# Chapter 8

# Diffusions and Stochastic Differential Equations

In Chapter ?? we remarked that the limiting (as the step size parameter $h$ approached zero) generator of our Langevin MCMC schemes corresponded to a class of continuous time stochastic processes. In this chapter we will explore that limiting class of stochastic processes, called diffusions, in more detail. In the MCMC context our only concern was that our schemes had approximately the correct ergodic measure. Here, we view a diffusion as a model of some physical process and we will be interested in the details of its evolution beyond its ergodic properties.

## 8.1 Brownian motion and Diffusion processes

The simplest diffusion process is the Brownian motion. A standard Brownian motion $W^{(t)}$ is a stochastic process with values in $\mathbb{R}$ that is continuous function of $t \in [0, \infty)$ with probability 1, and which has increments $\triangle_s^t W = W^{(t)} - W^{(s)}$ for $s < t$ that are distributed according to $\mathcal{N}(0, t-s)$ and that are independent in the sense that if $s_1 < t_1 < s_2 < t_2$ then $\triangle_{s_1}^{t_1} W$ and $\triangle_{s_2}^{t_2}$ are independent random variables. If not specified otherwise, it is customary to assume that $W^{(0)} = 0$. A standard $d$-dimensional Brownian motion is a vector $\left(W_1^{(t)}, W_2^{(t)}, \ldots, W_d^{(t)}\right)$ of $d$, independant 1 dimensional standard Brownian motions.

Notice that, from the properties of the Gaussian distribution,

$$\mathbf{E}\left[\left(\triangle_s^t W\right)^4\right] = \int x^4 \frac{e^{-\frac{x^2}{2(t-s)}}}{\sqrt{2\pi(t-s)}} dx = 3(t-s)^2$$

which seems to suggest continuity of the paths of the Brownian motion. Though a Brownian motion has, by definition, sample paths that are continuous with probability one, the expression in the last display would have been enough to guarantee the existence of another process, say $U^{(t)}$ having continuous sample paths with probability 1 and for which $\mathbf{P}\left[U^{(t)} = W^{(t)}\right]$ for all $t$, according to the a famous theorem by Kolmogorov. That theorem states that, for any stochastic process $X^{(t)}$, if there are positive constants $\alpha$ and $\beta$ such that for each $T \in [0, \infty)$ you can find a constant $C$ such that

$$\mathbf{E}\left[|X^{(t)} - X^{(s)}|^\alpha\right] \leq C|t-s|^{1+\beta}$$

for all $s \leq t \leq T$, then there is a process $Y^{(t)}$ satisfying

$$\sup_{s<t} \frac{|Y^{(t)} - Y^{(s)}|}{|t-s|^\gamma} < \infty \tag{8.1}$$

for any $\gamma < \beta/\alpha$ with probability 1 and $\mathbf{P}\left[Y^{(t)} = X^{(t)}\right]$ for all $t$. Such a process $Y^{(t)}$ is called a continuous modification of $X^{(t)}$. Condition (8.1) is referred to as Hölder continuity with exponent $\gamma$ and is stronger than continuity.

In fact, one can prove the existence of Brownian motion and show that, with probability 1, its sample paths are Hölder continuous with any exponent $\gamma < 1/2$. On the other hand, you may recall that if a function is differentiable in an interval, say $(a, b)$ containing a point $r$, then, by the mean value theorem, it must satisfy (8.1) with $\gamma = 1$ (with the supremum restricted to $a < s < t < b$). One can also show that, with probability 1, a Brownian motion is not differentiable at any $t$. This will pose a problem when we are constructing integrals with respect to $W$. The Brownian motion also has a number of additional special features. For example, for any $c > 0$, the process $X^{(t)} = \sqrt{c}W^{(t/c)}$ is also a Brownian motion (i.e. satisfy the requirements in the definition of a Brownian motion), as are the processes $X^{(t)} = \triangle_c^{t+c} W$, $X^{(t)} = \triangle_1^{1-t} W$ (for $t \in [0, 1]$), and $X^{(t)} = tW^{(1/t)}$ with $X^{(0)} = 0$.

The generator of the Brownian motion is specified by the formula

$$\begin{aligned}
\mathcal{L}f &= \lim_{h\to 0} \frac{E_x\left[f(W^{(h)})\right] - f(x)}{h} \\
&= \lim_{h\to 0} \frac{1}{h} \int (f(x+\sqrt{h}z) - f(x)) \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}} dz \\
&= \frac{1}{h} \int \sqrt{h}z \int_0^1 \int_0^1 \left(f'(x) + r\sqrt{h}z f''(x+sr\sqrt{h}z)\right) \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}} drdsdz
\end{aligned}$$

which, after noting that the term involving $f'$ vanishes and assuming that $f''$ is a continuous function, becomes

$$\mathcal{L}f(x) = \frac{1}{2} f''(x).$$

In $d$-dimensions the standard Brownian motion has generator defined by

$$\mathcal{L}f(x) = \frac{1}{2}\sum_{j=1}^{d} \partial_{x_j}^2 f(x)$$

In other words, if $u(t,x) = E_x\left[f(W^{(t)})\right]$ then $u$ solves the heat equation

$$\partial_t u = \frac{1}{2}\sum_{j=1}^{d} \partial_{x_j}^2 u, \qquad u(0,x) = f(x).$$

This is completely consistent with our discovery in Chapter **??** that the limiting generator corresponding to (5.8) is (**??**) (just note that $\xi = h^{-1/2}W^{(h)}$ is an $\mathcal{N}(0,1)$ random variable and take $C = 1/2$ in those expressions).

Despite its many very particular characteristics, the Brownian motion is the basic building block for a models of many physical processes. To see why this is the case, imagine a $d$-dimensional particle with no mass initially positioned at the origin and subject to no external force but experiencing random, independent perturbations at every instant. In other words, assume that the position of stochastic process is $X^{(t)}$ and that the increments of the process $\triangle_s^t X$ are independent in the same sense as are the increments of a Brownian motion. We will not assume that the increments of $X^{(t)}$ are Gaussian but we will assume that $X^{(t)}$ is continuous with probability 1. These are perfectly reasonable, and seemingly not very restrictive assumptions for our physical process.

Now choose $s < t$ and, for any integer $N > 0$ set $h = (t-s)/N$ and observe that

$$\triangle_s^t X = \sum_{j=0}^{N-1} \triangle_{jh}^{(j+1)h} X.$$

Because $X^{(t)}$ is continuous with probability 1, we know that, with probability 1,

$$\lim_{N\to\infty} \max_{j<N} \|\triangle_{jh}^{(j+1)h} X\|_2 = 0.$$

A version of the central limit theorem tells us that for any doubly indexed sequence $\left\{\xi^{(j,N)}\right\}_{\substack{j<N\\N>0}}$ if

$$S_N = \sum_{j=0}^{N-1} \xi^{(j,N)}$$

converge in distribution to some random variable $\xi$ and if the random sequence $\sup_{j<N} \|\xi^{(j,N)}\|_2$ converges to 0 in probability, then $\xi$ must be a Gaussian random variable. Setting $\xi^{(j,N)} = \triangle_{jh}^{(j+1)h} X$, the conditions of this CLT are satisfied, and since we $S_N = \triangle_s^t X$ we can conclude that the increments of $X^{(t)}$ must all be Gaussian. Letting $b^{(t)} = \mathbf{E}\left[X^{(t)}\right]$ and $a^{(t)} = \mathbf{cov}\left[X^{(t)}\right]$, we find that $\mathbf{E}\left[\triangle_s^t X\right] = \triangle_s^t b$ and, appealing to the independence of the increments of $X$,

$$\mathbf{cov}\left[\triangle_s^t X\right] = \mathbf{cov}\left[X^{(t)}\right] - \mathbf{cov}\left[X^{(s)}\right] = \triangle_s^t a$$

so that

$$\triangle_s^t X \sim \mathcal{N}\left(\triangle_s^t b, \triangle_s^t a\right).$$

If we also assume that the distribution of an increment of $X$ depends only on the length of the corresponding time interval (i.e. that increments of the same length are identically distributed), then the functions $b^{(t)}$ and $a^{(t)}$ depend linearly on $t$.

In summary, if we assume that the position of our particle varies continuously with time and has independent increments, then the distribution of our increments will necessarily also be Gaussian. Moreover, if increments corresponding to the time intervals of the equal length have the same distribution then $X^{(t)} = bt + \sigma W^{(t)}$ for a constant vector $b$ and matrix $\sigma$.

Most physical models involving a Brownian motion cannot be solved exactly in terms of the underlying Brownian motion. In the next few sections we will introduce a family of models that define a very general process $X^{(t)}$ as an implicit function of an underlying Brownian motion. Before doing that, however, we give a few important examples of diffusion processes that can be written as explicit functions of a Brownian motion. The first of these is the Ornstein–Uhlenbeck (OU) process

$$X^{(t)} = \mu + e^{-\lambda t}(x - \mu) + \frac{\sigma}{\sqrt{2\lambda}} e^{-\lambda t} W^{(e^{2\lambda t} - 1)}$$

for $\lambda > 0$, which, at time $t$, is an $\mathcal{N}\left(\mu + e^{-\lambda t}(x - \mu), \frac{\sigma^2}{2\lambda}\left(1 - e^{-2\lambda t}\right)\right)$ random variable and is the most basic mean reverting continuous time stochastic process.

**Exercise 80.** *Show that the OU process has invariant measure* $\mathcal{N}\left(\mu, \frac{\sigma^2}{2\lambda}\right)$.

**Exercise 81.** *Show that the generator for the OU process satisfies*

$$\mathcal{L}f(x) = \lambda(\mu - x)f'(x) + \frac{1}{2}\sigma^2 f''(x).$$

The geometric Brownian motion

$$X^{(t)} = e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W^{(t)}} x$$

is a simple model for the evolution of a stock price and is the basis for the famous Black–Scholes formula for the price of a stock option.

**Exercise 82.** *Find the generator of the geometric Brownian motion.*

One final stochastic process that can be written as an explicit function of a Brownian motion is the Brownian bridge

$$X^{(t)} = (1 - t)W^{(t/(1-t))}$$

for $t \in [0, 1]$. Notice that the Brownian bridge satisfies $X^{(0)} = X^{(1)} = 0$. In a later chapter we will learn that the Brownian bridge is actually distributed exactly as the Brownian motion conditioned on $W^{(1)} = 0$.

## 8.2 The Itô integral and calculus

A general diffusion process satisfies a relationship of the form

$$X^{(t)} = x + \int_0^t b^{(s)} ds + \int_0^t \sigma^{(s)} dW^{(s)} \tag{8.2}$$

where the drift coefficient $b^{(t)}$ and diffusion coefficient $\sigma^{(t)}$ are appropriate, possibly random, functions of time. We have not yet defined the integral $\int_0^t \sigma^{(s)} dW^{(s)}$. The goal of this section is to give meaning to this integral and to discuss the consequence of the integral's definition for the diffusion process satisfying (8.2).

First observe that if $W^{(s)}$ was almost surely differentiable then we could use the definition

$$\int_0^t \sigma^{(s)} dW^{(s)} = \int_0^t \sigma^{(s)} \frac{d}{dt} W^{(s)} ds$$

and the expression in (8.2) would be equivalent to

$$\frac{d}{dt} X^{(t)} = b^{(t)} + \frac{d}{dt} W^{(t)}, \qquad X^{(0)} = x.$$

We would like to imagine that $\frac{d}{dt} W^{(t)}$ is a white noise, that is that it gives an independent, identically distributed kick to the position of $X^{(t)}$ at each instant of time. Unfortunately, as we have already remarked the Brownian motion is, with probability 1, nowhere differentiable. We are therefore in need of a new definition of the integral.

To define the integral $\int_0^t Y^{(s)} dW^{(s)}$ of some real valued stochastic process $Y^{(t)}$ we need to insist $Y^{(t)}$ is adapted to some filtration $\{\mathcal{F}_t\}_{t \geq 0}$ and that, for any $t$,

$$\int_0^t \left(Y^{(s)}\right)^2 ds < \infty \tag{8.3}$$

with probability 1. In fact, we need to insist that $Y^{(t)}$ be progressively measurable, that is if we fix $t$, then on $[0, t]$ the function $Y^{(s)}(\omega)$ of $s \in [0, t]$ and $\omega \in \Omega$ is measurable with respect to $\mathcal{B}_t \otimes \mathcal{F}_t$, the smallest $\sigma$-algebra on $[0, t] \times \Omega$ containing all sets of the form $A \times B$ where $A \in \mathcal{F}_t$ and $B$ is in the Borel $\sigma$-algebra on $[0, t]$, $\mathcal{B}_t$. To make sense of (8.2) we will need to make these same assumptions on the drift and diffusion coefficients $b^{(t)}$ and $\sigma^{(t)}$ in that expression. In most cases that will be of interest to us, $b^{(t)}$ and $\sigma^{(t)}$ will be (possibly implicitly defined) functions of the underlying Brownian motion $W^{(t)}$.

We will also assume that the Brownian motion is adapted to $\{\mathcal{F}_t\}_{t\geq 0}$ and that any increment $\triangle_s^t W$ with $s < t$ is independent of the $\sigma$-algebra $\mathcal{F}_s$. When such a relationship between the Brownian motion $W^{(t)}$ and filtration $\mathcal{F}_t$ holds we say that $W^{(t)}$ is compatible with $\mathcal{F}_t$. With these assumptions the Brownian motion is an example of an $\mathcal{F}_t$-martingale, a stochastic process $X^{(t)}$ adapted to $\{\mathcal{F}_t\}_{t\geq 0}$ and satisfying

$$\mathbf{E}\left[X^{(t)} \,|\, \mathcal{F}_s\right] = X^{(s)}$$

for all $s \leq t$. Note that for any martingale $X^{(t)}$, the expectation of $X^{(t)}$ is a constant function of time. For example, in the case of the Brownian motion we already know that $\mathbf{E}\left[W^{(t)}\right] = 0$. Martingales play an important role in the theory of stochastic processes. They have a number of very useful properties and will play a key role in the developments below.

Let us suppose at first that $Y^{(t)}$ is continuous with probability 1 and that instead of (8.3) we have the stronger condition

$$\mathbf{E}\left[\int_0^t \left(Y^{(s)}\right)^2 ds\right] < \infty. \tag{8.4}$$

Suppose further that that $0 = t_0 < t_1 < \cdots < t_k = t$ is any partition of the interval $[0, t]$ and that, for each $j$, $t_j^* = t_j + \alpha(t_{j+1} - t_j)$ for some $\alpha \in [0, 1]$. Under these assumptions, the Riemann integral

$$\int_0^t Y^{(s)} \circ dW^{(s)} = \lim_{\max_{j<k} |t_{j+1}-t_j| \to 0} \sum_{j=0}^{k-1} Y^{(t_j^*)} \triangle_{t_j}^{t_{j+1}} W \tag{8.5}$$

can be shown to converge in probability to a continuous process. We will reserve the symbol $\int_0^t Y^{(s)} dW^{(s)}$ for the limit in (8.5) when $t_j^* = t_j$ (i.e. $\alpha = 0$). That limit is called the Itô stochastic integral. For more general $Y^{(t)}$, adapted to $\{\mathcal{F}_t\}_{t\geq 0}$ and satisfying (8.3), the stochastic integral can be defined as the limit of the stochastic integrals of a sequence of continuous processes approximating $Y^{(t)}$. Just like other integrals, stochastic integrals are linear operators and we can define

$$\int_r^t Y^{(s)} \circ dW^{(s)} = \int_0^t Y^{(s)} \circ dW^{(s)} - \int_0^r Y^{(s)} \circ dW^{(s)}.$$

For reasons we have already mentioned, if $W$ had differentiable paths then these Riemann integrals would be independent of the choice of $t_j^* \in [t_j, t_{j+1})$. To see that the stochastic integrals defined in (8.5) are not independent of the $t_j^*$, we can plug $Y^{(t)} = W^{(t)}$ into (8.5) and notice that

$$\begin{aligned} \int_0^t W^{(s)} \circ dW^{(s)} &= \lim_{\max_{j<k} |t_{j+1}-t_j| \to 0} \sum_{j=0}^{k-1} W^{(t_j^*)} \triangle_{t_j}^{t_{j+1}} W \\ &= \lim_{\max_{j<k} |t_{j+1}-t_j| \to 0} \sum_{j=0}^{k-1} W^{(t_j)} \triangle_{t_j}^{t_{j+1}} W \\ &\quad + \lim_{\max_{j<k} |t_{j+1}-t_j| \to 0} \sum_{j=0}^{k-1} \left( \triangle_{t_j}^{t_j^*} W \right)^2 \\ &\quad + \lim_{\max_{j<k} |t_{j+1}-t_j| \to 0} \sum_{j=0}^{k-1} \triangle_{t_j}^{t_j^*} W \triangle_{t_j^*}^{t_{j+1}} W \end{aligned} \tag{8.6}$$

since all three terms in the last display converge (at least in probability). The First limit in the last term of the previous display corresponds to the stochastic integral in (8.5) with $t_j^* = t_j$, i.e. to the Itô integral $\int_0^t W^{(s)} dW^{(s)}$. If the other two terms vanish then the stochastic integral in (8.5) is independent of the choice of $\alpha$. The third term is a sum of independent random variables each of which has mean zero and variance

$$\mathbf{E}\left[ \left( \triangle_{t_j}^{t_j^*} W \triangle_{t_j^*}^{t_{j+1}} W \right)^2 \right] = \mathbf{E}\left[ \left( \triangle_{t_j}^{t_j^*} W \right)^2 \right] \mathbf{E}\left[ \left( \triangle_{t_j^*}^{t_{j+1}} W \right)^2 \right] = (t_j^* - t_j)(t_{j+1} - t_j^*).$$

Since the term on the rand hand side of this expression is bounded by $\max_{j<k} |t_{j+1} - t_j|^2/4$, the sum of the variances of each summand converges to zero and the strong law of large numbers tells us that the third limit goes to zero with probability 1.

The second limit in (8.6) does not vanish however. This is already clear because each term in teh sum is non-negative. But we can be more precise.

Each term in that sum is an independent random variable with mean $t_j^* - t_j$ and variance $2(t_j^* - t_j)^2$. So again the sum of the variances of each summand converges to zero and so, with probability 1, the sum must converge to the limit of the sum of the means, which is

$$\lim_{\max_{j<k}|t_{j+1}-t_j|\to 0} \sum_{j=0}^{k-1} (t_j^* - t_j) = \alpha t$$

In other words we have shown that

$$\int_0^t W^{(s)} \circ dW^{(s)} = \int_0^t W^{(s)} dW^{(s)} + \alpha t,$$

and, more generally, that the stochastic integral in (8.5) depends on the choice of $\alpha$. Besides the Itô integral ($\alpha = 0$) one occasionally encounters the Stratonovich integral ($\alpha = 1/2$). From now on if we write $\int_0^t Y^{(s)} \circ dW^{(s)}$ we are referring to the Stratonovich integral. Repeating the above calculations for the Statonovich integral of $Y^{(t)} = f(W^{(t)})$ for a smooth function $f$ we would find that

$$\int_0^t f(W^{(s)}) \circ dW^{(s)} = \int_0^t f(W^{(s)}) dW^{(s)} + \frac{1}{2}\int_0^t f'(W^{(s)}) ds.$$

Many of the interesting properties of the Itô stochastic integral for general integrands can be deduced by considering the integral of so called elementary properties. A process $Y^{(t)}$ is elementary if, for some set of times $0 = t_0 < t_1 < t_2 < \cdots$, with $t_j \to \infty$ as $j \to \infty$,

$$Y^{(t)} = Y^{(0)} \mathbf{1}_{\{0\}}(t) + \sum_{j=0}^{\infty} \zeta^{(j)} \mathbf{1}_{(t_j, t_{j+1}]}(t)$$

where each $\zeta^{(j)}$ is $\mathcal{F}_{t_j}$-measurable and $\mathbf{E}\left[\left(\zeta^{(j)}\right)^2\right] < \infty$ for each $j$. Note that the $zeta^{(j)}$ are *not* assumed to be independent. The Itô stochastic integral of an elementary process, $Y^{(t)}$, is

$$\int_0^t Y^{(s)} dW^{(s)} = \sum_{j=0}^{\infty} \zeta^{(j)} \triangle_{t_j}^{t_{j+1}} W.$$

From this expression it is clear that the Itô integral of an elementary process is a martingale since, if $r \leq t$,

$$\begin{aligned}
\mathbf{E}\left[\int_0^t Y^{(s)} dW^{(s)} \mid \mathcal{F}_r\right] &= \mathbf{E}\left[\int_r^t Y^{(s)} dW^{(s)} \mid \mathcal{F}_r\right] + \int_0^r Y^{(s)} dW^{(s)} \\
&= \mathbf{E}\left[\sum_{j=0}^{\infty} \zeta^{(j)} \triangle_{(t_j \vee r)\wedge t}^{(t_{j+1}\vee r)\wedge t} W \mid \mathcal{F}_r\right] + \int_0^r Y^{(s)} dW^{(s)} \\
&= \sum_{j=0}^{\infty} \zeta^{(j)} \mathbf{E}\left[\triangle_{(t_j \vee r)\wedge t}^{(t_{j+1}\vee r)\wedge t} W \mid \mathcal{F}_r\right] + \int_0^r Y^{(s)} dW^{(s)} \\
&= \int_0^r Y^{(s)} dW^{(s)}
\end{aligned}$$

since $\zeta^{(t)}$ is $\mathcal{F}_{t_j}$-measurable and $\triangle_{(t_j \vee r)\wedge t}^{(t_{j+1}\vee r)\wedge t} W$ is independent of $\mathcal{F}_r$ and has mean 0. The Itô integral of more general processes is also a martingale.

Another interesting property of the Itô integral easily deducible when the integrand is an elementary process is the Itô isometry,

$$\mathbf{E}\left[\left(\int_0^t Y^{(s)} dW^{(s)}\right)^2\right] = \mathbf{E}\left[\int_0^t \left(Y^{(s)}\right)^2 ds\right],$$

valid for general Itô integrals whenever (8.4) is satisfied. Indeed, if $Y^{(t)}$ is elementary, then

$$\begin{aligned}
\mathbf{E}\left[\left(\int_0^t Y^{(s)} dW^{(s)}\right)^2\right] &= \sum_{j=0}^{\infty} \mathbf{E}\left[\left(\zeta^{(j)}\right)^2 \left(\triangle_{t_j\wedge t}^{t_{j+1}\wedge t} W\right)^2\right] \\
&\quad + 2\sum_{i<j} \mathbf{E}\left[\zeta^{(j)}\zeta^{(i)} \triangle_{t_j\wedge t}^{t_{j+1}\wedge t} W \triangle_{t_i\wedge t}^{t_{i+1}\wedge t} W\right] \\
&= \sum_{j=0}^{\infty} \mathbf{E}\left[\mathbf{E}\left[\left(\zeta^{(j)}\right)^2 \left(\triangle_{t_j\wedge t}^{t_{j+1}\wedge t} W\right)^2 \mid \mathcal{F}_{t_j}\right]\right] \\
&\quad + 2\sum_{i<j} \mathbf{E}\left[\mathbf{E}\left[\zeta^{(j)}\zeta^{(i)} \triangle_{t_j\wedge t}^{t_{j+1}\wedge t} W \triangle_{t_i\wedge t}^{t_{i+1}\wedge t} W \mid \mathcal{F}_{t_j}\right]\right]
\end{aligned}$$

Observing that $\zeta^{(j)}$ is $\mathcal{F}_{t_j}$ measurable and $\triangle_{t_j\wedge t}^{t_{j+1}\wedge t} W$ is independent of $\mathcal{F}_{t_j}$, the double sum in the last expression vanishes and the expression becomes the Itô isometry. In fact, for two Itô integrals we have

$$\mathbf{E}\left[\int_0^t X^{(s)}dW^{(s)}\int_0^t Y^{(s)}dW^{(s)}\right] = \mathbf{E}\left[\int_0^t X^{(s)}Y^{(s)}ds\right],$$

as long as $X^{(t)}$ and $Y^{(t)}$ satisfy (8.4).

An expression similar to the Itô isometry holds for the so called quadratic variation of an Itô integral. For any Markov process of the form in (8.2), $X^{(t)}$, if the limit

$$[X,X]^{(t)} = \lim_{\max_{j<k}|t_{j+1}-t_j|\to 0}\sum_{j=0}^{k-1}\left(\triangle_{t_j}^{t_{j+1}}X\right)^2$$

exists and does not depend on the particular partition of $[0,t]$, we call the result the quadratic variation of $X^{(t)}$. The discussion after (8.6) reveals that for Brownian motion the quadratic variation is $[W,W]^{(t)} = t$. In fact, this observation has a converse. Any continuous martingale with $[W,W]^{(t)} = t$ is a Brownian motion.

One can also check that any continuous process with finite variation, i.e. with

$$\lim_{\max_{j<k}|t_{j+1}-t_j|\to 0}\sum_{j=0}^{k-1}|\triangle_{t_j}^{t_{j+1}}X| < \infty$$

with probability 1 for any partition of $[0,1]$, must have $[X,X]^{(t)} = 0$. The covariation (not to be confused with the covariance) of a pair of Markov processes $X^{(t)}$ and $Y^{(t)}$ of the form in (8.2) is also an important quantity and can be defined either by

$$[X,Y]^{(t)} = \lim_{\max_{j<k}|t_{j+1}-t_j|\to 0}\sum_{j=0}^{k-1}\left(\triangle_{t_j}^{t_{j+1}}X\right)\left(\triangle_{t_j}^{t_{j+1}}Y\right)$$

or by

$$[X,Y]^{(t)} = \frac{1}{4}\left([X+Y,X+Y]^{(t)} - [X-Y,X-Y]^{(t)}\right).$$

Notice that the covariation is bilinear, i.e. $[X,Y]^{(t)}$ is linear in both its $X$ and $Y$ arguments.

Examining the quadratic variation of the Itô stochastic integral it is again useful to consider the integral of an elementary process, $Y^{(t)}$. In this case, if $M^{(t)} = \int_0^t Y^{(s)} dW^{(s)}$ then if for each $j$, $t_j = s_{j,0} < s_{j,1} < \cdots < s_{j,k} = t_{j+1}$,

$$\begin{aligned}
[M, M]^{(t)} &= \lim_{\max_{\{j,\ell:\, t_j<t,\, \ell<k\}} |s_{j,\ell+1}-s_{j,\ell}| \to 0} \sum_{j=0}^{\infty} \sum_{\ell=0}^{k} \left( \triangle_{s_{j,\ell}\wedge t}^{s_{j,\ell+1}\wedge t} M \right)^2 \\
&= \sum_{j=0}^{\infty} \zeta^{(j)} \lim_{\max_{\ell<k} |s_{j,\ell+1}-s_{j,\ell}| \to 0} \sum_{\ell=0}^{k} \left( \triangle_{s_{j,\ell}\wedge t}^{s_{j,\ell+1}\wedge t} W \right)^2 \\
&= \sum_{j=0}^{\infty} \left(\zeta^{(j)}\right)^2 [W, W]^{(t)} \\
&= \int_0^t \left(Y^{(s)}\right)^2 ds.
\end{aligned}$$

The expression

$$\left[ \int_0^{\cdot} Y^{(s)} dW^{(s)}, \int_0^{\cdot} Y^{(s)} dW^{(s)} \right]^{(t)} = \int_0^t \left(Y^{(s)}\right)^2 ds$$

holds for the general Itô integral as well. Similarly, the covariation of two Itô integrals is

$$\left[ \int_0^{\cdot} Y^{(s)} dW^{(s)}, \int_0^{\cdot} Z^{(s)} dW^{(s)} \right]^{(t)} = \int_0^t Y^{(s)} Z^{(s)} ds.$$

Now that we have defined the Itô stochastic integral with respect to a Brownian motion, we can extend the definition to an integral with respect to any process of the form in (8.2) by

$$\int_0^t Y^{(s)} dX^{(s)} = \int_0^t Y^{(s)} b^{(s)} ds + \int_0^t Y^{(s)} \sigma^{(s)} dW^{(s)}$$

under the same restrictions on $Y^{(t)}$ as for the integral with respect to the Brownian motion. In fact, a useful shorthand for this definition is

$$dX^{(s)} = b^{(s)} ds + \sigma^{(s)} dW^{(s)}.$$

As for the integral with respect to a Brownian motion we see that if $X^{(t)}$ is a continuous martingale (i.e. if $b^{(s)} = 0$), the Itô stochastic integral with respect to $X^{(t)}$ is also a martingale. Also as before, when the integrand is continuous we can write

$$\int_0^t Y^{(s)} dX^{(s)} = \lim_{\max_{j<k} |t_{j+1}-t_j| \to 0} \sum_{j=0}^{k-1} Y^{(t_j)} \triangle_{t_j}^{t_{j+1}} X$$

where $0 = t_0 < t_1 < \cdots < t_k = t$ is a partition of $[0, t]$ and convergence occurs in probability. Similarly, we can define the Stratanovich integral of a continuous process $Y^{(t)}$ against $X^{(t)}$ by

$$\int_0^t Y^{(s)} \circ dX^{(s)} = \lim_{\max_{j<k} |t_{j+1}-t_j| \to 0} \sum_{j=0}^{k-1} Y^{((t_j+t_{j+1})/2)} \triangle_{t_j}^{t_{j+1}} X$$

and find, as before, that it is related to the Itô integral by

$$\int_0^t f(X^{(s)}) \circ dX^{(s)} = \int_0^t f(X^{(s)}) dX^{(s)} + \frac{1}{2} \int_0^t f'(X^{(s)}) d[X, X]^{(s)} \tag{8.7}$$

any time $f$ is a smooth function.

Observe that, for any process $X^{(t)}$ and partition $0 = t_0 < t_1 < \cdots < t_k = t$ of $[0, t]$,

$$\sum_{j=0}^{k-1} \left( \triangle_{t_j}^{t_{j+1}} X \right)^2 = \left( X^{(t)} \right)^2 - \left( X^{(0)} \right)^2 - 2 \sum_{j=0}^{k-1} X^{(t_j)} \triangle_{t_j}^{t_{j+1}} X.$$

If $X^{(t)}$ is of the form in (8.2), it will be continuous and we will find upon taking finer and finer partitions, that

$$[X, X]^{(t)} = \left( X^{(t)} \right)^2 - \left( X^{(0)} \right)^2 - 2 \int_0^t X^{(s)} dX^{(s)}. \tag{8.8}$$

More generally, if $X$ and $Y$ are both of the form in (8.2) then, using the definition of the covariation of $X$ and $Y$ above we find that

$$[X, Y]^{(t)} = X^{(t)} Y^{(t)} - X^{(0)} Y^{(0)} - \int_0^t X^{(s)} dY^{(s)} - \int_0^t Y^{(s)} dX^{(s)}. \tag{8.9}$$

The relationship of this last expression, called the Itô product rule, with the standard integration by parts formula of calculus is striking. In the classical integration by parts formula the expression on the right hand side of the last display would be 0. More generally, for a vector of processes $X^{(t)} = (X_1^{(t)}, X_2^{(t)}, \ldots, X_d^{(t)}) \in \mathbb{R}^d$ and any twice continuously differentiable function $f$ on $\mathbb{R}^d$, we have Itô's formula

$$f(X^{(t)}) = f(X^{(0)}) + \int_0^t \nabla f(X^{(s)}) dX^{(s)} + \frac{1}{2}\text{trace}\left(\int_0^t D^2 f(X^{(s)} d[X,X]^{(s)}\right) \tag{8.10}$$

where $dX^{(s)}$ represents the vector with $j$th entry $dX_j^{(s)}$ and $d[X,X]^{(s)}$ represents the matrix with $i,j$th entry $d[X_i, X_j]^{(s)}$. Expressions (8.8), (8.9), and (8.10) can be usefully summarized as

$$d\left(X^{(t)}\right)^2 = 2X^{(t)} dX^{(t)} + d[X,X]^{(t)},$$

$$d\left(X^{(t)}Y^{(t)}\right) = X^{(t)} dY^{(t)} + Y^{(t)} dX^{(t)} + d[X,Y]^{(t)}$$

(hence the name product rule), and

$$df(X^{(t)}) = \nabla f(X^{(t)}) dX^{(t)} + \frac{1}{2}\text{trace}\left(D^2 f(X^{(t)}) d[X,X]^{(t)}\right)$$

(compare this expression to the usual chain rule of calculus).

**Exercise 83.** *Show that if Stratonovich integral obeys the usual chain rule, i.e. that*

$$df(X^{(t)}) = \nabla f(X^{(t)}) \circ dX^{(t)}.$$

Using the fact that the covariation is bilinear and that the standard Riemann integral $\int_0^t b^{(s)} ds$ has finite variation and therefore zero quadratic variation, if $X^{(t)}$ is of the form in (8.2) then

$$d[X,X]^{(t)} = \left[\int_0^{\cdot} \sigma^{(s)} dW^{(s)}, \int_0^{\cdot} \sigma^{(s)} dW^{(s)}\right]^{(t)} = \int_0^t \sigma^{(s)} \left(\sigma^{(s)}\right)^{\mathrm{T}} ds$$

and Itô's formula (in differential form) becomes

$$\begin{aligned} df(X^{(t)}) = \nabla f(X^{(t)}) b^{(t)} ds + \frac{1}{2}\text{trace}\left(D^2 f(X^{(t)}) \sigma^{(s)} \left(\sigma^{(s)}\right)^{\mathrm{T}}\right) ds \\ + \nabla f(X^{(t)}) \sigma^{(t)} dW^{(t)}. \end{aligned} \tag{8.11}$$

We close this section by mentioning one important consequence of Itô's formula that we will use later. Suppose that $v^{(t)}$ is a bounded processes that is adapted to the filtration, $\mathcal{F}_t$, generated by a Brownian motion $W$ and define

$$\hat{W}^{(t)} = W^{(t)} + \int_0^t v^{(s)} ds.$$

Let

$$Z^{(t)} = \exp\left(-\int_0^t \left(v^{(s)}\right)^{\mathrm{T}} dW^{(s)} - \frac{1}{2}\int_0^t \|v^{(s)}\|_2^2 ds\right).$$

Applying Itô's formula to $f(Y^{(t)})$ with $f(x) = e^x$ and

$$Y^{(t)} = \int_0^t \left(v^{(s)}\right)^{\mathrm{T}} dW^{(s)} + \frac{1}{2}\int_0^t \|v^{(s)}\|_2^2 ds$$

gives

$$dZ^{(t)} = -Z^{(t)} dY^{(t)} + \frac{1}{2} Z^{(t)} d[Y,Y]^{(t)} = -Z^{(t)} \left(v^{(s)}\right)^{\mathrm{T}} dW^{(s)}.$$

Since $Z$ is equal to a stochastic integral it is a martingale (actually we need $Z$ to be integrable too) and, since $Z^{(0)} = 1$, $\mathbf{E}\left[Z^{(t)}\right] = 1$. Moreover, $Z^{(t)}$ is positive. The process $Z$ can be regarded as a change of measure (like the ratio of two densities in importance sampling).

Suppose that the probability distribution of $W$ is $P$. From what we have already learned about $Z$, we can define a new probability distribution $Q$ on the $\sigma$-algebra $\mathcal{F}_\infty = \lim_{t\to\infty} \mathcal{F}_t$ by

$$Q\left[A\right] = E^P\left[Z^{(t)},\, A\right]$$

for any $A \in \mathcal{F}_t$. In other words, $Q$ is the probability distribution that is absolutely continuous with respect to $P$ and has change of measure (for paths on $[0,t]$),

$$\frac{dQ}{dP}^{(t)} = Z^{(t)}.$$

In fact, the restriction of $P$ to $\mathcal{F}_t$ is also absolutely continuos with respect to the corresponding restriction of $Q$. Note that on under $Q$, $W$ is no longer a Brownian motion.

For any observable $f$, by Itô's product rule we can write

$$\begin{aligned} d\left(\hat{W}^{(t)}Z^{(t)}\right) &= \hat{W}^{(t)}dZ^{(t)} + Z^{(t)}d\hat{W}^{(t)} + d[\hat{W}, Z]^{(t)} \\ &= \hat{W}^{(t)}dZ^{(t)} + Z^{(t)}dW^{(t)} + Z^{(t)}v^{(s)}ds + d[\hat{W}, Z]^{(t)} \\ &= \hat{W}^{(t)}dZ^{(t)} + Z^{(t)}dW^{(t)} \end{aligned}$$

since $d[\hat{W}, Z]^{(t)} = -Z^{(t)}v^{(t)}dt$. So we see that the process $M^{(t)} = \hat{W}^{(t)}Z^{(t)}$ is a continuous martingale under $P$. Now let $A$ be any event in $\mathcal{F}_s$ (and hence also in $\mathcal{F}_t$). The fact that $M^{(t)}$ is a martingale (along with the definition of conditional expectation) implies that

$$E^Q\left[\hat{W}^{(t)},\, A\right] = E^P\left[\hat{W}^{(t)}Z^{(t)},\, A\right] = E^P\left[\hat{W}^{(s)}Z^{(s)},\, A\right] = E^Q\left[\hat{W}^{(s)},\, A\right]$$

so that $\hat{W}^{(t)}$ is a martingale under $Q$. Since it also satisfies $[\hat{W}, \hat{W}]^{(t)} = t$, $X^{(t)}$ must be a Brownian motion under $Q$. This observation is known as Girsanov's theorem and it will have important consequences later in these notes. In fact, the boundedness assumption on $v^{(t)}$ can be relaxed to the so-called Novikov condition,

$$E^P\left[\exp\left(\frac{1}{2}\int_0^t \|v^{(s)}\|_2^2 ds\right)\right] < \infty \tag{8.12}$$

which is sufficient to guarantee that $Z^{(t)}$ is a martingale under $P$.

## 8.3 Stochastic differential equations and their solutions

A stochastic differential equation (SDE) is an equation of the form

$$X^{(t)} = X^{(t_0)} + \int_{t_0}^t b^{(s)}(X^{(s)})ds + \int_{t_0}^t \sigma^{(s)}(X^{(s)})dW^{(s)},\ t \geq t_0, \qquad X^{(t_0)} \sim \mu \tag{8.13}$$

where $b^{(t)}$ and $\sigma^{(t)}$ are now functions with values in $\mathbb{R}^d$ and $\mathbb{R}^{d\times d}$ respectively. At a minimum, the drift and diffusion coefficients $b$ and $\sigma$ have to satisfy a progressive measurability criterion as functions on $[0, t] \times \mathbb{R}^d$ similar to the one we required of the integrands in Itô integrals as functions on $[0,t] \times \Omega$. This is required to guarantee, that $Y^{(t)} = \sigma^{(t)}(X^{(t)})$ satisfies the progressive measurability criterion. The left hand side of this expression can be regarded as a special case of (8.2). Given a particular filtration $\{\mathcal{F}_t\}_{t\geq 0}$, the equation (8.13) has a solution when there is an $\mathcal{F}_t$ adapted process $X^{(t)}$ so that (8.13) holds with probability 1.

But this definition of solution leaves room for further specification. We will say that the SDE has a strong solution if, given *any* particular probability space $(\Omega, \mathcal{F}, \mathbf{P})$, filtration $\mathcal{F}_t$, and $\mathcal{F}_t$-Brownian motion $W$, we can find a continuous process $X^{(t)}$ solving (8.13) with probability 1. In contrast, the SDE has a weak solution if there is *some* probability space, filtration, and adapted Brownian motion for which we can find a continuous process solving (8.13) with probability 1. These two notions of solutions have two corresponding notions of uniqueness (recall that well-posedness of any equation requires that there be a solution and that that solution be unique). Solutions are said to be unique-in-law if, any two weak solutions of (8.13) have the same distribution. On the other hand, solutions are said to be pathwise-unique if any two solutions $X$ and $Y$ corresponding to the same probability space, filtration, and Brownian motion, and with $X^{(0)} = Y^{(0)}$ with probability 1, are equal with probability 1. It turns out that if weak solutions of (8.13) exist but are also pathwise unique then strong solutions exist and are unique and can be written as a function of $X^{(0)}$ and of the underlying Brownian motion, i.e. $X = F(X^{(0)}, W)$ for some $F$.

As you might expect, conditions guaranteeing existence of weak solution are weaker than conditions guaranteeing existence of strong solutions. If the drift and diffusion coefficients $b$, and $\sigma$, are merely continuous then weak solutions will exist (though they may not be unique). On the other hand, much more is typically required of the drift and diffusion coefficients to guarantee existence of strong solutions. For example, if the drift and diffusion coefficients are both Lipschitz (e.g. if for some constant $C$, $|b^{(t)}(x) - b^{(t)}(y)| \leq C|x-y|$ for all $x, y \in \mathbb{R}^d$) then we know that strong solutions exist and are pathwise unique. For most problems of practical interest, only the notions of weak solutions and uniqueness-in-law are relevant, though we can expect many of the models we encounter to have strong solutions and it is occasionally useful to think of the solution as a function of the initial condition and underlying Brownian motion as we can only do if we pathwise uniqueness.

A useful alternative characterization of stochastic differential equations is the notion of martingale problems. Defining the Kolmogorov operator $\mathcal{L}^{(t)}$ as in Chapter ?? by

$$\mathcal{L}^{(t)} f(x) = (\nabla f(x))\, b^{(t)}(x) + \frac{1}{2}\text{trace}\left(a^{(t)}(x) D^2 f(x)\right),$$

where

$$a^{(t)}(x) = \left(\sigma^{(t)}(x)\right)\left(\sigma^{(t)}(x)\right)^{\mathrm{T}},$$

a continuous process $X$ is said to solve the martingale problem for the drift diffusion pair $b, \sigma$ if, for any smooth function $f$,

$$M_f^{(t)} = f(X^{(t)}) - f(X^{(0)}) - \int_0^t \mathcal{L} f(X^{(s)}) ds$$

is a maringale. Note that if $X^{(t)}$ satisfies (8.13) then, by Itô's formula we expect $M_f$ to be an Itô stochastic integral and hence a martingale. In fact, equation (8.13) has a weak solution $X$ if and only if $X$ solves the martingale problem for $b, \sigma$. In most cases you can expect the operator

$$\mathcal{T}^t f = E_x\left[f(X^{(t)})\right]$$

to be a well behaved (so called Feller) semigroup with generator $\mathcal{L}$.

Returning to Girsanov's theorem established at the end of the last section, observe that if $\hat{W}$ is a Brownian motion under the measure $Q$ introduced there then the distribution of

$$\hat{X}^{(t)} = x + \int_0^t \left(b^{(s)} + \sigma^{(s)} v^{(s)}\right) ds + \int_0^t \sigma^{(s)} dW^{(s)}$$

under $Q$ is exactly the same as the distribution of $X^{(t)}$ defined by (8.2) under the original measure. In particular, if $f$ is a test function we can write

$$\mathbf{E}\left[f(\hat{X}^{(t)}) Z^{(t)}\right] = \mathbf{E}\left[f(X^{(t)})\right].$$

Consequently, in order to compute the expectation on the right of the last display we can generate samples of $f(\hat{X}^{(t)})Z^{(t)}$ instead of samples of $f(X^{(t)})$. We will see in Chapter ?? that by careful choice of the process $v^{(t)}$ we can arrange that the variance of $f(\hat{X}^{(t)})Z^{(t)}$ is much lower than the variance of $f(X^{(t)})$.

Before closing this section we revisit a few of the diffusions we introduced earlier and express them in terms of the SDE they solve. The OU process is the solution to the equation

$$dX^{(t)} = \lambda \int_0^t \left(\mu - X^{(t)}\right) dt + \sigma dW^{(t)}, \qquad X^{(0)} = x.$$

To see this consider the evolution of $Y^{(t)} = \left(X^{(t)} - \mu\right) e^{\lambda t}$. By Itô's formula

$$dY^{(t)} = \lambda e^{\lambda t} \left(X^{(t)} - \mu\right) + e^{\lambda t} dX^{(t)} = \sigma e^{\lambda t} dW^{(t)}.$$

From this formula we see that

$$X^{(t)} = \mu + e^{-\lambda t}(x - \mu) + \sigma e^{-\lambda t} \int_0^t e^{\lambda s} dW^{(s)}.$$

Examining the stochastic integral in this last expression we find that

$$\left[\int_0^{\cdot} e^{\lambda s} dW^{(s)}, \int_0^{\cdot} e^{\lambda s} dW^{(s)}\right]^{(t)} = \int_0^t e^{2\lambda s} ds = \frac{1}{2\lambda}\left(e^{2\lambda t} - 1\right).$$

So we see that $\int_0^{\cdot} e^{\lambda s} dW^{(s)}$ has the same distribution as $(2\lambda)^{-1/2} W^{(e^{2\lambda t}-1)}$ so that $X^{(t)}$ is equal in distribution to

$$x + e^{-\lambda t}(x - \mu) + \frac{\sigma}{\sqrt{2\lambda}} e^{-\lambda t} W^{(e^{2\lambda t}-1)}.$$

Now consider the stochastic differential equation

$$dX^{(t)} = \mu X^{(t)} dt + \sigma X^{(t)} dW^{(t)}, \qquad X^{(0)} = x.$$

We will see that the solution to this equation is the geometric Brownian motion. Letting $Y^{(t)} = \log X^{(t)}$, we find that

$$dY^{(t)} = \frac{1}{X^{(t)}} dX^{(t)} - \frac{1}{2}\sigma^2 dt = \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma dW^{(t)}.$$

so that

$$Y^{(t)} = \log x + \left(\mu - \frac{\sigma^2}{2}\right) t + \sigma W^{(t)}$$

or

$$X^{(t)} = e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W^{(t)}} x.$$

Finally, we turn to the Brownian bridge. It is the solution to the SDE

$$dX^{(t)} = -\frac{X^{(t)}}{1-t} dt + dW^{(t)}.$$

To see this consider the $Y^{(t)} = (1-t)^{-1} X^{(t)}$ for which

$$dY^{(t)} = \frac{1}{(1-t)^2} X^{(t)} + \frac{1}{1-t} dX^{(t)} = \frac{1}{1-t} dW^{(t)}.$$

So we see that $Y^{(t)}$ is a martingale. It has quadratic variation

$$[Y, Y]^{(t)} = \int_0^t \frac{1}{(1-s)^2} ds = \frac{t}{1-t}$$

and is therefore equivalent in distribution to $W^{(t/(1-t)}$ so that $X^{(t)}$ is equivalent in distribution to $(1-t)W^{(t/(1-t)}$.

## 8.4 Numerical approximation of SDE

As mentioned in the previous section, the PDE introduced there arise in a wide range of applications. When the independent variable $x$ is low dimensional (e.g. in $\mathbb{R}^d$ for $d \leq 3$), the best strategies for solving those PDE are classical and involve projecting the equation onto a finite basis (say of size $n$) and solving the resulting linear algebra problems. The cost of this strategy of course depends on $n$ and, in most cases, you can expect $n$ to grow exponentially with dimension (this is a version of the same problem we described for numerical integration in Chapter 1).

Fortunately, the probabilistic relationships in the previous section suggest an alternate route that is potentially independent of dimension. For example, appealing to (10.8), if we need to solve the equation (10.4) in high dimensions we can use the estimator

$$v(t, x) \approx \frac{1}{N} \sum_{j=0}^{N-1} f(X^{(j,t)}) e^{\int_0^t V(X^{(j,t)}) ds}$$

where each for each $j$, $X^{(j,t)}$ is an independent copy of the solution to (**??**) with $t_0 = 0$. Unfortunately, in most cases it is impossible to generate samples of the solution to (**??**). We can, however, approximate the solution. In fact, we have already seen one example of this in Section 5.1. There we argued that if $X_h^{(k)}$ was the chain generated by the Metropolis scheme with proposal density $\mathcal{N}(x, 2h)$ then, for $h = t/k$, the function

$$u_k(\ell h, x) = E_{\ell,x}\left[f(X_h^{(k)})\right]$$

converges to the solution $u$ of the backward Kolmogorov equation (10.4) with

$$b(x) = (\log \pi)'(x) \quad \text{and} \quad \sigma(x) = \sqrt{2}$$

which, as we saw in the last section, describes the moments of the solution to (**??**) via

$$u(r, x) = E_{r,x}\left[f(X^{(t)})\right].$$

There we also commented that the key requirement was that

$$\lim_{h\to 0} \frac{E_x\left[\triangle_0^1 X_h\right]}{h} = b(x)$$

and

$$\lim_{h\to 0} \frac{E_x\left[\left(\triangle_0^1 X_h\right)\left(\triangle_0^1 X_h\right)^{\mathrm{T}}\right]}{h} = a(x)$$

and that the limits converge sufficiently quickly.

In fact, as we will see in a moment, under assumptions on the drift and diffusion coefficients in (**??**), if $X_h^{(k)}$ is generated by the general iterative rule

$$X_h^{(k+1)} = X_h^{(k)} + F_h(k, X_h^{(k)}, \xi^{(k+1)}), \qquad X_h^{(0)} = x, \tag{8.14}$$

for some function $F_h$ and sequence of independent random vectors $\xi^{(k)}$, then the moments $E_x\left[f(X_h^{(k)})\right]$ converge to $E_x\left[f(X^{(t)})\right]$. For example, in one dimension, when we assume that $b$ and $\sigma$ are bounded and Lipschitz continuous and have bounded derivatives and that $f$ and its derivatives are also bounded, then if, for some constant $K$ and every $\ell \geq 0$,

$$\left|E_{\ell,x}\left[\left(\triangle_\ell^{\ell+1} X_h\right)^m - \left(\triangle_{\ell h}^{(\ell+1)h} X\right)^m\right]\right| \leq K h^{p+1} \text{ for every } m \leq 2p+1, \tag{8.15}$$

and

$$E_{\ell,x}\left[\left|\left(\triangle_\ell^{\ell+1}X_h\right)^{2p+2}\right|\right] \le Kh^{p+1} \tag{8.16}$$

(in these expectations we assume that $X^{(\ell h)} = X_h^{(\ell)} = x$) then, for some $t$-dependent constant $C_t$,

$$\left|E_x\left[f(X_h^{(k)})\right] - E_x\left[f(X^{(t)})\right]\right| \le C_t h^p \tag{8.17}$$

where $h = t/k$ and $p$ is referred to as the order of weak accuracy of the scheme. Convergence of the moments of the discrete process $X_h^{(k)}$ to the corresponding moments of $X^{(t)}$ is called weak convergence. The expression in (8.15) is called a consistency condition. It measures the contribution to the error at each time step. The loss of one order of accuracy in going from the single time step error bound in (8.15) to the final error bound over $\mathcal{O}(h^{-1})$ time steps is typical.

Just as in Section 5.1 we justify the claim above by noticing that, if $u(r,x) = E_{r,x}\left[f(X^{(t)})\right]$, then we can write

$$E_x\left[f(X_h^{(k)})\right] - E_x\left[f(X^{(t)})\right] = \sum_{s=0}^{k-1} E_x\left[E_{s,X_h^{(s)}}\left[u((s+1)h, X_h^{(s+1)})\right] - u(sh, X_h^{(s)})\right].$$

Now notice that, by the tower property of conditional expectations,

$$u(sh, X_h^{(s)}) = E_{sh,X_h^{(s)}}\left[E_{(s+1)h,X^{((s+1)h)}}\left[f(X^{(t)})\right]\right] = E_{sh,X_h^{(s)}}\left[u((s+1)h, X^{((s+1)h)})\right],$$

so that

$$\begin{aligned} &E_x\left[f(X_h^{(k)})\right] - E_x\left[f(X^{(t)})\right] \\ &= \sum_{s=0}^{k-1} E_x\left[E_{s,X_h^{(s)}}\left[u((s+1)h, X_h^{(s+1)})\right] - E_{sh,X_h^{(s)}}\left[u((s+1)h, X^{((s+1)h)})\right]\right]. \end{aligned}$$

Our assumptions on $b$, $\sigma$, and $f$ imply that $u$ is also bounded and has bounded derivatives. Consequently, upon Taylor expanding both $E_{s,x}\left[u((s+1)h, X_h^{(s+1)})\right]$ and $E_{sh,x}\left[u((s+1)h, X^{((s+1)h)})\right]$ around $(s,x)$ and using (8.15) and (8.16) we find that

$$\left|E_x\left[f(X_h^{(k)})\right] - E_x\left[f(X^{(t)})\right]\right| \le Ckh^{p+1}$$

for some constant $C$. Using the fact that $kh = t$, we arrive at (8.17).

The Euler-Maryuma scheme

$$X_h^{(k+1)} = X_h^{(k)} + hb(kh, X_h^{(k)}) + \sqrt{h}\sigma(kh, X_h^{(k)})\xi^{(k+1)} \tag{8.18}$$

for an independent sequence of random variables $\xi^{(k)}$ with $\mathbf{E}\left[\xi^{(k)}\right] = 0$ and $\mathbf{E}\left[\xi^{(k)}\left(\xi^{(k)}\right)^{\mathrm{T}}\right] = I$ is perhaps the simplest consistent discretization scheme. The reader will recognize the Euler-Maryuma scheme as the basis for many of the overdamped Langevin schemes introduced in Chapter 5. Indeed, by Itô's formula,

$$\begin{aligned} b^{(t)}(X^{(t)}) = b^{(t_0)}(x) &+ \int_{t_0}^{t} \left(\partial_t b^{(s)}(X^{(s)}) + \mathcal{L}^{(s)} b^{(s)}(X^{(s)})\right) ds \\ &+ \int_{t_0}^{t} \nabla b^{(s)}(X^{(s)})\sigma^{(s)}(X^{(s)}) dW^{(s)} \end{aligned} \tag{8.19}$$

and (in 1 dimension)

$$\begin{aligned} \sigma^{(t)}(X^{(t)}) = \sigma^{(t_0)}(x) &+ \int_{t_0}^{t} \left(\partial_t \sigma^{(s)}(X^{(s)}) + \mathcal{L}^{(s)} \sigma^{(s)}(X^{(s)})\right) ds \\ &+ \int_{t_0}^{t} \nabla \sigma^{(s)}(X^{(s)})\sigma^{(s)}(X^{(s)}) dW^{(s)}. \end{aligned} \tag{8.20}$$

Plugging these formulas into the SDE solved by $X^{(t)}$,

$$\begin{aligned} X^{(t)} &= x + \int_{t_0}^{t} b^{(s)}(X^{(s)}) ds + \int_{t_0}^{t} \sigma^{(s)}(X^{(s)}) dW^{(s)} \\ &= x + \int_{t_0}^{t} \end{aligned}$$

we can verify the consistency conditions in (8.15).

**Exercise 84.** *Verify that the Euler-Maruyama scheme is weakly consistent with $p = 1$.*

In fact, we can derive higher order schemes by further Itô-Taylor expanding the functions appearing in the integrals in (8.19) and (8.20) around their values and time and position pair $(t_0, x)$. For example, in 1 dimension and when $b$ and $\sigma$ depend only on position, the scheme

$$\begin{aligned}X_h^{(k+1)} = X_h^{(k)} + hb(X_h^{(k)}) + \sigma(X_h^{(k)})\xi^{(k+1)} + \frac{1}{2}\sigma(X_h^{(k)})\sigma'(X_h^{(k)})\left(\left(\xi^{(k+1)}\right)^2 - h\right)\\ + \frac{1}{2}\left(b'(X_h^{(k)})\sigma(X_h^{(k)}) + b(X_h^{(k)})\sigma'(X_h^{(k)}) + \frac{1}{2}\sigma''(X_h^{(k)})\sigma^2(X_h^{(k)})\right)\xi^{(k+1)}h\\ + \frac{1}{2}\left(b(X_h^{(k)})b'(X_h^{(k)}) + \frac{1}{2}\sigma''(X_h^{(k)})\sigma^2(X_h^{(k)})\right)h^2 \quad (8.21)\end{aligned}$$

where the $\xi^{(k)}$ are independent and satisfy

$$\mathbf{P}\left[\xi^{(k)} = \pm\sqrt{3h}\right] = \frac{1}{6}, \qquad \mathbf{P}\left[\xi^{(k)} = 0\right] = \frac{2}{3},$$

satisfies (8.15) with $p = 2$.

**Exercise 85.** *Show this.*

As you can imagine, schemes such as (8.21) that require derivatives of the drift and diffusion coefficients are not very useful in high dimensions where those derivatives are seldom available. Instead methods that approximate derivatives using finite differences are preferred. Such methods are referred to as Runge-Kutta schemes. For example, again in 1 dimension, the scheme

$$\begin{aligned}X_h^{(k+1)} = X_h^{(k)} + \frac{1}{2}\left(b(X_h^{(k)}) + b(Y)\right)h + \frac{1}{4}\left(\sigma(Y^+) + \sigma(Y^-) + 2\sigma(X_h^{(k)})\right)\xi^{(k+1)}\\ + \frac{1}{4\sqrt{h}}\left(\sigma(Y^+) - \sigma(Y^-)\right)\left(\left(\xi^{(k+1)}\right)^2 - h\right) \quad (8.22)\end{aligned}$$

where, at each iteration,

$$Y = X_h^{(k)} + hb(X_h^{(k)}) + \sigma(X_h^{(k)})\xi^{(k+1)}$$

and

$$Y^{\pm} = X_h^{(k)} + hb(X_h^{(k)}) \pm \sigma(X_h^{(k)})\sqrt{h},$$

satisfies (8.15) with $p = 2$.

**Exercise 86.** *Show this when $\sigma$ is constant.*

Occasionally one is interested in the accuracy of the path of the discrete approximation of an SDE given a particular path of the driving Brownian motion. In this setup the scheme is constructed on the same probability space as the Brownian motion and it makes sense to asses the order of strong accuracy, $p$, defined by the relationship

$$E_x\left[\sup_{\ell\le k}|X_h^{(\ell)} - X^{(\ell h)}|\right] \le C_t h^p. \tag{8.23}$$

Again the simplest example is the Euler-Maruyama scheme which we now write as

$$X_h^{(k+1)} = X_h^{(k)} + hb^{(kh)}(X_h^{(k)}) + \sqrt{h}\sigma^{(kh)}(X_h^{(k)})\triangle_{kh}^{(k+1)h}W. \tag{8.24}$$

Notice that we have replaced the more general $\xi^{(k)}$ in (8.24) by increments of the same underlying Brownian motion appearing in the equation solved by $X^{(t)}$. We have already seen that this scheme is weakly accurate with order $p = 1$. It has order of strong accuracy only $p = 1/2$. It is typical that obtaining high order strong accuracy scheme is more difficult than deriving high order weak accuracy schemes. Fortunately, in most applications, the relevant notion of accuracy is weak accuracy.

Because obtaining high order schemes involves generating random variables with distributional properties of high order (repeated) Itô integrals, they can become very complicated very quickly. Moreover, in most applications involving diffusion models of physical processes, the model is not an accurate representation of the underlying process to warrant a high order approximation. Instead, the stability of the scheme becomes a limiting characteristic. By stability we mean that, if the solution to the SDE does not blow up, then neither does the discrete approximation. In most cases stability is discussed in terms of the geometric Brownian motion,

$$X^{(t)} = e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W^{(t)}}x$$

which we have seen solves the equation

$$dX^{(t)} = \mu X^{(t)}dt + \sigma X^{(t)}dW^{(t)}, \qquad X^{(0)} = x.$$

Since the process

$$e^{-\frac{\sigma^2}{2}t + \sigma W^{(t)}}$$

is a martingale (in fact, it is the change of measure appearing in Girsanov's formula), the geometric Brownian motion satisfies

$$E_x\left[X^{(t)}\right] = e^{\mu t}x.$$

Similarly, one can compute

$$E_x\left[\left(X^{(t)}\right)^2\right] = e^{(2\mu+\sigma^2)t}x^2$$

so that the second moment remains stable only when

$$2\mu + \sigma^2 \leq 0. \tag{8.25}$$

The Euler-Maruyama approximation of the geometric Brownian motion is

$$X_h^{(k+1)} = \left(1 + h\mu + \sqrt{h}\sigma\xi^{(k+1)}\right) X_h^{(k)}$$

which satisfies

$$E_x\left[|X_h^{(k)}|^2\right] = \left(1 + h^2\mu^2 + h(2\mu+\sigma^2)\right)^k x^2.$$

The region of linear stability for the scheme is the set of $h$ for which the quantity in the last display converges to zero when (8.25) holds. In other words, assuming (8.25), the region of linear stability is

$$\left\{h :\ 1 + h^2\mu^2 + h(2\mu+\sigma^2) < 1\right\}.$$

Fortunately, under (8.25) we know that the region of linear stability is non-empty (because $h^2\mu^2$ will be smaller than the other terms when $h$ is small enough). Unfortunately, if $\mu$ is large or if $|2\mu+\sigma^2|$ is very small, then the region of linear stability may be very small. In other words, we may need to choose an extremely small step size $h$. As a consequence, to approximate the solution to the SDE on the same physical time scale we will need to take many more steps of the approximate scheme, which implies many more evaluations of the drift and diffusion coefficients (typically the major expense in approximating the solution to an SDE). When this is the case it is sometimes useful to use an implicit scheme such as an implicit $\theta$ method which generalizes the Euler–Maruyama scheme to

$$X_h^{(k+1)} = X_h^{(k)} + h\left(\theta b^{((k+1)h)}(X_h^{(k+1)}) + (1-\theta)b^{(kh)}(X_h^{(k)})\right) + \sqrt{h}\sigma(kh, X_h^{(k)})\xi^{(k+1)}$$

for $\theta \in [0,1]$. When $\theta = 0$ the scheme is our familiar "explicit" Euler–Maruyama scheme and when $\theta = 1$ it is called the "implicit" Euler–Maruyama scheme. Both of these schemes are weakly accurate with order $p = 1$. The implicit midpoint scheme corresponds to $\theta = 1/2$ and is weakly accurate with order $p = 2$ when $\sigma$ is constant. The advantage of the $\theta$ methods is that, if $\theta \geq 1/2$, their region of linear stability is the entire positive half line as can be seen by observing that, when applied to the geometric Brownian motion, the $\theta$ method approximation becomes

$$X_h^{(k+1)} = \frac{1 + h(1-\theta)\mu + \sqrt{h}\sigma\xi^{(k+1)}}{1 - h\theta\mu} X_h^{(k)}$$

and proceeding as before.

**Exercise 87.** *Show that the $\theta$ method with $\theta \geq 1/2$ is unconditionally linearly stable.*

## 8.5 General diffusion limits and noisy coefficients

The last section and Section 5.1 were related to the convergence of a discrete time Markov process to the solution of a stochastic differential equation. In the last section the goal was to show that a discrete time process was an accurate approximation of the the solution to (**??**). But it is often interesting to consider the diffusion limit of some sequence of stochastic processes $X_n^{(t)}$ even when our primary interest is in the $X_n^{(t)}$ themselves and not the limiting process, because knowledge of the limiting process can provide useful qualitative information about the $X_n^{(t)}$.

In this section we will consider diffusion limits of sequences of continuous time stochastic (not necessarily Markovian) processes. We will assume that the process $X_n^{(t)} \in \mathbb{R}^d$ is adapted to a filtration $\mathcal{F}_t^n$ and has samples paths that are (with probability 1) r.c.l.l, i.e. that are right continuous (i.e. $\lim_{s\to t^+} X_n^{(s)} = X_n^{(t)}$) and have left hand limits (i.e. $X^{(t^-)} = \lim_{s\to t^-} X_n^{(s)}$ exists). We will also assume that, for each $n$, there is a process $B_n^{(t)} \in \mathbb{R}^d$ and a $d \times d$ matrix valued process $A_n^{(t)}$, both of which are r.c.l.l. and $\mathcal{F}_t^n$-adapted, such that for any $t > s \geq 0$, $A^{(t)} - A^{(s)}$ is positive semi-definite. Moreover, suppose that, for each $n$,

$$M_n^{(t)} = X_n^{(t)} - B_n^{(t)} \quad \text{and} \quad M_n^{(t)} \left(M_n^{(t)}\right)^{\mathrm{T}} - A_n^{(t)}$$

are both $\mathcal{F}_t^n$ martingales. For simplicity we will also assume that, with probability 1, the $X_n^{(t)}$ remain in some bounded set for all time. We will consider the convergence of the sequence $X_n^{(t)}$ to the solution $X^{(t)}$ of the SDE

$$dX^{(t)} = b(X^{(t)})dt + \sigma(X^{(t)})dW^{(t)}, \quad X^{(0)} = x \tag{8.26}$$

which we assume has continuous drift and diffusion coefficients.

Before giving the conditions guaranteeing convergence to $X^{(t)}$, observe that, since $X^{(t)}$ solves (8.26), and if

$$B^{(t)} = \int_0^t b(X^{(s)})ds$$

then $B^{(t)}$ is continuous (and hence r.c.l.l.) and

$$M^{(t)} = X^{(t)} - B^{(t)} = x + \int_0^t \sigma(X^{(s)})dW^{(s)}$$

is continuous a martingale. Moreover, by Itô's product rule,

$$M^{(t)} \left(M^{(t)}\right)^{\mathrm{T}} - [M, M]^{(t)}$$

is a martingale. Finally, observe that

$$A^{(t)} = [M, M]^{(t)} = \int_0^t a(X^{(s)})ds$$

has positive semi-definite increments. It will not come as a complete surprise to the reader then that the final conditions needed to guarantee convergence of $X_n^{(t)}$ to $X^{(t)}$ (in law) include conditions concerning the limiting continuity of $X_n^{(t)}$ and the convergence of $B_n^{(t)}$ to $B^{(t)}$ and $A_n^{(t)}$ to $A^{(t)}$. The general theorem on convergence of $X_n$ to $X$ state that if, for each $t \geq 0$,

$$\lim_{n\to\infty} E_x\left[\sup_{s\leq t}\|X_n^{(s)} - X_n^{(s^-)}\|_2^2\right] = 0, \tag{8.27}$$

$$\lim_{n\to\infty} E_x\left[\sup_{s\leq t}\|B_n^{(s)} - B_n^{(s^-)}\|_2^2\right] = 0, \tag{8.28}$$

$$\lim_{n\to\infty} E_x\left[\sup_{s\leq t}\left|\left(A_n^{(s)}\right)_{ij} - \left(A_n^{(s^-)}\right)_{ij}\right|\right] = 0, \tag{8.29}$$

$$\sup_{s\leq t}\left|\left(B_n^{(s)}\right)_i - \int_0^s b_i(X_n^{(r)})dr\right| \to^P 0, \tag{8.30}$$

and

$$\sup_{s\leq t}\left|\left(A_n^{(s)}\right)_{ij} - \int_0^s a_{ij}(X_n^{(r)})dr\right| \to^P 0, \tag{8.31}$$

(here $\to^P$ means convergence in probability) then $X_n$ converges in distribution to the weak solution of the stochastic differential equation Conditions (8.27), (8.28), and (8.29), all concern the continuity (in the limit) of the processes $X_n$, $B_n$, and $A_n$. Conditions (8.30) and (8.31) concern the mean drift and quadratic variation of $X_n$ in the limit.

One could use this theorem to show that, for example, the Euler iteration converges to solutions of (8.26). But the theorem would not provide a rate of convergence (how small is $X_h^{(k)} - X^{(t)}$ in terms of powers of $h$) and we would not be fully exploiting its generality. Instead we will use the theorem to show convergence of the sequence of diffusions

$$dX_h^{(t)} = b(X_h^{(t)}, \eta^{(k+1)})dt + \sigma(X_h^{(t)}, \mu^{(k+1)})dW^{(t)}, \quad t \in [hk, h(k+1)), \tag{8.32}$$

where, for each $k$, $\eta^{(k)}$ and $\mu^{(k)}$ are independent (of each other, of $W$, and of $\mathcal{F}_{h(k-1)}$), identically distributed sequences of random variables, to the solution of the equation

$$dX^{(t)} = \hat{b}(X^{(t)})dt + \hat{\sigma}(X^{(t)})dW^{(t)} \tag{8.33}$$

where

$$\hat{b}(x) = \mathbf{E}\left[b(x, \eta^{(1)})\right] \quad \text{and} \quad \hat{\sigma}(x)\left(\hat{\sigma}(x)\right)^{\mathrm{T}} = \mathbf{E}\left[a(x, \mu^{(1)})\right].$$

We will assume that $b$, $\sigma$, $\hat{b}$, and $\hat{\sigma}$, are bounded and that all of these functions are continuos in their $x$ arguments.

Before we apply this section's limit theorem, observe that, even with stronger assumptions on the smoothness of the various coefficients we could not apply the limit theorem for numerical approximations introduced in the last section because $E_{\ell h,x}\left[\left(\triangle_{\ell h}^{(\ell+1)h} X_h\right)^2\right]$ is not a sufficiently accurate approximation of $E_{\ell h,x}\left[\left(\triangle_{\ell h}^{(\ell+1)h} X\right)^2\right]$.

**Exercise 88.** *Check this by assuming that $b$ and $\sigma$ have no dependence on $x$ and comparing the moments of the increments of* (8.32) *and* (8.33) *as in* (8.15) *with $p = 1$.*

To apply the diffusion limit theorem introduced in this section, first notice that, for $t \in [hk, h(k+1))$,

$$X_h^{(t)} = x + \sum_{j=0}^{k-1} \int_{hj}^{h(j+1)} b(X_h^{(s)}, \eta^{(j+1)}) ds + \sum_{j=0}^{k-1} \int_{hj}^{h(j+1)} \sigma(X_h^{(s)}, \mu^{(j+1)}) dW^{(s)}$$
$$+ \int_{hj}^{t} b(X_h^{(s)}, \eta^{(k+1)}) ds + \int_{hj}^{t} \sigma(X_h^{(s)}, \mu^{(k+1)}) dW^{(s)}.$$

If, for $t \in [hk, h(k+1))$, we define

$$B_h^{(t)} = \sum_{j=0}^{k-1} \int_{hj}^{h(j+1)} b(X_h^{(s)}, \eta^{(j+1)}) ds + \int_{hj}^{t} b(X_h^{(s)}, \eta^{(k+1)}) ds$$

then the process

$$M_h^{(t)} = X_h^{(t)} - B_h^{(t)} = \sum_{j=0}^{k-1} \int_{hj}^{h(j+1)} \sigma(X_h^{(s)}, \mu^{(j+1)}) dW^{(s)} + \int_{hj}^{t} \sigma(X_h^{(s)}, \mu^{(k+1)}) dW^{(s)}$$

is a martingale. We can also let

$$A_h^{(t)} = [M_h, M_h]^{(t)}$$

and note that

$$A_h^{(t)} = \sum_{j=0}^{k-1} \int_{jh}^{(j+1)h} a(X_h^{(s)}, \mu^{(j+1)}) ds + \int_{kh}^{t} a(X_h^{(s)}, \mu^{(k+1)}) ds$$

when $t \in [hk, h(k+1))$ and where $a = \sigma\sigma^{\mathrm{T}}$.

**Exercise 89.** *Establish the formula for $A^{(t)}$ in the last dispaly.*

Since $a$ is symmetric and positive semi-definite, $A_h^{(t)}$ has the required properties to apply the general diffusion limit theorem above. Moreover, $X_h$, $B_h$, and $A_h$ are all continuous so that conditions (8.27), (8.28), and (8.29), are all satisfied automatically.

For $s \in [hj, h(j+1))$, we can write

$$B_h^{(s)} - \int_0^s \hat{b}(X_h^{(r)})dr = \sum_{\ell=0}^{j-1} \int_{\ell h}^{(\ell+1)h} Z_h^{(r)} dr + \int_{jh}^{s} Z_h^{(r)} dr$$

where, for $r \in [\ell h, (\ell+1)h)$,

$$Z_h^{(r)} = b(X_h^{(r)}, \eta^{(\ell+1)}) - \hat{b}(X_h^{(r)}).$$

Now define the martingale

$$J_h^{(j)} = \sum_{\ell=0}^{j-1} \int_{\ell h}^{(\ell+1)h} Z_h^{(r)} dr$$

and, if $t \in [kh, (k+1)h)$, the remainder term

$$R_h = \sup_{j \le k} \int_{jh}^{(j+1)h} \|Z_h^{(r)}\|_2 dr.$$

Since both $b$ and $\hat{b}$ are assumed to be bounded, $R_h \le Ch$ for some constant $C$. As a consequence, for $s \in [hj, h(j+1))$,

$$\left\| B_h^{(s)} - \int_0^s \hat{b}(X_h^{(r)})dr \right\|_2 \le \|J_h^{(j)}\|_2 + Ch.$$

Since $J_h^{(j)}$ is a martingale, Doob's maximal inequality implies that

$$\sqrt{\mathbf{E}\left[\sup_{j \le k} \|J_h^{(j)}\|_2^2\right]} \le 2\sqrt{\mathbf{E}\left[\|J_h^{(k)}\|_2^2\right]}$$

and, since if $r < u$,

$$\mathbf{E}\left[Z_h^{(r)} Z_h^{(u)}\right] = \mathbf{E}\left[\mathbf{E}\left[Z_h^{(r)} Z_h^{(u)} \mid \mathcal{F}_u\right]\right] = \mathbf{E}\left[Z_h^{(r)} \mathbf{E}\left[Z_h^{(u)} \mid \mathcal{F}_u\right]\right] = 0, \quad (8.34)$$

$$\mathbf{E}\left[\|J_h^{(k)}\|_2^2\right] = \sum_{j=0}^{k-1} \mathbf{E}\left[\left(\int_{jh}^{(j+1)h} Z_h^{(r)} dr\right)^2\right] \leq C^2 k h^2.$$

Putting these bounds together we find that

$$\mathbf{E}\left[\sup_{s\leq t}\left\|B_h^{(s)} - \int_0^s \hat{b}(X_h^{(r)})dr\right\|_2^2\right] \leq Ch\left(\sqrt{k}+1\right).$$

Finally, noting that $k \leq t/h$, the bound becomes $C\sqrt{h}\left(\sqrt{t}+\sqrt{h}\right)$ which goes to zero when $h$ goes to zero. Consequently, (8.30) is satisfied. A similar argument shows that (8.31) is also satisfied and we conclude that the solution to the SDE (8.32) with noisy coefficients converges in distribution to the solution to the SDE (8.33).

This argument shows that, despite the fact that $b(x, \xi^{(\ell)})$ will typically not be close to to $\hat{b}$, over many time intervals the errors $Z_h^{(r)} = b(X_h^{(r)}, \eta^{(\ell+1)}) - \hat{b}(X_h^{(r)})$ (which have mean zero) have a chance to "average away." This is an example of a phenomenon called dynamic self averaging and is closely related to the averaging results we will use in Section **??** of the next chapter to reduce the dimension of a system of SDE. In that section however, we will use more problem specific arguments that yield stronger conclusions in the specific context considered there.

## 8.6 Bibliography