Chapter 8
Stochastic Differential Equations

This chapter is devoted to stochastic differential equations, which motivated Itô's construction of stochastic integrals. After giving the general definitions, we provide a detailed treatment of the Lipschitz case, where strong existence and uniqueness statements hold. Still in the Lipschitz case, we show that the solution of a stochastic differential equation is a Markov process with a Feller semigroup, whose generator is a second-order differential operator. By results of Chap. 6, the Feller property immediately gives the strong Markov property of solutions of stochastic differential equations. The last section presents a few important examples. This chapter can be read independently of Chap. 7.

8.1 Motivation and General Definitions

The goal of stochastic differential equations is to provide a model for a differential equation perturbed by a random noise. Consider an ordinary differential equation of the form

$$y'(t) = b(y(t)),$$

or, in differential form,

$$\mathrm{d}y_t = b(y_t)\,\mathrm{d}t.$$

Such an equation is used to model the evolution of a physical system. If we take random perturbations of the system into account, we add a noise term, which is typically of the form $\sigma\,\mathrm{d}B_t$, where B denotes a Brownian motion, and σ is a constant corresponding to the intensity of the noise. Note that the use of Brownian motion here is justified by its property of independence of increments, corresponding to the

J.-F. Le Gall, *Brownian Motion, Martingales, and Stochastic Calculus*,
Graduate Texts in Mathematics 274, DOI 10.1007/978-3-319-31089-3_8

fact that the random perturbations affecting disjoint time intervals are assumed to be independent.

In this way, we arrive at a stochastic differential equation of the form

$$\mathrm{d}y_t = b(y_t)\,\mathrm{d}t + \sigma\,\mathrm{d}B_t,$$

or in integral form, the only one with a rigorous mathematical meaning,

$$y_t = y_0 + \int_0^t b(y_s)\,\mathrm{d}s + \sigma\,B_t.$$

We generalize the preceding equation by allowing σ to depend on the state of the system at time t :

$$\mathrm{d}y_t = b(y_t)\,\mathrm{d}t + \sigma(y_t)\,\mathrm{d}B_t,$$

or, in integral form,

$$y_t = y_0 + \int_0^t b(y_s)\,\mathrm{d}s + \int_0^t \sigma(y_s)\,\mathrm{d}B_s.$$

Because of the integral in $\mathrm{d}B_s$, the preceding equation only makes sense thanks to the theory of stochastic integrals developed in Chap. 5. We can still generalize the preceding equation by allowing σ and b to depend on the time parameter t. This leads to the following definition.

Definition 8.1 Let d and m be positive integers, and let σ and b be locally bounded measurable functions defined on $\mathbb{R}_+ \times \mathbb{R}^d$ and taking values in $M_{d\times m}(\mathbb{R})$ and in $\mathbb{R}^d$ respectively, where $M_{d\times m}(\mathbb{R})$ is the set of all $d \times m$ matrices with real coefficients. We write $\sigma = (\sigma_{ij})_{1\le i\le d, 1\le j\le m}$ and $b = (b_i)_{1\le i\le d}$.

A *solution of the stochastic differential equation*

$$E(\sigma, b) \qquad\qquad \mathrm{d}X_t = \sigma(t, X_t)\,\mathrm{d}B_t + b(t, X_t)\,\mathrm{d}t$$

consists of:

- a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\in[0,\infty]}, P)$ (where the filtration is always assumed to be complete);
- an m-dimensional $(\mathscr{F}_t)$-Brownian motion $B = (B^1, \ldots, B^m)$ started from 0;
- an $(\mathscr{F}_t)$-adapted process $X = (X^1, \ldots, X^d)$ with values in $\mathbb{R}^d$, with continuous sample paths, such that

$$X_t = X_0 + \int_0^t \sigma(s, X_s)\,\mathrm{d}B_s + \int_0^t b(s, X_s)\,\mathrm{d}s,$$

meaning that, for every $i \in \{1, \ldots, d\}$,

$$X_t^i = X_0^i + \sum_{j=1}^m \int_0^t \sigma_{ij}(s, X_s)\, \mathrm{d}B_s^j + \int_0^t b_i(s, X_s)\, \mathrm{d}s.$$

If additionally $X_0 = x \in \mathbb{R}^d$, we say that X is a solution of $E_x(\sigma, b)$.

Note that, when we speak about a solution of $E(\sigma, b)$, we do not fix a priori the filtered probability space and the Brownian motion B. When we fix these objects, we will say so explicitly.

There are several notions of existence and uniqueness for stochastic differential equations.

Definition 8.2 For the equation $E(\sigma, b)$ we say that there is

- *weak existence* if, for every $x \in \mathbb{R}^d$, there exists a solution of $E_x(\sigma, b)$;
- *weak existence and weak uniqueness* if in addition, for every $x \in \mathbb{R}^d$, all solutions of $E_x(\sigma, b)$ have the same law;
- *pathwise uniqueness* if, whenever the filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t), P)$ and the $(\mathscr{F}_t)$-Brownian motion B are fixed, two solutions X and X' such that $X_0 = X'_0$ a.s. are indistinguishable.

Furthermore, we say that a solution X of $E_x(\sigma, b)$ is a *strong solution* if X is adapted with respect to the completed canonical filtration of B.

Remark It may happen that weak existence and weak uniqueness hold but pathwise uniqueness fails. For a simple example, consider a real Brownian motion β started from $\beta_0 = y$, and set

$$B_t = \int_0^t \operatorname{sgn}(\beta_s)\, \mathrm{d}\beta_s,$$

where $\operatorname{sgn}(x) = 1$ if $x > 0$ and $\operatorname{sgn}(x) = -1$ if $x \leq 0$. Then, one immediately gets from the "associativity" of stochastic integrals that

$$\beta_t = y + \int_0^t \operatorname{sgn}(\beta_s)\, \mathrm{d}B_s.$$

Moreover, B is a continuous martingale with quadratic variation $\langle B, B \rangle_t = t$, and Theorem 5.12 shows that B is a Brownian motion started from 0. We thus see that β solves the stochastic differential equation

$$\mathrm{d}X_t = \operatorname{sgn}(X_t)\, \mathrm{d}B_t, \qquad X_0 = y,$$

and it follows that weak existence holds for this equation. Theorem 5.12 again shows that any other solution of this equation must be a Brownian motion, which gives weak uniqueness. On the other hand, pathwise uniqueness fails. In fact, taking $y = 0$ in the construction, one easily sees that both β and $-\beta$ solve the preceding stochastic differential equation with the same Brownian motion B and initial value 0 (note that $\int_0^t \mathbf{1}_{\{\beta_s=0\}}\, \mathrm{d}s = 0$, which implies $\int_0^t \mathbf{1}_{\{\beta_s=0\}}\, \mathrm{d}B_s = 0$). One can also show that β is not a strong solution: One verifies that the canonical filtration of B coincides with the canonical filtration of $|\beta|$, which is strictly smaller than that of β (we omit the proof, which is a simple application of formula (9.18) in Chap. 9).

The next theorem links the different notions of existence and uniqueness.

Theorem (Yamada–Watanabe) *If both weak existence and pathwise uniqueness hold, then weak uniqueness also holds. Moreover, for any choice of the filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t), P)$ and of the $(\mathscr{F}_t)$-Brownian motion B, there exists for every $x \in \mathbb{R}^d$ a (unique) strong solution of $E_x(\sigma, b)$.*

We omit the proof (see Yamada and Watanabe [83]) because we will not need this result. In the Lipschitz case that we will consider, we will establish directly the properties given by the Yamada–Watanabe theorem.

8.2 The Lipschitz Case

In this section, we work under the following assumptions.

Assumptions The functions σ are b are continuous on $\mathbb{R}_+ \times \mathbb{R}^d$ and Lipschitz in the variable x: There exists a constant K such that, for every $t \geq 0, x, y \in \mathbb{R}^d$,

$$|\sigma(t,x) - \sigma(t,y)| \leq K\,|x-y|,$$
$$|b(t,x) - b(t,y)| \leq K\,|x-y|.$$

Theorem 8.3 *Under the preceding assumptions, pathwise uniqueness holds for $E(\sigma, b)$, and, for every choice of the filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t), P)$ and of the $(\mathscr{F}_t)$-Brownian motion B, for every $x \in \mathbb{R}^d$, there exists a (unique) strong solution of $E_x(\sigma, b)$.*

The theorem implies in particular that weak existence holds for $E(\sigma, b)$. Weak uniqueness will follow from the next theorem (it can also be deduced from pathwise uniqueness using the Yamada–Watanabe theorem).

Remark One can "localize" the Lipschitz assumption on σ and b, meaning that the constant K may depend on the compact set on which the parameters t and x, y are considered. In that case, it is, however, necessary to keep a condition of linear growth of the form

$$|\sigma(t,x)| \leq K(1+|x|), \quad |b(t,x)| \leq K(1+|x|).$$

This kind of condition, which avoids the blow-up of solutions, already appears in ordinary differential equations.

Proof For the sake of simplicity, we consider only the case $d = m = 1$. The reader will be able to check that the general case follows from exactly the same arguments, at the cost of a heavier notation. Let us start by proving pathwise uniqueness. We consider (on the same filtered probability space, with the same Brownian motion B) two solutions X and X' such that $X_0 = X'_0$. Fix $M > 0$ and set

$$\tau = \inf\{t \geq 0 : |X_t| \geq M \text{ or } |X'_t| \geq M\}.$$

Then, for every $t \geq 0$,

$$X_{t\wedge\tau} = X_0 + \int_0^{t\wedge\tau} \sigma(s, X_s)\, \mathrm{d}B_s + \int_0^{t\wedge\tau} b(s, X_s)\, \mathrm{d}s$$

and an analogous equation holds for $X'_{t\wedge\tau}$. Fix a constant $T > 0$. By considering the difference between the two equations and using the bound (5.14), we get, for $t \in [0, T]$,

$$\begin{aligned}
&E[(X_{t\wedge\tau} - X'_{t\wedge\tau})^2] \\
&\leq 2E\Big[\Big(\int_0^{t\wedge\tau} (\sigma(s, X_s) - \sigma(s, X'_s))\, \mathrm{d}B_s\Big)^2\Big] + 2E\Big[\Big(\int_0^{t\wedge\tau} (b(s, X_s) - b(s, X'_s))\, \mathrm{d}s\Big)^2\Big] \\
&\leq 2\left(E\Big[\int_0^{t\wedge\tau} (\sigma(s, X_s) - \sigma(s, X'_s))^2 \mathrm{d}s\Big] + T\, E\Big[\int_0^{t\wedge\tau} (b(s, X_s) - b(s, X'_s))^2 \mathrm{d}s\Big]\right) \\
&\leq 2K^2(1 + T)\, E\Big[\int_0^{t\wedge\tau} (X_s - X'_s)^2 \mathrm{d}s\Big] \\
&\leq 2K^2(1 + T)\, E\Big[\int_0^{t} (X_{s\wedge\tau} - X'_{s\wedge\tau})^2 \mathrm{d}s\Big].
\end{aligned}$$

Hence the function $h(t) = E[(X_{t\wedge\tau} - X'_{t\wedge\tau})^2]$ satisfies

$$h(t) \leq C \int_0^t h(s)\, \mathrm{d}s$$

for every $t \in [0, T]$, with $C = 2K^2(1 + T)$.

Lemma 8.4 (Gronwall's lemma) *Let $T > 0$ and let g be a nonnegative bounded measurable function on $[0, T]$. Assume that there exist two constants $a \geq 0$ and $b \geq 0$ such that, for every $t \in [0, T]$,*

$$g(t) \leq a + b \int_0^t g(s)\, \mathrm{d}s.$$

Then, we also have, for every $t \in [0, T]$,

$$g(t) \le a \exp(bt).$$

Proof of the lemma By iterating the condition on g, we get,

$$\begin{aligned} g(t) &\le a + a(bt) + b^2 \int_0^t \mathrm{d}s \int_0^s \mathrm{d}r\, g(r) \\ &\le a + a(bt) + a\frac{(bt)^2}{2} + \cdots + a\frac{(bt)^n}{n!} + b^{n+1} \int_0^t \mathrm{d}s_1 \int_0^{s_1} \mathrm{d}s_2 \cdots \int_0^{s_n} \mathrm{d}s_{n+1} g(s_{n+1}), \end{aligned}$$

for every $n \ge 1$. If A is a constant such that $0 \le g \le A$, the last term in the right-hand side is bounded above by $A(bt)^{n+1}/(n+1)!$, hence tends to 0 as $n \to \infty$. The desired result now follows. □

Let us return to the proof of the theorem. The function h is bounded above by $4M^2$ and the assumption of the lemma holds with $a = 0, b = C$. We thus get $h = 0$, so that $X_{t\wedge\tau} = X'_{t\wedge\tau}$. By letting M tend to ∞, we get $X_t = X'_t$, which completes the proof of pathwise uniqueness.

For the second assertion, we construct a solution using Picard's approximation method. We define by induction

$$\begin{aligned} X_t^0 &= x, \\ X_t^1 &= x + \int_0^t \sigma(s, x)\, \mathrm{d}B_s + \int_0^t b(s, x)\, \mathrm{d}s, \\ X_t^n &= x + \int_0^t \sigma(s, X_s^{n-1})\, \mathrm{d}B_s + \int_0^t b(s, X_s^{n-1})\, \mathrm{d}s. \end{aligned}$$

The stochastic integrals are well defined since one verifies by induction that, for every n, the process X^n is adapted and has continuous sample paths.

It is enough to show that, for every $T > 0$, there is a strong solution of $E_x(\sigma, b)$ on the time interval $[0, T]$. Indeed, the uniqueness part of the argument will then allow us to get a (unique) strong solution on $\mathbb{R}_+$ that will coincide with the solution on $[0, T]$ up to time T.

We fix $T > 0$ and, for every $n \ge 1$ and every $t \in [0, T]$, we set

$$g_n(t) = E\Big[\sup_{0\le s\le t} |X_s^n - X_s^{n-1}|^2\Big].$$

We will bound the functions g_n by induction on n (at present, it is not yet clear that these functions are finite). The fact that the functions $\sigma(\cdot, x)$ and $b(\cdot, x)$ are continuous, hence bounded, over $[0, T]$ implies that there exists a constant C'_T such that $g_1(t) \le C'_T$ for every $t \in [0, T]$ (use Doob's inequality in L^2 for the stochastic integral term).

Then we observe that

$$X_t^{n+1} - X_t^n = \int_0^t (\sigma(s, X_s^n) - \sigma(s, X_s^{n-1}))\,\mathrm{d}B_s + \int_0^t (b(s, X_s^n) - b(s, X_s^{n-1}))\,\mathrm{d}s.$$

Hence, using the case $p = 2$ of the Burkholder–Davis–Gundy inequalities in the second bound (and writing $C_{(2)}$ for the constant in this inequality),

$$\begin{aligned} E\Big[\sup_{0\le s\le t} |X_s^{n+1} - X_s^n|^2\Big] &\le 2\,E\Big[\sup_{0\le s\le t}\Big|\int_0^s (\sigma(u, X_u^n) - \sigma(u, X_u^{n-1}))\,\mathrm{d}B_u\Big|^2 \\ &\qquad + \sup_{0\le s\le t}\Big|\int_0^s (b(u, X_u^n) - b(u, X_u^{n-1}))\,\mathrm{d}u\Big|^2\Big] \\ &\le 2\Big(C_{(2)}\,E\Big[\int_0^t (\sigma(u, X_u^n) - \sigma(u, X_u^{n-1}))^2\,\mathrm{d}u\Big] \\ &\qquad + T\,E\Big[\int_0^t (b(u, X_u^n) - b(u, X_u^{n-1}))^2\,\mathrm{d}u\Big]\Big) \\ &\le 2(C_{(2)} + T)K^2\,E\Big[\int_0^t |X_u^n - X_u^{n-1}|^2\,\mathrm{d}u\Big] \\ &\le C_T\,E\Big[\int_0^t \sup_{0\le r\le u} |X_r^n - X_r^{n-1}|^2\,\mathrm{d}u\Big] \end{aligned}$$

where $C_T = 2(C_{(2)} + T)K^2$. We have thus obtained that, for every $n \ge 1$,

$$g_{n+1}(t) \le C_T \int_0^t g_n(u)\,\mathrm{d}u. \tag{8.1}$$

Recalling that $g_1(t) \le C'_T$, an induction argument using (8.1) shows that, for every $n \ge 1$ and $t \in [0, T]$,

$$g_n(t) \le C'_T (C_T)^{n-1} \frac{t^{n-1}}{(n-1)!}.$$

In particular, $\sum_{n=1}^{\infty} g_n(T)^{1/2} < \infty$, which implies that

$$\sum_{n=0}^{\infty} \sup_{0\le t\le T} |X_t^{n+1} - X_t^n| < \infty, \quad \text{a.s.}$$

Hence the sequence $(X_t^n, 0 \le t \le T)$ converges uniformly on $[0, T]$, a.s., to a limiting process $(X_t, 0 \le t \le T)$, which has continuous sample paths. By induction, one also verifies that, for every n, X^n is adapted with respect to the (completed) canonical filtration of B, and the same holds for X.

Finally, from the fact that σ and b are Lipschitz in the variable x, we also get that, for every $t \in [0, T]$,

$$\lim_{n\to\infty}\Big(\int_0^t \sigma(s, X_s)\,\mathrm{d}B_s - \int_0^t \sigma(s, X_s^n)\,\mathrm{d}B_s\Big) = 0,$$
$$\lim_{n\to\infty}\Big(\int_0^t b(s, X_s)\,\mathrm{d}s - \int_0^t b(s, X_s^n)\,\mathrm{d}s\Big) = 0,$$

in probability (to deal with the stochastic integrals, we may use Proposition 5.8, noting that $|X_s^n - X_s|$ is dominated by $\sum_{k=0}^{\infty} \sup_{0\le r\le s} |X_r^{k+1} - X_r^k|$). By passing to the limit in the induction equation defining X^n, we get that X solves $E_x(\sigma, b)$ on $[0, T]$. This completes the proof of the theorem. □

In the following statement, $W(\mathrm{dw})$ stands for the Wiener measure on the canonical space $C(\mathbb{R}_+, \mathbb{R}^m)$ of all continuous functions from $\mathbb{R}_+$ into $\mathbb{R}^m$ ($W(\mathrm{dw})$ is the law of $(B_t, t \ge 0)$ if B is an m-dimensional Brownian motion started from 0).

Theorem 8.5 *Under the assumptions of the preceding theorem, there exists, for every $x \in \mathbb{R}$, a mapping $F_x : C(\mathbb{R}_+, \mathbb{R}^m) \longrightarrow C(\mathbb{R}_+, \mathbb{R}^d)$, which is measurable when $C(\mathbb{R}_+, \mathbb{R}^m)$ is equipped with the Borel σ-field completed by the W-negligible sets, and $C(\mathbb{R}_+, \mathbb{R}^d)$ is equipped with the Borel σ-field, such that the following properties hold:*

(i) *for every $t \ge 0$, $F_x(\mathrm{w})_t$ coincides $W(\mathrm{dw})$ a.s. with a measurable function of $(\mathrm{w}(r), 0 \le r \le t)$;*
(ii) *for every $\mathrm{w} \in C(\mathbb{R}_+, \mathbb{R}^m)$, the mapping $x \mapsto F_x(\mathrm{w})$ is continuous;*
(iii) *for every $x \in \mathbb{R}^d$, for every choice of the (complete) filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t), P)$ and of the m-dimensional $(\mathscr{F}_t)$-Brownian motion B with $B_0 = 0$, the process X_t defined $X_t = F_x(B)_t$ is the unique solution of $E_x(\sigma, b)$; furthermore, if U is an $\mathscr{F}_0$-measurable real random variable, the process $F_U(B)_t$ is the unique solution with $X_0 = U$.*

Remark Assertion (iii) implies in particular that weak uniqueness holds for $E(\sigma, b)$: any solution of $E_x(\sigma, b)$ must be of the form $F_x(B)$ and its law is thus uniquely determined as the image of $W(\mathrm{dw})$ under F_x.

Proof Again we consider only the case $d = m = 1$. Let $\mathscr{N}$ be the class of all W-negligible sets in $C(\mathbb{R}_+, \mathbb{R})$, and, for every $t \in [0, \infty]$, set

$$\mathscr{G}_t = \sigma(\mathrm{w}(s), 0 \le s \le t) \vee \mathscr{N}.$$

For every $x \in \mathbb{R}$, we write X^x for the solution of $E_x(\sigma, b)$ corresponding to the filtered probability space $(C(\mathbb{R}_+, \mathbb{R}), \mathscr{G}_\infty, (\mathscr{G}_t), W)$ and the (canonical) Brownian motion $B_t(\mathrm{w}) = \mathrm{w}(t)$. This solution exists and is unique (up to indistinguishability) by Theorem 8.3, noting that the filtration $(\mathscr{G}_t)$ is complete by construction.

Let $x, y \in \mathbb{R}$ and let T_n be the stopping time defined by

$$T_n = \inf\{t \geq 0 : |X_t^x| \geq n \text{ or } |X_t^y| \geq n\}.$$

Let $p \geq 2$ and $T \geq 1$. Using the Burkholder–Davis–Gundy inequalities (Theorem 5.16) and then the Hölder inequality, we get, for $t \in [0, T]$,

$$\begin{aligned}
&E\Big[\sup_{s\leq t} |X^x_{s\wedge T_n} - X^y_{s\wedge T_n}|^p\Big] \\
&\quad\leq C_p\Big(|x-y|^p + E\Big[\sup_{s\leq t}\Big|\int_0^{s\wedge T_n}(\sigma(r,X_r^x)-\sigma(r,X_r^y))\,\mathrm{d}B_r\Big|^p\Big] \\
&\qquad + E\Big[\sup_{s\leq t}\Big|\int_0^{s\wedge T_n}(b(r,X_r^x)-b(r,X_r^y))\,\mathrm{d}r\Big|^p\Big]\Big) \\
&\quad\leq C_p\Big(|x-y|^p + C_p' E\Big[\Big(\int_0^{t\wedge T_n}(\sigma(r,X_r^x)-\sigma(r,X_r^y))^2\mathrm{d}r\Big)^{p/2}\Big] \\
&\qquad + E\Big[\Big(\int_0^{t\wedge T_n}|b(r,X_r^x)-b(r,X_r^y)|\,\mathrm{d}r\Big)^p\Big]\Big) \\
&\quad\leq C_p\Big(|x-y|^p + C_p' t^{\frac{p}{2}-1} E\Big[\int_0^t |\sigma(r\wedge T_n, X^x_{r\wedge T_n}) - \sigma(r\wedge T_n, X^y_{r\wedge T_n})|^p\mathrm{d}r\Big] \\
&\qquad + t^{p-1}\, E\Big[\int_0^t |b(r\wedge T_n, X^x_{r\wedge T_n}) - b(r\wedge T_n, X^y_{r\wedge T_n})|^p\mathrm{d}r\Big]\Big) \\
&\quad\leq C_p''\Big(|x-y|^p + T^p\int_0^t E[|X^x_{r\wedge T_n} - X^y_{r\wedge T_n}|^p]\,\mathrm{d}r\Big),
\end{aligned}$$

where the constants $C_p, C_p', C_p'' < \infty$ depend on p (and on the constant K appearing in our assumption on σ and b) but not on n or on x, y and T.

As the function $t \mapsto E\Big[\sup_{s\leq t} |X^x_{s\wedge T_n} - X^y_{s\wedge T_n}|^p\Big]$ is bounded, Lemma 8.4 implies that, for $t \in [0, T]$,

$$E\Big[\sup_{s\leq t} |X^x_{s\wedge T_n} - X^y_{s\wedge T_n}|^p\Big] \leq C_p''|x-y|^p \, \exp(C_p'' T^p t),$$

hence, letting n tend to ∞,

$$E\Big[\sup_{s\leq t} |X^x_s - X^y_s|^p\Big] \leq C_p''|x-y|^p \, \exp(C_p'' T^p t).$$

The topology on the space $C(\mathbb{R}_+, \mathbb{R})$ is defined by the distance

$$\mathbf{d}(\mathrm{w}, \mathrm{w}') = \sum_{k=1}^{\infty} \alpha_k \Big(\sup_{s\leq k} |\mathrm{w}(s) - \mathrm{w}'(s)| \wedge 1\Big),$$

where the sequence of positive reals α_k can be chosen in an arbitrary way, provided that the series $\sum \alpha_k$ converges. We may choose the coefficients α_k so that

$$\sum_{k=1}^{\infty} \alpha_k \exp(C''_p k^{p+1}) < \infty.$$

For every $x \in \mathbb{R}$, we consider X^x as a random variable with values in $C(\mathbb{R}_+, \mathbb{R})$. The preceding estimates and Jensen's inequality then show that

$$E[\mathbf{d}(X^x, X^y)^p] \leq \Big(\sum_{k=1}^{\infty} \alpha_k\Big)^{p-1} \sum_{k=1}^{\infty} \alpha_k\, E\Big[\sup_{s\leq k} |X_s^x - X_s^y|^p\Big] \leq \bar{C}_p\, |x-y|^p,$$

with a constant $\bar{C}_p$ independent of x and y. By Kolmogorov's lemma (Theorem 2.9), applied to the process $(X^x, x \in \mathbb{R})$ with values in the space $E = C(\mathbb{R}_+, \mathbb{R})$ equipped with the distance $\mathbf{d}$, we get that $(X^x, x \in \mathbb{R})$ has a modification with continuous sample paths, which we denote by $(\tilde{X}^x, x \in \mathbb{R})$. We set $F_x(\mathrm{w}) = \tilde{X}^x(\mathrm{w}) = (\tilde{X}_t^x(\mathrm{w}))_{t\geq 0}$. Property (ii) is then obvious.

The mapping $\mathrm{w} \mapsto F_x(\mathrm{w})$ is measurable from $C(\mathbb{R}_+, \mathbb{R})$ equipped with the σ-field $\mathscr{G}_\infty$ into $C(\mathbb{R}_+, \mathbb{R})$ equipped with the Borel σ-field $\mathscr{C} = \sigma(\mathrm{w}(s), s \geq 0)$. Moreover, for every $t \geq 0$, $F_x(\mathrm{w})_t = \tilde{X}_t^x(\mathrm{w}) \overset{\text{a.s.}}{=} X_t^x(\mathrm{w})$ is $\mathscr{G}_t$-measurable hence coincides $W(\mathrm{dw})$ a.s. with a measurable function of $(\mathrm{w}(s), 0 \leq s \leq t)$. Thus property (i) holds.

Let us now prove the first part of assertion (iii). To this end, we fix the filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t), P)$ and the $(\mathscr{F}_t)$-Brownian motion B. We need to verify that the process $(F_x(B)_t)_{t\geq 0}$ then solves $E_x(\sigma, b)$. This process (trivially) has continuous sample paths, and is also adapted since $F_x(B)_t$ coincides a.s. with a measurable function of $(B_r, 0 \leq r \leq t)$, by (i), and since the filtration $(\mathscr{F}_t)$ is complete. On the other hand, by the construction of F_x (and because $\tilde{X}^x = X^x$ a.s.), we have, for every $t \geq 0$, $W(\mathrm{dw})$ a.s.

$$F_x(\mathrm{w})_t = x + \int_0^t \sigma(s, F_x(\mathrm{w})_s)\, \mathrm{dw}(s) + \int_0^t b(s, F_x(\mathrm{w})_s)\mathrm{d}s,$$

where the stochastic integral $\int_0^t \sigma(s, F_x(\mathrm{w})_s)\mathrm{dw}(s)$ can be defined by

$$\int_0^t \sigma(s, F_x(\mathrm{w})_s)\, \mathrm{dw}(s) = \lim_{k\to\infty} \sum_{i=0}^{2^{n_k}-1} \sigma(\frac{it}{2^{n_k}}, F_x(\mathrm{w})_{it/2^{n_k}})\, (\mathrm{w}(\frac{(i+1)t}{2^{n_k}}) - \mathrm{w}(\frac{it}{2^{n_k}})),$$

$W(\mathrm{dw})$ a.s. Here $(n_k)_{k\geq 1}$ is a suitable subsequence, and we used Proposition 5.9. We can now replace w by B (whose distribution is $W(\mathrm{dw})$!) and get a.s.

$$\begin{aligned}F_x(B)_t &= x + \lim_{k\to\infty} \sum_{i=0}^{2^{n_k}-1} \sigma(\frac{it}{2^{n_k}}, F_x(B)_{it/2^{n_k}})\,(B_{(i+1)t/2^{n_k}} - B_{it/2^{n_k}}) + \int_0^t b(s, F_x(B)_s)\mathrm{d}s \\ &= x + \int_0^t \sigma(s, F_x(B)_s)\mathrm{d}B_s + \int_0^t b(s, F_x(B)_s)\mathrm{d}s,\end{aligned}$$

again thanks to Proposition 5.9. We thus obtain that $F_x(B)$ is the desired solution.

We still have to prove the second part of assertion (iii). We again fix the filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t), P)$ and the $(\mathscr{F}_t)$-Brownian motion B. Let U be an $\mathscr{F}_0$-measurable random variable. If in the stochastic integral equation satisfied by $F_x(B)$ we formally substitute U for x, we obtain that $F_U(B)$ solves $E(\sigma, b)$ with initial value U. However, this formal substitution is not so easy to justify, and we will argue with some care.

We first observe that the mapping $(x, \omega) \mapsto F_x(B)_t$ is continuous with respect to the variable x (if ω is fixed) and $\mathscr{F}_t$-measurable with respect to ω (if x is fixed). It easily follows that this mapping is measurable for the σ-field $\mathscr{B}(\mathbb{R}) \otimes \mathscr{F}_t$. Since U is $\mathscr{F}_0$-measurable, we get that $F_U(B)_t$ is $\mathscr{F}_t$-measurable. Hence the process $F_U(B)$ is adapted. For $x \in \mathbb{R}$ and $\mathrm{w} \in C(\mathbb{R}_+, \mathbb{R})$, we define $G(x, \mathrm{w}) \in C(\mathbb{R}_+, \mathbb{R})$ by the formula

$$G(x, \mathrm{w})_t = \int_0^t b(s, F_x(\mathrm{w})_s)\,\mathrm{d}s.$$

We also set $H(x, \mathrm{w}) = F_x(\mathrm{w}) - x - G(x, \mathrm{w})$. We have already seen that, for every $x \in \mathbb{R}$, we have $W(\mathrm{dw})$ a.s.,

$$H(x, \mathrm{w})_t = \int_0^t \sigma(s, F_x(\mathrm{w})_s)\,\mathrm{dw}(s).$$

Hence, if

$$H_n(x, \mathrm{w})_t = \sum_{i=0}^{2^n-1} \sigma(\frac{it}{2^n}, F_x(\mathrm{w})_{it/2^n})\,(\mathrm{w}(\frac{(i+1)t}{2^n}) - \mathrm{w}(\frac{it}{2^n})),$$

Proposition 5.9 shows that

$$H(x, \mathrm{w})_t = \lim_{n\to\infty} H_n(x, \mathrm{w})_t,$$

in probability under $W(\mathrm{dw})$, for every $x \in \mathbb{R}$. Using the fact that U and B are independent (because U is $\mathscr{F}_0$-measurable), we infer from the latter convergence that

$$H(U,B)_t = \lim_{n\to\infty} H_n(U,B)_t$$

in probability. Thanks again to Proposition 5.9, the limit must be the stochastic integral

$$\int_0^t \sigma(s, F_U(B)_s)\, \mathrm{d}B_s.$$

We have thus proved that

$$\int_0^t \sigma(s, F_U(B)_s)\, \mathrm{d}B_s = H(U,B)_t = F_U(B)_t - U - \int_0^t b(s, F_U(B)_s)\, \mathrm{d}s\,,$$

which shows that $F_U(B)$ solves $E(\sigma, b)$ with initial value U. □

A consequence of Theorem 8.5, especially of property (ii) in this theorem, is the continuity of solutions with respect to the initial value. Given the filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t), P)$ and the $(\mathscr{F}_t)$-Brownian motion B, one can construct, for every $x \in \mathbb{R}^d$, the solution X^x of $E_x(\sigma, b)$ in such a way that, for every $\omega \in \Omega$, the mapping $x \mapsto X^x(\omega)$ is continuous. More precisely, the arguments of the previous proof give, for every $\varepsilon \in (0,1)$ and for every choice of the constants $A > 0$ and $T > 0$, a (random) constant $C_{\varepsilon,A,T}(\omega)$ such that, if $|x|, |y| \le A$,

$$\sup_{t\le T} |X_t^x(\omega) - X_t^y(\omega)| \le C_{\varepsilon,A,T}(\omega)\, |x-y|^{1-\varepsilon}$$

(in fact the version of Kolmogorov's lemma in Theorem 2.9 gives this only for $d = 1$, but there is an analogous version of Kolmogorov's lemma for processes indexed by a multidimensional parameter, see [70, Theorem I.2.1]).

8.3 Solutions of Stochastic Differential Equations as Markov Processes

In this section, we consider the homogeneous case where $\sigma(t,y) = \sigma(y)$ and $b(t,y) = b(y)$. As in the previous section, we assume that σ and b are Lipschitz: There exists a constant K such that, for every $x, y \in \mathbb{R}^d$,

$$|\sigma(x) - \sigma(y)| \le K|x-y|\,,\ |b(x) - b(y)| \le K|x-y|.$$

Let $x \in \mathbb{R}^d$, and let X^x be a solution of $E_x(\sigma, b)$. Since weak uniqueness holds, for every $t \ge 0$, the law of X_t^x does not depend on the choice of the solution. In fact, this law is the image of Wiener measure on $C(\mathbb{R}_+, \mathbb{R}^d)$ under the mapping $\mathrm{w} \mapsto F_x(\mathrm{w})_t$, where the mappings F_x were introduced in Theorem 8.5. The latter theorem also shows that the law of X_t^x depends continuously on the pair (x, t).

Theorem 8.6 *Assume that $(X_t)_{t\geq 0}$ is a solution of $E(\sigma, b)$ on a (complete) filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t), P)$. Then $(X_t)_{t\geq 0}$ is a Markov process with respect to the filtration $(\mathscr{F}_t)$, with semigroup $(Q_t)_{t\geq 0}$ defined by*

$$Q_t f(x) = E[f(X_t^x)],$$

where X^x is an arbitrary solution of $E_x(\sigma, b)$.

Remark With the notation of Theorem 8.5, we have also

$$Q_t f(x) = \int f(F_x(\mathrm{w})_t)\, W(\mathrm{dw}). \tag{8.2}$$

Proof We first verify that, for any bounded measurable function f on $\mathbb{R}^d$, and for every $s, t \geq 0$, we have

$$E[f(X_{s+t}) \mid \mathscr{F}_s] = Q_t f(X_s),$$

where $Q_t f$ is defined by (8.2). To this end, we fix $s \geq 0$ and we write, for every $t \geq 0$,

$$X_{s+t} = X_s + \int_s^{s+t} \sigma(X_r)\, \mathrm{d}B_r + \int_s^{s+t} b(X_r)\, \mathrm{d}r \tag{8.3}$$

where B is an $(\mathscr{F}_t)$-Brownian motion starting from 0. We then set, for every $t \geq 0$,

$$X'_t = X_{s+t}\,,\ \mathscr{F}'_t = \mathscr{F}_{s+t}\,,\ B'_t = B_{s+t} - B_s.$$

We observe that the filtration $(\mathscr{F}'_t)$ is complete (of course $\mathscr{F}'_\infty = \mathscr{F}_\infty$), that the process X' is adapted to $(\mathscr{F}'_t)$, and that B' is an m-dimensional $(\mathscr{F}'_t)$-Brownian motion. Furthermore, using the approximation results for the stochastic integral of adapted processes with continuous sample paths (Proposition 5.9), one easily verifies that, a.s. for every $t \geq 0$,

$$\int_s^{s+t} \sigma(X_r)\, \mathrm{d}B_r = \int_0^t \sigma(X'_u)\, \mathrm{d}B'_u$$

where the stochastic integral in the right-hand side is computed in the filtration $(\mathscr{F}'_t)$. It follows from (8.3) that

$$X'_t = X_s + \int_0^t \sigma(X'_u)\, \mathrm{d}B'_u + \int_0^t b(X'_u)\, \mathrm{d}u.$$

Hence X' solves $E(\sigma, b)$, on the space $(\Omega, \mathscr{F}, (\mathscr{F}'_t), P)$ and with the Brownian motion B', with initial value $X'_0 = X_s$ (note that X_s is $\mathscr{F}'_0$-measurable). By the last assertion of Theorem 8.5, we must have $X' = F_{X_s}(B')$, a.s.

Consequently, for every $t \geq 0$,

$$E[f(X_{s+t})|\mathscr{F}_s] = E[f(X'_t)|\mathscr{F}_s] = E[f(F_{X_s}(B')_t)|\mathscr{F}_s] = \int f(F_{X_s}(\mathrm{w})_t)\, W(\mathrm{dw})$$
$$= Q_t f(X_s),$$

by the definition of $Q_t f$. In the third equality, we used the fact that B' is independent of $\mathscr{F}_s$, and distributed according to $W(\mathrm{dw})$, whereas X_s is $\mathscr{F}_s$-measurable.

We still have to verify that $(Q_t)_{t\geq 0}$ is a transition semigroup. Properties (i) and (iii) of the definition are immediate (for (iii), we use the fact that the law of X_t^x depends continuously on the pair (x, t)). For the Chapman–Kolmogorov relation, we observe that, by applying the preceding considerations to X^x, we have, for every $s, t \geq 0$,

$$Q_{t+s} f(x) = E[f(X_{s+t}^x)] = E[E[f(X_{s+t}^x)|\mathscr{F}_s]] = E[Q_t f(X_s^x)] = \int Q_s(x, \mathrm{d}y) Q_t f(y).$$

This completes the proof. □

We write $C_c^2(\mathbb{R}^d)$ for the space of all twice continuously differentiable functions with compact support on $\mathbb{R}^d$.

Theorem 8.7 *The semigroup $(Q_t)_{t\geq 0}$ is Feller. Furthermore, its generator L is such that*

$$C_c^2(\mathbb{R}^d) \subset D(L)$$

and, for every $f \in C_c^2(\mathbb{R}^d)$,

$$Lf(x) = \frac{1}{2} \sum_{i,j=1}^{d} (\sigma\sigma^*)_{ij}(x) \frac{\partial^2 f}{\partial x_i \partial x_j}(x) + \sum_{i=1}^{d} b_i(x) \frac{\partial f}{\partial x_i}(x)$$

where σ^* denotes the transpose of the matrix σ.

Proof For the sake of simplicity, we give the proof only in the case when σ and b are bounded. We fix $f \in C_0(\mathbb{R}^d)$ and we first verify that $Q_t f \in C_0(\mathbb{R}^d)$. Since the mappings $x \mapsto F_x(\mathrm{w})$ are continuous, formula (8.2) and dominated convergence show that $Q_t f$ is continuous. Then, since

$$X_t^x = x + \int_0^t \sigma(X_s^x)\, \mathrm{d}B_s + \int_0^t b(X_s^x)\, \mathrm{d}s,$$

and σ and b are assumed to be bounded, we get the existence of a constant C, which does not depend on t, x, such that

$$E[(X_t^x - x)^2] \le C(t + t^2). \tag{8.4}$$

Using Markov's inequality, we have thus, for every $t \ge 0$,

$$\sup_{x \in \mathbb{R}^d} P(|X_t^x - x| > A) \underset{A \to \infty}{\longrightarrow} 0.$$

Writing

$$|Q_t f(x)| = |E[f(X_t^x)]| \le |E[f(X_t^x)\, \mathbf{1}_{\{|X_t^x - x| \le A\}}]| + \|f\|\, P(|X_t^x - x| > A),$$

we get, using our assumption $f \in C_0(\mathbb{R}^d)$,

$$\limsup_{x \to \infty} |Q_t f(x)| \le \|f\| \sup_{x \in \mathbb{R}^d} P(|X_t^x - x| > A),$$

and thus, since A was arbitrary,

$$\lim_{x \to \infty} Q_t f(x) = 0,$$

which completes the proof of the property $Q_t f \in C_0(\mathbb{R}^d)$.

Let us show similarly that $Q_t f \longrightarrow f$ when $t \to 0$. For every $\varepsilon > 0$,

$$\sup_{x \in \mathbb{R}^d} |E[f(X_t^x)] - f(x)| \le \sup_{x, y \in \mathbb{R}^d, |x - y| \le \varepsilon} |f(x) - f(y)| + 2\|f\| \sup_{x \in \mathbb{R}^d} P(|X_t^x - x| > \varepsilon).$$

However, using (8.4) and Markov's inequality again, we get

$$\sup_{x \in \mathbb{R}^d} P(|X_t^x - x| > \varepsilon) \underset{t \to 0}{\longrightarrow} 0,$$

hence

$$\limsup_{t \to 0} \|Q_t f - f\| = \limsup_{t \to 0} \Big(\sup_{x \in \mathbb{R}^d} |E[f(X_t^x)] - f(x)| \Big) \le \sup_{x, y \in \mathbb{R}^d, |x - y| \le \varepsilon} |f(x) - f(y)]$$

which can be made arbitrarily close to 0 by taking ε small.

Let us prove the second assertion of the theorem. Let $f \in C_c^2(\mathbb{R}^d)$. We apply Itô's formula to $f(X_t^x)$, recalling that, if $X_t^x = (X_t^{x,1}, \ldots, X_t^{x,d})$, we have, for every $i \in \{1, \ldots, d\}$,

$$X_t^{x,i} = x_i + \sum_{j=1}^{m} \int_0^t \sigma_{ij}(X_s^x)\, \mathrm{d}B_s^j + \int_0^t b_i(X_s^x)\, \mathrm{d}s.$$

We get

$$f(X_t^x) = f(x) + M_t + \sum_{i=1}^{d} \int_0^t b_i(X_s^x) \frac{\partial f}{\partial x_i}(X_s^x) \mathrm{d}s + \frac{1}{2} \sum_{i,i'=1}^{d} \int_0^t \frac{\partial^2 f}{\partial x_i \partial x_{i'}}(X_s^x) \mathrm{d}\langle X^{x,i}, X^{x,i'} \rangle_s$$

where M is a continuous local martingale. Moreover, if $i, i' \in \{1, \ldots, d\}$,

$$\mathrm{d}\langle X^{x,i}, X^{x,i'} \rangle_s = \sum_{j=1}^{m} \sigma_{ij}(X_s^x) \sigma_{i'j}(X_s^x) \, \mathrm{d}s = (\sigma\sigma^*)_{ii'}(X_s^x) \, \mathrm{d}s.$$

We thus see that, if g is the function defined by

$$g(x) = \frac{1}{2} \sum_{i,i'=1}^{d} (\sigma\sigma^*)_{ii'}(x) \, \frac{\partial^2 f}{\partial x_i \partial x_{i'}}(x) + \sum_{i=1}^{d} b_i(x) \, \frac{\partial f}{\partial x_i}(x),$$

the process

$$M_t = f(X_t^x) - f(x) - \int_0^t g(X_s^x) \, \mathrm{d}s$$

is a continuous local martingale. Since f and g are bounded, Proposition 4.7 (ii) shows that M is a martingale. It now follows from Theorem 6.14 that $f \in D(L)$ and $Lf = g$. □

Corollary 8.8 *Suppose that $(X_t)_{t\geq 0}$ solves $E(\sigma, b)$ on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t), P)$. Then $(X_t)_{t\geq 0}$ satisfies the strong Markov property: If T is a stopping time and if Φ is a Borel measurable function from $C(\mathbb{R}_+, \mathbb{R}^d)$ into $\mathbb{R}_+$,*

$$E[\mathbf{1}_{\{T<\infty\}} \Phi(X_{T+t}, t \geq 0) \mid \mathscr{F}_T] = \mathbf{1}_{\{T<\infty\}} \, \mathbb{E}_{X_T}[\Phi],$$

where, for every $x \in \mathbb{R}^d$, $\mathbb{P}_x$ denotes the law on $C(\mathbb{R}_+, \mathbb{R}^d)$ of an arbitrary solution of $E_x(\sigma, b)$.

Proof It suffices to apply Theorem 6.17. Alternatively, we could also argue in a similar manner as in the proof of Theorem 8.6, letting the stopping time T play the same role as the deterministic time s in the latter proof, and using the strong Markov property of Brownian motion. □

Markov processes with continuous sample paths that are obtained as solutions of stochastic differential equations are sometimes called *diffusion processes* (certain authors call a diffusion process any strong Markov process with continuous sample paths in $\mathbb{R}^d$ or on a manifold). Note that, even in the Lipschitz setting considered here, Theorem 8.7 does not completely identify the generator L, but only its action on a subset of the domain $D(L)$: As we already mentioned in Chap. 6, it is often very difficult to give a complete description of the domain. However, in many instances, one can show that a partial knowledge of the generator, such as the one given by Theorem 8.7, suffices to characterize the law of the process. This observation is at the core of the powerful theory of martingale problems, which is developed in the classical book [77] by Stroock and Varadhan.

At least when restricted to $C_c^2(\mathbb{R}^d)$, the generator L is a second order differential operator. The stochastic differential equation $E(\sigma, b)$ allows one to give a probabilistic approach (as well as an interpretation) to many analytic results concerning this differential operator, in the spirit of the connections between Brownian motion and the Laplace operator described in the previous chapter. We refer to Durrett [18, Chapter 9] and Friedman [26, 27] for more about links between stochastic differential equations and partial differential equations. These connections between probability and analysis were an important motivation for the definition and study of stochastic differential equations.

8.4 A Few Examples of Stochastic Differential Equations

In this section, we briefly discuss three important examples, all in dimension one. In the first two examples, one can obtain an explicit formula for the solution, which is of course not the case in general.

8.4.1 The Ornstein–Uhlenbeck Process

Let $\lambda > 0$. The (one-dimensional) Ornstein–Uhlenbeck process is the solution of the stochastic differential equation

$$\mathrm{d}X_t = \mathrm{d}B_t - \lambda X_t \,\mathrm{d}t.$$

This equation is solved by applying Itô's formula to $\mathrm{e}^{\lambda t} X_t$, and we get

$$X_t = X_0 \mathrm{e}^{-\lambda t} + \int_0^t \mathrm{e}^{-\lambda(t-s)} \,\mathrm{d}B_s.$$

Note that the stochastic integral is a Wiener integral (the integrand is deterministic), which thus belongs to the Gaussian space generated by B.

First consider the case where $X_0 = x \in \mathbb{R}$. By the previous remark, X is a (non-centered) Gaussian process, whose mean function is $m(t) = E[X_t] = x\,\mathrm{e}^{-\lambda t}$, and whose covariance function is also easy to compute:

$$K(s,t) = \operatorname{cov}(X_s, X_t) = \frac{\mathrm{e}^{-\lambda|t-s|} - \mathrm{e}^{-\lambda(t+s)}}{2\lambda}.$$

It is also interesting to consider the case when X_0 is distributed according to $\mathcal{N}(0, \frac{1}{2\lambda})$. In that case, X is a centered Gaussian process with covariance function

$$\frac{1}{2\lambda}\,\mathrm{e}^{-\lambda|t-s|}.$$

Notice that this is a stationary covariance function. In that case, the Ornstein–Uhlenbeck process X is both a stationary Gaussian process (indexed by $\mathbb{R}_+$) and a Markov process.

8.4.2 Geometric Brownian Motion

Let $\sigma > 0$ and $r \in \mathbb{R}$. The geometric Brownian motion with parameters σ and r is the solution of the stochastic differential equation

$$\mathrm{d}X_t = \sigma X_t\,\mathrm{d}B_t + rX_t\,\mathrm{d}t.$$

One solves this equation by applying Itô's formula to $\log X_t$ (say in the case where $X_0 > 0$), and it follows that:

$$X_t = X_0\,\exp\Big(\sigma B_t + (r - \frac{\sigma^2}{2})t\Big).$$

Note in particular that, if the initial value X_0 is (strictly) positive, the solution remains so at every time $t \geq 0$. Geometric Brownian motion is used in the celebrated Black–Scholes model of financial mathematics. The reason for the use of this process comes from an economic assumption of independence of the successive ratios

$$\frac{X_{t_2} - X_{t_1}}{X_{t_1}}, \frac{X_{t_3} - X_{t_2}}{X_{t_2}}, \ldots, \frac{X_{t_n} - X_{t_{n-1}}}{X_{t_{n-1}}}$$

corresponding to disjoint time intervals: From the explicit formula for X_t, we see that this is nothing but the property of independence of increments of Brownian motion.

8.4.3 Bessel Processes

Let $m \geq 0$ be a real number. The m-dimensional squared Bessel process is the real process taking nonnegative values that solves the stochastic differential equation

$$\mathrm{d}X_t = 2\sqrt{X_t}\,\mathrm{d}B_t + m\,\mathrm{d}t\,. \tag{8.5}$$

Notice that this equation does not fit into the Lipschitz setting studied in this chapter, because the function $\sigma(x) = 2\sqrt{x}$ is not Lipschitz over $\mathbb{R}_+$ (one might also observe that this function is only defined on $\mathbb{R}_+$ and not on $\mathbb{R}$, but this is a minor point because one can replace $2\sqrt{x}$ by $2\sqrt{|x|}$ and check a posteriori that a solution starting from a nonnegative value stays nonnegative). However, there exist (especially in dimension one) criteria weaker than our Lipschitz continuity assumptions, which apply to (8.5) and give the existence and pathwise uniqueness of solutions of (8.5). See in particular Exercise 8.14 for a criterion of pathwise uniqueness that applies to (8.5).

One of the main reasons for studying Bessel processes comes from the following observation. If $d \geq 1$ is an integer and $\beta = (\beta^1, \ldots, \beta^d)$ is a d-dimensional Brownian motion, an application of Itô's formula shows that the process

$$|\beta_t|^2 = (\beta_t^1)^2 + \cdots + (\beta_t^d)^2$$

is a d-dimensional squared Bessel process: See Exercise 5.33. Furthermore, one can also check that, when $m = 0$, the process $(\frac{1}{2}X_t)_{t\geq 0}$ has the same distribution as Feller's branching diffusion discussed at the end of Chap. 6 (see Exercise 8.11).

Suppose from now on that $m > 0$ and $X_0 = x > 0$. For every $r \geq 0$, set $T_r := \inf\{t \geq 0 : X_t = r\}$. If $r > x$, we have $P(T_r < \infty) = 1$. To get this, use (8.5) to see that $X_{t\wedge T_r} = x + m(t \wedge T_r) + Y_{t\wedge T_r}$, where $E[(Y_{t\wedge T_r})^2] \leq 4\,r\,t$. By Markov's inequality, $P(Y_{t\wedge T_r} > t^{3/4}) \longrightarrow 0$ as $t \to \infty$, and if we assume that $P(T_r = \infty) > 0$ the preceding expression for $X_{t\wedge T_r}$ gives a contradiction.

Set, for every $t \in [0, T_0)$,

$$M_t = \begin{cases} (X_t)^{1-\frac{m}{2}} & \text{if } m \neq 2, \\ \log(X_t) & \text{if } m = 2. \end{cases}$$

It follows from Itô's formula that, for every $\varepsilon \in (0, x)$, $M_{t\wedge T_\varepsilon}$ is a continuous local martingale. This continuous local martingale is bounded over the time interval $[0, T_\varepsilon \wedge T_A]$, for every $A > x$, and an application of the optional stopping theorem (using the fact that $T_A < \infty$ a.s.) gives, if $m \neq 2$,

$$P(T_\varepsilon < T_A) = \frac{A^{1-\frac{m}{2}} - x^{1-\frac{m}{2}}}{A^{1-\frac{m}{2}} - \varepsilon^{1-\frac{m}{2}}},$$

and if $m = 2$,

$$P(T_\varepsilon < T_A) = \frac{\log A - \log x}{\log A - \log \varepsilon}.$$

When $m = d$ is an integer, we recover the formulas of Proposition 7.16.

Let us finally concentrate on the case $m \geq 2$. Letting ε go to 0 in the preceding formulas, we obtain that $P(T_0 < \infty) = 0$. If we let A tend to ∞, we also get that $P(T_\varepsilon < \infty) = 1$ if $m = 2$ (as we already noticed in Chap. 7) and $P(T_\varepsilon < \infty) = (\varepsilon/x)^{(m/2)-1}$ if $m > 2$.

It then follows from the property $P(T_0 < \infty) = 0$ that the process M_t is well-defined for every $t \geq 0$ and is a continuous local martingale. When $m > 2$, M_t takes nonnegative values and is thus a supermartingale (Proposition 4.7 (i)), which converges a.s. as $t \to \infty$ (Proposition 3.19). The limit must be 0, since we already noticed that $P(T_A < \infty) = 1$ for every $A > x$, and we conclude that X_t converges a.s. to ∞ as $t \to \infty$ when $m > 2$. One can show that the continuous local martingale M_t is not a (true) martingale (cf. Question **8.** in Exercise 5.33 in the case $m = 3$).

Exercise 5.31 in Chap. 5 gives a number of important calculations related to squared Bessel processes. We refer to Chapter XI in [70] for a thorough study of this class of processes.

Remark The m-dimensional Bessel process is (of course) obtained by taking $Y_t = \sqrt{X_t}$, and, when $m = d$ is a positive integer, it corresponds to the norm of d-dimensional Brownian motion. When $m > 1$, the process Y also satisfies a stochastic differential equation, which is however less tractable than (8.5): See Exercise 8.13 below.

Exercises

Exercise 8.9 (*Time change method*) We consider the stochastic differential equation

$E(\sigma, 0)$ $$\mathrm{d}X_t = \sigma(X_t)\,\mathrm{d}B_t$$

where the function $\sigma : \mathbb{R} \longrightarrow \mathbb{R}$ is continuous and there exist constants $\varepsilon > 0$ and M such that $\varepsilon \leq \sigma \leq M$.

1. In this question and the next one, we assume that X solves $E(\sigma, 0)$ with $X_0 = x$. We set, for every $t \geq 0$,

$$A_t = \int_0^t \sigma(X_s)^2\,\mathrm{d}s \quad , \quad \tau_t = \inf\{s \geq 0 : A_s > t\}.$$

Justify the equalities

$$\tau_t = \int_0^t \frac{\mathrm{d}r}{\sigma(X_{\tau_r})^2}\,, \quad A_t = \inf\{s \geq 0 : \int_0^s \frac{\mathrm{d}r}{\sigma(X_{\tau_r})^2} > t\}.$$

2. Show that there exists a real Brownian motion $\beta = (\beta_t)_{t\geq 0}$ started from x such that, a.s. for every $t \geq 0$,

$$X_t = \beta_{\inf\{s\geq 0: \int_0^s \sigma(\beta_r)^{-2}\mathrm{d}r > t\}}.$$

3. Show that weak existence and weak uniqueness hold for $E(\sigma, 0)$. (*Hint*: For the existence part, observe that, if X is defined from a Brownian motion β by the formula of question **2.**, X is (in an appropriate filtration) a continuous local martingale with quadratic variation $\langle X, X\rangle_t = \int_0^t \sigma(X_s)^2 \mathrm{d}s$.)

Exercise 8.10 We consider the stochastic differential equation

$E(\sigma, b)$ $\qquad \mathrm{d}X_t = \sigma(X_t)\,\mathrm{d}B_t + b(X_t)\,\mathrm{d}t$

where the functions $\sigma, b : \mathbb{R} \longrightarrow \mathbb{R}$ are bounded and continuous, and such that $\int_{\mathbb{R}} |b(x)|dx < \infty$ and $\sigma \geq \varepsilon$ for some constant $\varepsilon > 0$.

1. Let X be a solution of $E(\sigma, b)$. Show that there exists a monotone increasing function $F : \mathbb{R} \longrightarrow \mathbb{R}$, which is also twice continuously differentiable, such that $F(X_t)$ is a martingale. Give an explicit formula for F in terms of σ and b.

2. Show that the process $Y_t = F(X_t)$ solves a stochastic differential equation of the form $dY_t = \sigma'(Y_t)\,\mathrm{d}B_t$, with a function σ' to be determined.

3. Using the result of the preceding exercise, show that weak existence and weak uniqueness hold for $E(\sigma, b)$. Show that pathwise uniqueness also holds if σ is Lipschitz.

Exercise 8.11 We suppose that, for every $x \in \mathbb{R}_+$, one can construct on the same filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t), P)$ a process X^x taking nonnegative values, which solves the stochastic differential equation

$$\begin{cases} \mathrm{d}X_t = \sqrt{2X_t}\,dB_t \\ X_0 = x \end{cases}$$

and that the processes X^x are Markov processes with values in $\mathbb{R}_+$, with the same semigroup $(Q_t)_{t\geq 0}$, with respect to the filtration $(\mathscr{F}_t)$. (This is, of course, close to Theorem 8.6, which however cannot be applied directly because the function $\sqrt{2x}$ is not Lipschitz.)

1. We fix $x \in \mathbb{R}_+$, and a real $T > 0$. We set, for every $t \in [0, T]$

$$M_t = \exp\Big(-\frac{\lambda X_t^x}{1 + \lambda(T - t)}\Big).$$

Show that the process $(M_{t\wedge T})_{t\geq 0}$ is a martingale.

2. Show that $(Q_t)_{t\geq 0}$ is the semigroup of Feller's branching diffusion (see the end of Chap. 6).

Exercise 8.12 We consider two sequences $(\sigma_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$ of real functions defined on $\mathbb{R}$. We assume that:

(i) There exists a constant $C > 0$ such that $|\sigma_n(x)| \leq C$ and $|b_n(x)| \leq C$ for every $n \geq 1$ and $x \in \mathbb{R}$.
(ii) There exists a constant $K > 0$ such that, for every $n \geq 1$ and $x, y \in \mathbb{R}$,

$$|\sigma_n(x) - \sigma_n(y)| \leq K|x-y| \quad , \quad |b_n(x) - b_n(y)| \leq K|x-y|.$$

Let B be an $(\mathscr{F}_t)$-Brownian motion and, for every $n \geq 1$, let X^n be the unique adapted process satisfying

$$X_t^n = \int_0^t \sigma_n(X_s^n)\,\mathrm{d}B_s + \int_0^t b_n(X_s^n)\,\mathrm{d}s.$$

1. Let $T > 0$. Show that there exists a constant $A > 0$ such that, for every real $M > 0$ and for every $n \geq 1$,

$$P\Big(\sup_{t\leq T} |X_t^n| \geq M\Big) \leq \frac{A}{M^2}.$$

2. We assume that the sequences (σ_n) and (b_n) converge uniformly on every compact subset of $\mathbb{R}$ to limiting functions denoted by σ and b respectively. Justify the existence of an adapted process $X = (X_t)_{t\geq 0}$ with continuous sample paths, such that

$$X_t = \int_0^t \sigma(X_s)\,\mathrm{d}B_s + \int_0^t b(X_s)\,\mathrm{d}s,$$

then show that there exists a constant A' such that, for every real $M > 0$, for every $t \in [0, T]$ and $n \geq 1$,

$$E\Big[\sup_{s\leq t}(X_s^n - X_s)^2\Big] \leq 4(4+T)K^2 \int_0^t E[(X_s^n - X_s)^2]\,\mathrm{d}s + \frac{A'}{M^2}$$
$$+4T\Big(4 \sup_{|x|\leq M} (\sigma_n(x) - \sigma(x))^2 + T \sup_{|x|\leq M} (b_n(x) - b(x))^2\Big).$$

3. Infer from the preceding question that

$$\lim_{n\to\infty} E\Big[\sup_{s\leq T}(X_s^n - X_s)^2\Big] = 0.$$

Exercise 8.13 Let $\beta = (\beta_t)_{t\geq 0}$ be an $(\mathscr{F}_t)$-Brownian motion started from 0. We fix two real parameters α and r, with $\alpha > 1/2$ and $r > 0$. For every integer $n \geq 1$ and every $x \in \mathbb{R}$, we set

$$f_n(x) = \frac{1}{|x|} \wedge n \,.$$

1. Let $n \geq 1$. Justify the existence of the unique semimartingale Z^n that solves the equation

$$Z_t^n = r + \beta_t + \alpha \int_0^t f_n(Z_s^n)\,\mathrm{d}s.$$

2. We set $S_n = \inf\{t \geq 0 : Z_t^n \leq 1/n\}$. After observing that, for $t \leq S_n \wedge S_{n+1}$,

$$Z_t^{n+1} - Z_t^n = \alpha \int_0^t \Big(\frac{1}{Z_s^{n+1}} - \frac{1}{Z_s^n}\Big)\mathrm{d}s\,,$$

show that $Z_t^{n+1} = Z_t^n$ for every $t \in [0, S_n \wedge S_{n+1}]$, a.s. Infer that $S_{n+1} \geq S_n$.

3. Let g be a twice continuously differentiable function on $\mathbb{R}$. Show that the process

$$g(Z_t^n) - g(r) - \int_0^t \Big(\alpha g'(Z_s^n) f_n(Z_s^n) + \frac{1}{2} g''(Z_s^n)\Big)\mathrm{d}s$$

is a continuous local martingale.

4. We set $h(x) = x^{1-2\alpha}$ for every $x > 0$. Show that, for every integer $n \geq 1$, $h(Z_{t\wedge S_n}^n)$ is a bounded martingale. Infer that, for every $t \geq 0$, $P(S_n \leq t)$ tends to 0 as $n \to \infty$, and consequently $S_n \to \infty$ a.s. as $n \to \infty$.

5. Infer from questions **2.** and **4.** that there exists a unique positive semimartingale Z such that, for every $t \geq 0$,

$$Z_t = r + \beta_t + \alpha \int_0^t \frac{\mathrm{d}s}{Z_s}.$$

6. Let $d \geq 3$ and let B be a d-dimensional Brownian motion started from $y \in \mathbb{R}^d \backslash \{0\}$. Show that $Y_t = |B_t|$ satisfies the stochastic equation in question **5.** (with an appropriate choice of β) with $r = |y|$ and $\alpha = (d-1)/2$. One may use the results of Exercise 5.33.

Exercise 8.14 (*Yamada–Watanabe uniqueness criterion*) The goal of the exercise is to get pathwise uniqueness for the one-dimensional stochastic differential equation

$$\mathrm{d}X_t = \sigma(X_t)\mathrm{d}B_t + b(X_t)\mathrm{d}t$$

when the functions σ and b satisfy the conditions

$$|\sigma(x) - \sigma(y)| \leq K\sqrt{|x-y|}\,, \quad |b(x) - b(y)| \leq K\,|x-y|\,,$$

for every $x, y \in \mathbb{R}$, with a constant $K < \infty$.

1. Preliminary question. Let Z be a semimartingale such that $\langle Z, Z\rangle_t = \int_0^t h_s\,\mathrm{d}s$, where $0 \leq h_s \leq C\,|Z_s|$, with a constant $C < \infty$. Show that, for every $t \geq 0$,

$$\lim_{n\to\infty} n\,E\Big[\int_0^t \mathbf{1}_{\{0<|Z_s|\leq 1/n\}}\,\mathrm{d}\langle Z, Z\rangle_s\Big] = 0.$$

(*Hint*: Observe that, for every $n \geq 1$,

$$E\Big[\int_0^t |Z_s|^{-1}\mathbf{1}_{\{0<|Z_s|\leq 1\}}\,\mathrm{d}\langle Z, Z\rangle_s\Big] \leq C\,t < \infty.\,)$$

2. For every integer $n \geq 1$, let φ_n be the function defined on $\mathbb{R}$ by

$$\varphi_n(x) = \begin{cases} 0 & \text{if } |x| \geq 1/n, \\ 2n(1-nx) & \text{if } 0 \leq x \leq 1/n, \\ 2n(1+nx) & \text{if } -1/n \leq x \leq 0. \end{cases}$$

Also write F_n for the unique twice continuously differentiable function on $\mathbb{R}$ such that $F_n(0) = F_n'(0) = 0$ and $F_n'' = \varphi_n$. Note that, for every $x \in \mathbb{R}$, one has $F_n(x) \longrightarrow |x|$ and $F_n'(x) \longrightarrow \mathrm{sgn}(x) := \mathbf{1}_{\{x>0\}} - \mathbf{1}_{\{x<0\}}$ when $n \to \infty$.

Let X and X' be two solutions of $E(\sigma, b)$ on the same filtered probability space and with the same Brownian motion B. Infer from question **1.** that

$$\lim_{n\to\infty} E\Big[\int_0^t \varphi_n(X_s - X_s')\,\mathrm{d}\langle X - X', X - X'\rangle_s\Big] = 0.$$

3. Let T be a stopping time such that the semimartingale $X_{t\wedge T} - X'_{t\wedge T}$ is bounded. By applying Itô's formula to $F_n(X_{t\wedge T} - X'_{t\wedge T})$, show that

$$E[|X_{t\wedge T} - X'_{t\wedge T}|] = E[|X_0 - X_0'|] + E\Big[\int_0^{t\wedge T} (b(X_s) - b(X_s'))\,\mathrm{sgn}(X_s - X_s')\,\mathrm{d}s\Big].$$

4. Using Gronwall's lemma, show that, if $X_0 = X_0'$, one has $X_t = X_t'$ for every $t \geq 0$, a.s.

Notes and Comments

As already mentioned, the treatment of stochastic differential equations motivated Itô's invention of stochastic differential equations. For further reading on this topic, the reader may consult the classical books of Ikeda and Watanabe [43] and Stroock and Varadhan [77], the latter studying stochastic differential equations in connection with martingale problems. Øksendal's book [66] emphasizes the applications of stochastic differential equations in other fields. The books [26, 27] of Friedman focus on connections with partial differential equations. We have chosen to concentrate on the Lipschitz case, where the main results of existence and uniqueness were already obtained by Itô [37, 38]. In dimension one, the criteria ensuring pathwise uniqueness can be weakened significantly (see in particular Yamada and Watanabe [83], which inspired Exercise 8.14) but this is no longer the case in higher dimensions. Chapter XI of [70] contains a lot of information about Bessel processes.