a damping coefficient modeling frictional damping, and  $f(t)$  is the external forcing. The oscillator  $X$  could represent for example the angle of a pendulum, such as a swing, under small perturbations from its rest state.

From mechanics, we know that when an undamped ( $\gamma = 0$ ) harmonic oscillator is forced periodically,  $f(t) = \sin \lambda t$ , then when the forcing frequency does not equal the resonant frequency,  $\lambda \neq \tilde{k} = \sqrt{k/m}$ , 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,  $f(t) = \sin(\tilde{k}t)$ , the oscillations will grow without bound – on a frictionless swingset, you could make the swing go all the way around the swingset.

What happens when the forcing is stochastic? If you pump your legs completely at random, will you swing?

We answer this by letting  $f(t) = \sigma \eta(t)$ , where  $\eta(t)$  is a white noise, and finding the solution to (8). This



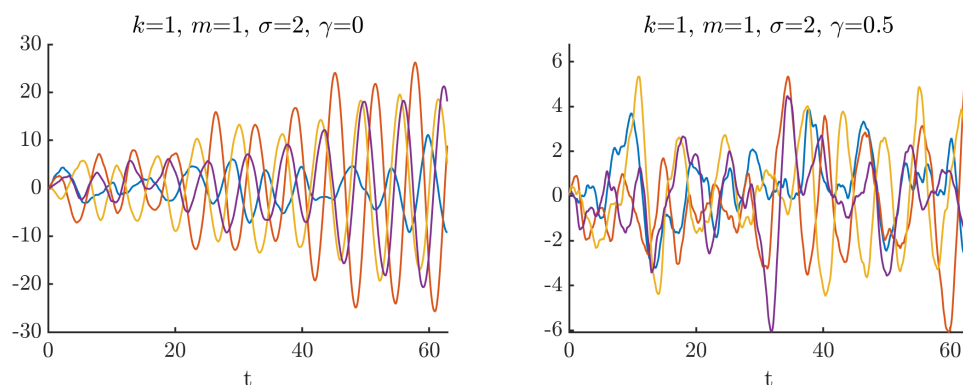


Figure 2: Trajectories of a harmonic oscillator, both without (left,  $\gamma = 0$ ) and with (right,  $\gamma \neq 0$ ) damping. Each plot shows 4 independent solutions to (8), over a time period equal to 10 natural periods ( $T = 10 \cdot 2\pi/k$ ), with parameters given in the title.

equation is a second-order ODE but we can write it as an SDE by letting  $V_t = \frac{dX_t}{dt}$ :

$$\begin{aligned} dX_t &= V_t dt \\ mdV_t &= (-kX_t - \gamma V_t)dt + \sigma dW_2 \end{aligned} \quad \Rightarrow \quad d \begin{pmatrix} X_t \\ V_t \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{\gamma}{m} \end{pmatrix}}_{-A} \begin{pmatrix} X_t \\ V_t \end{pmatrix} dt + \underbrace{\begin{pmatrix} 0 & 0 \\ 0 & \frac{\sigma}{m} \end{pmatrix}}_B \begin{pmatrix} dW_t^{(1)} \\ dW_t^{(2)} \end{pmatrix}.$$

This equation has the form

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

where  $U = \begin{pmatrix} X_t \\ V_t \end{pmatrix}$ ,  $W_t = \begin{pmatrix} W_t^{(1)} \\ W_t^{(2)} \end{pmatrix}$ , and  $A, B$  are constant matrices. This is a 2-dimensional Ornstein-Uhlenbeck process. We can solve it in the same way as in Example 8.3, using an integrating factor:

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

Consider how the solution behaves under different assumptions on the damping.

- Case  $\gamma = 0$  (no damping). Then  $A$  has eigenvalues  $\lambda = \pm i\tilde{k}$  and corresponding eigenvectors  $u_1 = (1, -i\tilde{k})^T$ ,  $u_2 = (1, i\tilde{k})^T$ . Then  $e^{At} = Ue^{Dt}U^{-1}$ , with  $U = (u_1 \ u_2)$ ,  $D = \text{diag}(i\tilde{k}, -i\tilde{k})$ . Therefore

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

If the oscillator initially starts at rest,  $X_0 = 0$ ,  $X'_0 = 0$ , then

$$X_t = \frac{\sigma}{m\tilde{k}} \int_0^t \sin \tilde{k}(s-t) dW_s^{(2)}.$$

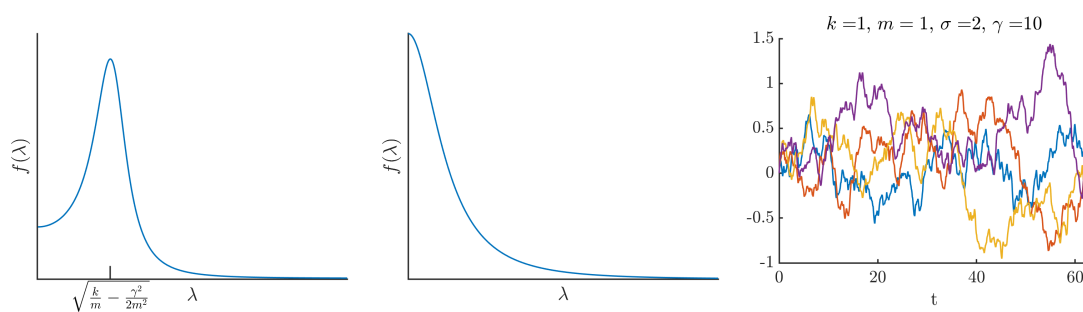


Figure 3: Spectral density for a stationary swing, with small damping ( $\gamma$  small, left), and large damping  $\gamma$  large, middle). Right: trajectories with large damping parameter.

What happens to this solution as  $t \rightarrow \infty$ ? We claim it grows without bound. To see why, calculate the variance of the solution, observing first that  $\mathbb{E}X_t = 0$  by the nonanticipating property of the Itô integral. The variance therefore is

$$\begin{aligned} \mathbb{E}X_t^2 &= \frac{\sigma}{m\tilde{k}} \int_0^t (\sin \tilde{k}(s-t))^2 ds && \text{(Itô isometry)} \\ &= \frac{\sigma}{m\tilde{k}} \left( \frac{t}{2} - \frac{\sin 2\tilde{k}t}{4\tilde{k}} \right) \\ &\sim \frac{\sigma}{m\tilde{k}} \frac{t}{2} \quad \rightarrow \infty && \text{as } t \rightarrow \infty. \end{aligned}$$

A stochastically forced swingset will swing! If you pump your legs completely at random, you can make a frictionless swing go as high as you want. It is somewhat remarkable that although you are forcing all frequencies equally, and all of these frequencies except one 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 leads to exponential damping. The imaginary part leads to oscillations, with a frequency that is slightly shifted from the resonant frequency  $\tilde{k}$ .

**Exercise 8.2.** Write down the solution to (8) explicitly, when  $\gamma \neq 0$ .

One way to gain insight into the properties of the solution is to assume that a stationary solution exists, to look for its covariance function  $C(t)$ , and then calculate the spectral density  $f(\lambda)$ . To this aim, consider (8) at times 0 and  $t$ , and multiply these equations together to get

$$(mX_t'' + kX_t + \gamma X_t') (mX_0'' + kX_0 + \gamma X_0') = \sigma^2 \eta(t) \eta(0).$$

Now take the expectation, and use the relationships

$$\begin{aligned} \mathbb{E}X_t' X_0 &= -\mathbb{E}X_t X_0' = C'(t), & \mathbb{E}X_t' X_0' &= -C''(t), & \mathbb{E}X_t X_0'' &= \mathbb{E}X_t'' X_0 = C''(t), \\ \mathbb{E}X_t'' X_0' &= -\mathbb{E}X_t' X_0'' = C'''(t), & \mathbb{E}X_t'' X_0'' &= C^{(4)}(t). \end{aligned}$$

These relationships are obtained by differentiating  $C(t) = \mathbb{E}X_{t+s}X_s = \mathbb{E}X_sX_{s-t}$  in time. We get

$$m^2 C^{(4)}(t) + k^2 C(t) - \gamma^2 C''(t) + 2mk C''(t) = \sigma^2 \delta(t).$$

Taking the Fourier transform of this equations gives an equation for the spectral density  $f(\lambda)$ :

$$m^2 \lambda^4 f(\lambda) + k^2 f(\lambda) + \gamma^2 \lambda^2 f(\lambda) - 2mk \lambda^2 f(\lambda) = \frac{\sigma^2}{2\pi}.$$

Solving for  $f(\lambda)$  gives

$$f(\lambda) = \frac{1}{2\pi} \frac{\sigma^2}{(m\lambda^2 - k)^2 + \lambda^2 \gamma^2}.$$

This spectral density is plotted in Figure 3 for small and large damping  $\gamma$ . For small enough damping, the density has a peak at  $\tilde{\lambda} = \sqrt{\frac{k}{m} - \frac{\gamma^2}{2m^2}}$ , which is less than the resonant frequency  $\sqrt{k/m}$ . The swing swings, with random oscillations near frequency  $\tilde{\lambda}$ , but the oscillations stay bounded. For large damping, the density is peaked at  $\tilde{\lambda} = 0$ , and trajectories are much more jagged (Figure 3). Notice that when  $\gamma = 0$  the spectral density is not integrable, so we wouldn't expect a stationary solution.

### 8.3 Stratonovich Integral

The Stratonovich integral is another useful stochastic integral. There is a simple formula that relates it to the Itô integral.

**Theorem.** *Suppose  $X$  solves the Stratonovich equation*

$$dX_t = b(t, X_t)dt + \sigma(t, X_t) \circ dW_t. \quad (9)$$

*Then  $X$  also solves the Itô equation*

$$dX_t = \left( b(t, X_t) + \frac{1}{2} \sigma(t, X_t) \frac{\partial}{\partial x} \sigma(t, X_t) \right) dt + \sigma(t, X_t) dW_t. \quad (10)$$

*Conversely, suppose  $X$  solves the Itô equation*

$$dX_t = b(t, X_t)dt + \sigma(t, X_t) dW_t. \quad (11)$$

*Then  $X$  also solves the Stratonovich equation*

$$dX_t = \left( b(t, X_t) - \frac{1}{2} \sigma(t, X_t) \frac{\partial}{\partial x} \sigma(t, X_t) \right) dt + \sigma(t, X_t) \circ dW_t. \quad (12)$$

Therefore, a Stratonovich SDE is equivalent to an Itô SDE with an additional drift term. The drift arises only when the noise is multiplicative, and it depends on the rate of change of the magnitude of the noise. Heuristically, there is an extra drift because the Stratonovich integral can “see” into the future so it needs to account for the changing magnitude of the noise.

*Remark.* To transform a Stratonovich SDE to an Itô SDE and vice versa, the diffusion term  $\sigma$  must be differentiable. However, we do not need such an assumption to show existence and uniqueness of the solutions to the SDEs.

*Proof sketch.* (Pavliotis (2014), p.62.) Here is a sketch of the proof. To make it rigorous, one needs to control the error terms; this is left as an exercise.

Suppose  $X$  solves (9), which recall is shorthand for the integral equation

$$X_t = \int_0^t b(s, X_s) ds + \int \sigma(s, X_s) \circ dW_s.$$

The Riemann integral is unaffected by the definition of the stochastic integral. Therefore we consider how the Stratonovich integral in the expression above, can be transformed into an Itô integral. Let  $\{t_j\}_{j=0}^n$  be a partition of  $[0, t]$  and write  $\sigma_j \equiv \sigma(t_j, X_{t_j})$  (and similarly for  $W, X$ ), with increments  $\Delta W_j = W_{t_{j+1}} - W_{t_j}$ ,  $\Delta X_j = X_{t_{j+1}} - X_{t_j}$ ,  $\Delta t_j = t_{j+1} - t_j$ . Let  $\Delta t = \sup_j \Delta t_j$ . The Stratonovich integral is

$$\int_0^t \sigma(s, X_s) \circ dW_s = \text{m. s. lim}_{\Delta t \rightarrow 0} \sum_j \left[ \frac{\sigma_j + \sigma_{j+1}}{2} \right] (W_{j+1} - W_j). \quad (13)$$

Now we estimate  $\sigma_{j+1}$  using a Taylor expansion about  $\sigma_j$ , keeping only terms up to  $O(\sqrt{\Delta t_j})$ , since we will be multiplying them by  $\Delta W_j$ .

$$\begin{aligned} \sigma_{j+1} &= \sigma_j + \frac{\partial \sigma_j}{\partial x} \Delta X_j + O((\Delta X_j)^2) \\ &= \sigma_j + \frac{\partial \sigma_j}{\partial x} \left( b_j \Delta t_j + \left( \frac{\sigma_j + \sigma_{j+1}}{2} \right) \Delta W_j + O(\Delta t) \right) + O(\Delta t_j) \quad \text{using (9) to approximate } \Delta X_j \\ &= \sigma_j + \frac{\partial \sigma_j}{\partial x} \left( b_j \Delta t_j + \left( \frac{\sigma_j + \sigma_j + \frac{\partial \sigma_j}{\partial x} \Delta X_j}{2} \right) \Delta W_j \right) + O(\Delta t_j) \quad \text{substitute for } \sigma_{j+1} \text{ from first line} \\ &= \sigma_j + \frac{\partial \sigma_j}{\partial x} \sigma_j \Delta W_j + O(\Delta t_j). \end{aligned}$$

Now substitute the approximation for  $\sigma_{j+1}$  into (13) to obtain

$$\begin{aligned} \int_0^t \sigma(s, X_s) \circ dW_s &= \text{m. s. lim}_{\Delta t \rightarrow 0} \sum_j \left( \sigma_j + \frac{1}{2} \frac{\partial \sigma_j}{\partial x} \sigma_j \Delta W_j \right) \Delta W_j + O((\Delta t_j)^{3/2}) \\ &= \int_0^t \sigma(s, X_s) dW_s + \int_0^t \frac{1}{2} \frac{\partial \sigma(s, X_s)}{\partial x} \sigma(s, X_s) ds. \end{aligned}$$

This gives us the extra drift term in (10). Therefore  $X$  also solves the Itô equation (10).  $\square$

**Example 8.6** (Geometric Brownian Motion, revisited) Consider the Stratonovich equation

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

This is equivalent to the Itô equation

$$dN_t = \left( r + \frac{1}{2} \alpha^2 \right) N_t dt + \alpha N_t dW_t,$$

whose solution we found earlier to be  $N_t = N_0 e^{rt + \alpha W_t}$ . This is also the solution we would have found using the regular chain rule,  $d(\log N_t) = dN_t/N_t$ .

Why use the Stratonovich integral? There are several advantages:

- The regular chain rule holds, i.e.  $df(X_t) = f'(X_t) \circ dX_t$ . We will investigate this more on the homework.
- If you start with smooth, non-white noise in a one-dimensional ODE, and take a limit to make the noise white, you typically end up with a Stratonovich integral. That is, consider an ODE of the form

$$\frac{dx^\varepsilon}{dt} = b(x^\varepsilon, t) + \sigma(x^\varepsilon, t) \xi^\varepsilon(t),$$

where  $\xi^\varepsilon$  is a family of stationary stochastic processes parameterized by  $\varepsilon > 0$ , and such that as  $\varepsilon \rightarrow 0$  their covariance functions approach the covariance function for white noise,  $C_\varepsilon(t) \rightarrow \delta(t)$ . Then  $x^\varepsilon(t)$  approaches the solution to the Stratonovich SDE  $dX_t = b(X_t, t) + \sigma(X_t, t) \circ dW_t$ . (See Pavliotis (2014), Section 5.1, for an example where  $\xi^\varepsilon$  is a family of Ornstein-Uhlenbeck processes, and see Evans (2013), p.119, for an example with general  $\xi^\varepsilon$  where the solution to the SDE can be worked out explicitly.)

- 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  $d$ -dimensional sphere is the solution to

$$dX_t = P(X_t) \circ dB_t,$$

where  $B \in \mathbb{R}^d$  is a  $d$ -dimensional Brownian motion and  $P(x)$  is the orthogonal projection matrix onto the tangent space to the surface of the unit sphere at point  $x$  on the unit sphere.

What are the disadvantages?

- The Itô isometry no longer holds.
- The non-anticipating property no longer holds; the Stratonovich integral “looks into the future.”

These losses make rigorously analyzing the Stratonovich integral significantly harder. Mathematically, the Itô integral is a “martingale” but the Stratonovich integral is not; since there are many powerful tools developed for martingales it is more convenient to use the Itô integral to develop the theory of diffusion processes.

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

In multiple dimensions the conversion is given as follows

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

$$dX_t = b(t, X_t)dt + \sigma(t, X_t) \circ dW_t \tag{14}$$

is equivalent to the Itô equation

$$dX_t = (b(t, X_t) + h(t, X_t))dt + \sigma(t, X_t)dW_t. \tag{15}$$

The additional drift term is

$$h = \frac{1}{2} (\nabla \cdot (\sigma \sigma^T) - \sigma \nabla \cdot \sigma^T), \quad \text{with components} \quad h_i = \frac{1}{2} \sum_{j=1}^n \sum_{k=1}^m \sigma_{jk} \frac{\partial \sigma_{ik}}{\partial x_j}. \tag{16}$$

The additional drift also satisfies the relationship  $h \cdot v = \frac{1}{2} \sigma^T : \nabla(\sigma^T v)$  for all  $v \in \mathbb{R}^n$ .

**Example 8.7** To apply the Itô-Stratonovich conversion, it is *very important* that  $X_t$  solves an SDE – the coefficients of the equation, must depend only on  $X_t, t$  and *not* on any other processes. For example, suppose you are given the following one-dimensional equation:

$$dY_t = b(X_t) \circ dW_t, \quad \text{where} \quad dX_t = \alpha(X_t)dt + \beta(X_t)dW_t. \quad (17)$$

(The same Brownian motion is used for both processes.) It is tempting, but wrong, to convert the equation for  $Y_t$  to Itô form using the conversion rule for a one-dimensional diffusion, as

$$\text{(Incorrect!)} \quad dY_t = \frac{1}{2} b'(X_t) b(X_t) dt + b(X_t) dW_t.$$

In fact, the conversion rule is

$$dY_t = \frac{1}{2} b'(X_t) \beta(X_t) dt + b(X_t) dW_t, \quad (18)$$

which can be shown using the multidimensional conversion rule.

**Exercise 8.3.** Derive the Itô equation (18) for  $Y$  defined in Example 8.7 above. Do this in two ways: (i) start from a discrete approximation of the integrals, as in the derivation of the Stratonovich conversion rule; and (ii) use the Stratonovich conversion rule for multidimensional diffusions.

*Remark.* Here is a more general approach to the Stratonovich integral. Given processes<sup>3</sup>  $X_t, Y_t$ , the Stratonovich integral is defined in terms of the Itô integral as

$$\int_0^t Y_t \circ dX_t = \int_0^t Y_t dX_t + \frac{1}{2} \langle X, Y \rangle_t.$$

Here  $\langle X, Y \rangle_t$  is the quadratic covariation of  $X_t, Y_t$ , defined as

$$\langle X, Y \rangle_t = \text{m. s. lim}_{\Delta t \rightarrow 0} \sum_{j: t_j < t} (X_{j+1} - X_j)(Y_{j+1} - Y_j).$$

The quadratic covariation satisfies  $\langle X, Y \rangle_t = \frac{1}{2}([X + Y]_t - [X]_t - [Y]_t)$ , where  $[X]_t = \langle X, X \rangle_t$  is the quadratic variation of  $X$ . For an Itô process  $X_t = \int_0^t f(s) ds + \int_0^t g(s) dW_s$ , we have that  $[X]_t = \int_0^t g^2(s) ds$ .

Consider Example 8.7. To apply this formalism we must calculate  $\langle b(X), W \rangle_t$ . Defining  $Z_t = b(X_t)$ , we have from the Itô formula that

$$Z_t = \int_0^t (b'(X_s) \alpha(X_s) + \frac{1}{2} b''(X_s) \beta^2(X_s)) ds + \int_0^t b'(X_s) \beta(X_s) dW_t.$$

The quadratic covariation is  $\langle Z, W \rangle_t = \int_0^t b'(X_s) \beta(X_s) ds$ , giving

$$Y_t = Y_0 + \int_0^t b(X_t) dW_t + \frac{1}{2} \int_0^t b'(X_s) \beta(X_s) ds \quad \Leftrightarrow \quad dY_t = b(X_t) dW_t + \frac{1}{2} b'(X_t) \beta(X_t) dt.$$

<sup>3</sup> $X_t, Y_t$  must be continuous semimartingales, with respect to the filtration generated by  $W_t$ . A process  $X_t$  is a semimartingale if it can be decomposed as  $X_t = M_t + A_t$ , where  $M_t$  is a local martingale, and  $A_t$  is a progressively measurable process of locally bounded variation.

## 8.4 Appendix

**Theorem** (Gronwall's Inequality). *Let  $\phi(t)$ ,  $b(t) \geq 0$  be nonnegative, continuous functions defined for  $0 \leq t \leq T$ , and let  $a \geq 0$  be a constant. If*

$$\phi(t) \leq a + \int_0^t b(s)\phi(s)ds \quad \text{for all } 0 \leq t \leq T ,$$

then

$$\phi(t) \leq ae^{\int_0^t b(s)ds} \quad \text{for all } 0 \leq t \leq T .$$

*Proof.* (Theorem and proof as stated in Evans (2013), section 5.B.3) Let  $\Phi(t) = a + \int_0^t b(s)\phi(s)ds$ . Then  $\Phi' = b\phi \leq b\Phi$ , so

$$\left( e^{-\int_0^t bds} \Phi \right)' = (\Phi' - b\Phi)e^{-\int_0^t bds} \leq (b\phi - b\Phi)e^{-\int_0^t bds} = 0 .$$

Therefore

$$\Phi(t)e^{-\int_0^t bds} \leq \Phi(0) = a \quad \implies \quad \phi(t) \leq \Phi(t) \leq ae^{\int_0^t bds} .$$

□

## References

- Evans, L. C. (2013). *An Introduction to Stochastic Differential Equations*. American Mathematical Society.
- Gardiner, C. (2009). *Stochastic methods: A handbook for the natural sciences*. Springer, 4th edition.
- Karatzas, I. and Shreve, S. E. (1991). *Brownian Motion and Stochastic Calculus*. Springer.
- Koralov, L. B. and Sinai, Y. G. (2010). *Theory of Probability and Random Processes*. Springer.
- Oksendal, B. (2005). *Stochastic Differential Equations*. Springer, 6 edition.
- Pavliotis, G. A. (2014). *Stochastic Processes and Applications*. Springer.
- Varadhan, S. R. S. (2007). *Stochastic Processes*, volume 16 of *Courant Lecture Notes*. American Mathematical Society.