Stochastic Differential Equations

Steven P. Lalley

December 2, 2016

1 SDEs: Definitions

1.1 Stochastic differential equations

Many important continuous-time Markov processes — for instance, the Ornstein-Uhlenbeck process and the Bessel processes — can be defined as solutions to *stochastic differential equations* with drift and diffusion coefficients that depend only on the current value of the process. The general form of such an equation (for a one-dimensional process with a one-dimensional driving Brownian motion) is

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t, \tag{1}$$

where $\{W_t\}_{t\geq 0}$ is a standard Wiener process.

Definition 1. Let $\{W_t\}_{t\geq 0}$ be a standard Brownian motion on a probability space $(\Omega, \mathcal{F}, P)$ with an admissible filtration $\mathbb{F} = \{\mathcal{F}_t\}_{t\geq 0}$. A *strong solution* of the stochastic differential equation (1) with initial condition $x \in \mathbb{R}$ is an adapted process $X_t = X_t^x$ with continuous paths such that for all $t \geq 0$,

$$X_t = x + \int_0^t \mu(X_s)\,ds + \int_0^t \sigma(X_s)\,dW_s \quad \text{a.s.} \tag{2}$$

At first sight this definition seems to have little content except to give a more-or-less obvious interpretation of the differential equation (1). However, there are a number of subtle points involved: First, the existence of the integrals in (2) requires some degree of regularity on X_t and the functions μ and σ; in particular, it must be the case that for all $t \geq 0$, with probability one,

$$\int_0^t |\mu(X_s)|\,ds < \infty \quad \text{and} \quad \int_0^t \sigma^2(X_s)\,ds < \infty. \tag{3}$$

Second, the solution is required to exist for all $t < \infty$ with probability one. In fact, there are interesting cases of (1) for which solutions can be constructed up to a finite, possibly random time $T < \infty$, but not beyond; this often happens because the solution X_t *explodes* (that is, runs off to $\pm\infty$) in finite time. Third, the definition requires that the process X_t live on the same probability space as the given Wiener process W_t, and that it be adapted to the given filtration. It turns out (as we will see) that for certain coefficient functions μ and σ, solutions to the stochastic integral equation equation (2) may exist for *some* Wiener processes and *some* admissible filtrations but not for others.

Definition 2. A *weak solution* of the stochastic differential equation (1) with initial condition x is a continuous stochastic process X_t defined on *some* probability space $(\Omega, \mathcal{F}, P)$ such that for some Wiener process W_t and some admissible filtration $\mathbb{F}$ the process $X(t)$ is adapted and satisfies the stochastic integral equation (2).

2 Existence and Uniqueness of Solutions

2.1 Itô's existence/uniqueness theorem

The basic result, due to Itô, is that for *uniformly Lipschitz* functions $\mu(x)$ and $\sigma(x)$ the stochastic differential equation (1) has strong solutions, and that for each initial value $X_0 = x$ the solution is unique.

Theorem 1. *Assume that $\mu : \mathbb{R} \to \mathbb{R}$ and $\sigma : \mathbb{R} \to \mathbb{R}_+$ are uniformly Lipschitz, that is, there exists a constant $C < \infty$ such that for all $x, y \in \mathbb{R}$,*

$$|\mu(x) - \mu(y)| \leq C|x - y| \quad \text{and} \tag{4}$$

$$|\sigma(x) - \sigma(y)| \leq C|x - y|. \tag{5}$$

Then the stochastic differential equation (1) *has strong solutions: In particular, for any standard Brownian motion $\{W_t\}_{t \geq 0}$, any admissible filtration $\mathbb{F} = \{\mathcal{F}_t\}_{t \geq 0}$, and any initial value $x \in \mathbb{R}$ there exists a unique adapted process $X_t = X_t^x$ with continuous paths such that*

$$X_t = x + \int_0^t \mu(X_s)\,ds + \int_0^t \sigma(X_s)\,dW_s \quad \text{a.s.} \tag{6}$$

Furthermore, the solutions depend continuously on the initial data x, that is, the two-parameter process X_t^x is jointly continuous in t and x.

This parallels the main existence/uniqueness result for *ordinary* differential equations, or more generally finite systems of ordinary differential equations

$$x'(t) = F(x(t)), \tag{7}$$

which asserts that unique solutions exist for each initial value $x(0)$ provided the function F is uniformly Lipschitz. Without the hypothesis that the function F is Lipschitz, the theorem may fail in any number of ways, even for ordinary differential equations.

Example 1. Consider the equation $x' = 2\sqrt{|x|}$. This is the special case of equation (7) with $F(x) = \sqrt{x}$. This function fails the Lipschitz property at $x = 0$. Correspondingly, uniqueness of solutions fails for the initial value $x(0) = 0$: the functions

$$x(t) \equiv 0 \quad \text{and} \quad y(t) = t^2$$

are both solutions of the ordinary differential equation with initial value 0.

Example 2. Consider the equation $x' = x^2$, the special case of (7) where $F(x) = x^2$. The function F is C^∞, hence Lipschitz on any finite interval, but it is not uniformly Lipschitz, as uniformly Lipschitz functions cannot grow faster than linearly. For any initial value $x_0 > 0$, the function

$$x(t) = (x_0^{-1} - t)^{-1}$$

solves the differential equation and has the right initial value, and it can be shown that there is no other solution. The difficulty is that the function $x(t)$ blows up as $t \to 1/x_0$, so the solution does not exist for all time $t > 0$. The same difficulty can arise with stochastic differential equations whose coefficients grow too quickly: for stochastic differential equations, when solutions travel to $\pm\infty$ in finite time they are said to *explode*.

2.2 Gronwall inequalities

The proof of Theorem 1 will make use of several basic results concerning the solutions of simple differential inequalities due to Gronwall. These are also useful in the theory of ordinary differential equations.

Lemma 1. *Let $y(t)$ be a nonnegative function that satisfies the following condition: For some $T \leq \infty$ there exist constants $A, B \geq 0$ such that*

$$y(t) \leq A + B \int_0^t y(s)\, ds < \infty \quad \textit{for all} \quad 0 \leq t \leq T. \tag{8}$$

Then

$$y(t) \leq Ae^{Bt} \quad \textit{for all} \quad 0 \leq t \leq T. \tag{9}$$

Proof. Without loss of generality, we may assume that $C := \int_0^T y(s)\, ds < \infty$ and that $T < \infty$. It then follows since y is nonnegative, that $y(t)$ is bounded by $D := A + BC$ on the interval $[0, T]$. Iterate the inequality (8) to obtain

$$\begin{aligned}
y(t) &\leq A + B \int_0^t y(s)\, ds \\
&\leq A + B \int_0^t (A + B) \int_0^s y(r)\, drds \\
&\leq A + BAt + B^2 \int_0^t \int_0^s (A + B \int_0^r y(q)\, dq) drds \\
&\leq A + BAt + B^2 At^2/2! + B^3 \int_0^t \int_0^s \int_0^r (A + B \int_0^q y(p)\, dp) dqdrds \\
&\leq \cdots .
\end{aligned}$$

After k iterations, one has the first k terms in the series for Ae^{Bt} plus a $(k+1)-$fold iterated integral I_k. Because $y(t) \leq D$ on the interval $[0, T]$, the integral I_k is bounded by $B^k Dt^{k+1}/(k+1)!$. This converges to zero uniformly for $t \leq T$ as $k \to \infty$. Hence, inequality (9) follows. □

Lemma 2. *Let $y_n(t)$ be a sequence of nonnegative functions such that for some constants $B, C < \infty$,*

$$\begin{aligned}
y_0(t) &\leq C \quad \textit{for all} \quad t \leq T \quad \textit{and} \\
y_{n+1}(t) &\leq B \int_0^t y_n(s)\, ds < \infty \quad \textit{for all} \quad t \leq T \quad \textit{and} \quad n = 0, 1, 2, \dots .
\end{aligned} \tag{10}$$

Then

$$y_n(t) \leq CB^n t^n/n! \quad \textit{for all} \quad t \leq T. \tag{11}$$

Proof. Exercise. □

2.3 Proof of Theorem 1: Constant σ

It is instructive to first consider the special case where the function $\sigma(x) \equiv \sigma$ is constant. (This includes the possibility $\sigma \equiv 0$, which the stochastic differential equation reduces to an *ordinary* differential equation $x' = \mu(x)$.) In this case the Gronwall inequalities can be used *pathwise* to prove all three assertions of the theorem (existence, uniqueness, and continuous dependence on initial conditions). First, uniqueness: suppose that for some initial value x there are two continuous solutions

$$X_t = x + \int_0^t \mu(X_s)\,ds + \int_0^t \sigma\,dW_s \quad \text{and}$$
$$Y_t = x + \int_0^t \mu(Y_s)\,ds + \int_0^t \sigma\,dW_s.$$

Then the difference satisfies

$$Y_t - X_t = \int_0^t (\mu(Y_s) - \mu(X_s))\,ds,$$

and since the drift coefficient μ is uniformly Lipschitz, it follows that for some constant $B < \infty$,

$$|Y_t - X_t| \le B \int_0^t |Y_s - X_s|\,ds$$

for all $t < \infty$. Lemma 1 now implies that $Y_t - X_t \equiv 0$. Thus, the stochastic differential equation can have at most one solution for any particular initial value x. A similar argument shows that solutions depend continuously on initial conditions $X_0 = x$.

Existence of solutions is proved by a variant of Picard's method of successive approximations. Fix an initial value x, and define a sequence of adapted process $X_n(t)$ by

$$X_0(t) = x \quad \text{and} \quad X_{n+1}(t) = x + \int_0^t \mu(X_n(s))\,ds + \sigma W(t).$$

The processes $X_n(t)$ are all well-defined and have continuous paths, by induction on n (using the hypothesis that the function $\mu(y)$ is continuous). The strategy will be to show that the sequence $X_n(t)$ converges uniformly on compact time intervals. It will then follow, by the dominated convergence theorem and the continuity of μ, that the limit process $X(t)$ solves the stochastic integral equation (6). Because $\mu(y)$ is Lipschitz,

$$|X_{n+1}(t) - X_n(t)| \le B \int_0^t |X_n(s) - X_{n-1}(s)|\,ds,$$

and so Lemma 2 implies that for any $T < \infty$,

$$|X_{n+1}(t) - X_n(t)| \le CB^n T^n / n! \quad \text{for all } t \le T$$

It follows that the processes $X_n(t)$ converge uniformly on compact time intervals $[0, T]$, and therefore that the limit process $X(t)$ has continuous trajectories. $\square$

2.4 Proof of Theorem 1. General Case: Existence

The proof of Theorem 1 in the general case is more complicated, because when differences of solutions or approximate solutions are taken, the Itô integrals no longer vanish. Thus, the Gronwall inequalities cannot be applied directly. Instead, we will use Gronwall to control second moments. Different arguments are needed for existence and uniqueness. Continuous dependence on initial conditions can be proved using arguments similar to those used for the uniqueness proof; the details are left as an exercise.

To prove existence of solutions we use the same iterative method as in the case of constant σ to generate approximate solutions:

$$X_0(t) = x \quad \text{and} \quad X_{n+1}(t) = x + \int_0^t \mu(X_n(s))\,ds + \int_0^t \sigma(X_n(s))\,dW_s. \tag{12}$$

By induction, the processes $X_n(t)$ are well-defined and have continuous paths. The problem is to show that these converge uniformly on compact time intervals, and that the limit process is a solution to the stochastic differential equation.

First we will show that for each $t \geq 0$ the sequence of random variables $X_n(t)$ converges in L^2 to a random variable $X(t)$, necessarily in L^2. The first two terms of the sequence are $X_0(t) \equiv x$ and $X_1(t) = x + \mu(x)t + \sigma(x)W_t$; for both of these the random variables $X_j(t)$ are uniformly bounded in L^2 for t in any bounded interval $[0, T]$, and so for each $T < \infty$ there exists $C = C_T < \infty$ such that

$$E(X_1(t) - X_0(t))^2 \leq C \quad \text{for all } \ t \leq T.$$

Now by hypothesis, the functions μ and σ are uniformly Lipschitz, and hence, for a suitable constant $B < \infty$,

$$\begin{aligned} |\mu(X_n(t)) - \mu(X_{n-1}(t))| &\leq B|X_n(t) - X_{n-1}(t)| \quad \text{and} \\ |\sigma(X_n(t)) - \sigma(X_{n-1}(t))| &\leq B|X_n(t) - X_{n-1}(t)| \end{aligned} \tag{13}$$

for all $t \geq 0$. Thus, by Cauchy-Schwartz and the Itô isometry, together with the elementary inequality $(x+y)^2 \leq 2x^2 + 2y^2$,

$$\begin{aligned} E|X_{n+1}(t) - X_n(t)|^2 &\leq E\left(\int_0^t (\mu(X_n(s)) - \mu(X_{n-1}(s))\, ds + \int_0^t (\sigma(X_n(s)) - \sigma(X_{n-1}(s)))\, dW_s\right)^2 \\ &\leq 2E\left(\int_0^t (\mu(X_n(s)) - \mu(X_{n-1}(s))\, ds\right)^2 + 2E\left(\int_0^t (\sigma(X_n(s)) - \sigma(X_{n-1}(s))\, dW_s\right)^2 \\ &\leq 2B^2 E\left(\int_0^t |X_n(s) - X_{n-1}(s)|\, ds\right)^2 + 2B^2 \int_0^t E|X_n(s) - X_{n-1}(s)|^2\, ds \\ &\leq 2B^2 E\left(t\int_0^t |X_n(s) - X_{n-1}(s)|^2\, ds\right) + 2B^2 \int_0^t E|X_n(s) - X_{n-1}(s)|^2\, ds \\ &\leq 2B^2(T+1)\int_0^t E|X_n(s) - X_{n-1}(s)|^2\, ds \ \ \forall\, t \leq T. \end{aligned}$$

Lemma 2 now applies to $y_n(t) := E|X_{n+1}(t) - X_n(t)|^2$ (recall that $E|X_1(t) - X_0(t)|^2 \leq C = C_T$ for all $t \leq T$), yielding

$$E(X_{n+1}(t) - X_n(t))^2 \leq C(4B^2 + 4B^2T)^n/n! \quad \forall\, t \leq T. \tag{14}$$

This clearly implies that for each $t \leq T$ the random variables $X_n(t)$ converge in L^2. Furthermore, this L^2-convergence is uniform for $t \leq T$ (because the bounds in (14) hold uniformly for $t \leq T$), and the limit random variables $X(t) := L^2 - \lim_{n\to\infty} X_n(t)$ are bounded in L^2 for $t \leq T$.

It remains to show that the limit process $X(t)$ satisfies the stochastic differential equation (6). To this end, consider the random variables $\mu(X_n(t))$ and $\sigma(X_n(t))$. Since $X_n(t) \to X(t)$ in L^2, the Lipschitz bounds (13) imply that

$$\lim_{n\to\infty} \left(E|\mu(X_n(t)) - \mu(X(t))|^2 + E|\sigma(X_n(t)) - \sigma(X(t))|^2\right) = 0$$

uniformly for $t \leq T$. Hence, by the Itô isometry,

$$L^2 - \lim_{n\to\infty} \int_0^t \sigma(X_n(s))\, dW_s = \int_0^t \sigma(X(s))\, dW_s$$

for each $t \leq T$. Similarly, by Cauchy-Schwartz and Fubini,

$$L^2 - \lim_{n\to\infty} \int_0^t \mu(X_n(s))\, ds = \int_0^t \mu(X(s))\, ds.$$

Thus, (12) implies that

$$X(t) = x + \int_0^t \mu(X(s))\,ds + \int_0^t \sigma(X(s))\,dW_s.$$

This shows that the process $X(t)$ satisfies the stochastic integral equation (6). Both of the integrals in this equation are continuous in t, and therefore so is $X(t)$.

□

2.5 Proof of Theorem 1. General Case: Uniqueness

Suppose as before that for some initial value x there are two continuous solutions

$$X_t = x + \int_0^t \mu(X_s)\,ds + \int_0^t \sigma(X_s)\,dW_s \quad \text{and}$$
$$Y_t = x + \int_0^t \mu(Y_s)\,ds + \int_0^t \sigma(Y_s)\,dW_s.$$

Then the difference satisfies

$$Y_t - X_t = \int_0^t (\mu(Y_s) - \mu(X_s))\,ds + \int_0^t (\sigma(Y_s) - \sigma(X_s))\,dW_s \tag{15}$$

Although the second integral cannot be bounded pathwise, its second moment can be bounded, since $\sigma(y)$ is Lipschitz:

$$E\left\{\int_0^t (\sigma(Y_s) - \sigma(X_s))\,dW_s\right\}^2 \leq B^2 \int_0^t E(Y_s - X_s)^2\,ds,$$

where B is the Lipschitz constant. Of course, we have no way of knowing that the expectations $E(Y_s - X_s^2)$ are finite, so the integral on the right side of the inequality may be ∞. Nevertheless, taking second moments on both sides of (15), using the inequality $(a+b)^2 \leq 2a^2 + 2b^2$ and the Cauchy-Schwartz inequality, we obtain

$$E(Y_t - X_t)^2 \leq (2B^2 + 2B^2T) \int_0^t E(Y_s - X_s)^2\,ds$$

If the function $f(t) := E(Y_t - X_t)^2$ were known to be finite and integrable on compact time intervals, then the Gronwall inequality (9) would imply that $f(t) \equiv 0$, and the proof of uniqueness would be complete.[1] To circumvent this difficulty, we use a localization argument: Define the stopping time

$$\tau := \tau_A = \inf\{t\,:\,X_t^2 + Y_t^2 \geq A\}.$$

Since X_t and Y_t are defined and continuous for all t, they are a.s. bounded on compact time intervals, and so $\tau_A \to \infty$ as $A \to \infty$. Hence, with probability one, $t \wedge \tau_A = t$ for all sufficiently large A. Next, starting from the identity (15), stopping at time $\tau = \tau_A$, and proceeding as in the last paragraph, we obtain

$$E(Y_{t\wedge\tau} - X_{t\wedge\tau})^2 \leq (2B^2 + 2B^2T) \int_0^t E(Y_{s\wedge\tau} - X_{s\wedge\tau})^2\,ds \quad \text{for all } t \leq T$$

[1] OKSENDAL seems to have fallen prey to this trap: In his proof of Theorem 5.2.1 he fails to check that the second moment is finite.

By definition of τ, both sides are finite, and so Gronwall's inequality (9) implies that

$$E(Y_{t\wedge\tau} - X_{t\wedge\tau})^2 = 0$$

Since this is true for every $\tau = \tau_A$, it follows that $X_t = Y_t$ a.s., for each $t \geq 0$. SInce X_t and Y_t have continuous sample paths, it follows that with probability one, $X_t = Y_t$ for all $t \geq 0$. A similar argument proves continuous dependence on initial conditions.

□

3 Example: The Feller diffusion

The Feller diffusion $\{Y_t\}_{t\geq 0}$ is a continuous-time Markov process on the half-line $[0, \infty)$ with absorption at 0 that satisfies the stochastic differential equation

$$dY_t = \sigma\sqrt{Y_t}\,dW_t \tag{16}$$

up until the time $\tau = \tau_0$ of the first visit to 0. Here $\sigma > 0$ is a positive parameter. The Itô existence/uniqueness theorem does not apply, at least directly, because the function $\sqrt{y}$ is not Lipschitz. However, the localization lemma of Itô calculus can be used in a routine fashion to show that for any initial value $y > 0$ there is a continuous process Y_t such that

$$Y_{t\wedge\tau} = y + \int_0^{t\wedge\tau} \sqrt{Y_s}\,dW_s \quad \text{where} \quad \tau = \inf\{t > 0 \,:\, Y_t = 0\}.$$

(Exercise: Fill in the details.)

The importance of the Feller diffusion stems from the fact that it is the natural continuous-time analogue[2] of the critical Galton-Watson process. The Galton-Watson process is a discrete-time Markov chain Z_n on the nonnegative integers that evolves according to the following rule: Given that $Z_n = k$ and any realization of the past up to time $n-1$, the random variable Z_{n+1} is distributed as the sum of k independent, identically distributed random variables with common distribution F, called the *offspring distribution*. The process is said to be *critical* if F has mean 1. Assume also that F has finite variance σ^2; then the evolution rule implies that the increment $Z_{n+1} - Z_n$ has conditional expectation 0 and conditional variance $\sigma^2 Z_n$, given the history of the process to time n. This corresponds to the stochastic differential equation (16), which roughly states that the increments of Y_t have conditional expectation 0 and conditional variance $\sigma^2 Y_t\,dt$, given $\mathcal{F}_t$.

A natural question to ask about the Feller diffusion is this: If $Y_0 = y > 0$, does the trajectory Y_t reach the endpoint 0 of the state space in finite time? (That is, is $\tau < \infty$ w.p.1?) To see that it does, consider the process $Y_t^{1/2}$. By Itô's formula, if Y_t satisfies (16), or more precisely, if it satisfies

$$Y_t = y + \int_0^t \sigma\sqrt{Y_s}\mathbf{1}_{[0,\tau]}(s)\,dW_s, \tag{17}$$

then

$$\begin{aligned} dY_t^{1/2} &= \frac{1}{2}Y_t^{-1/2}dY_t - \frac{1}{8}Y_t^{-3/2}d[Y]_t \\ &= \frac{\sigma}{2}dW_t - \frac{1}{8}Y_t^{-1/2}\,dt \end{aligned}$$

up to time τ. Thus, up to the time of the first visit to 0 (if any), the process $Y_t^{1/2}$ is a Brownian motion plus a negative drift. Since a Brownian motion started at $\sqrt{y}$ will reach 0 in finite time, with probability one, so will $Y_t^{1/2}$.

[2]Actually, the Feller diffusion is more than just an analogue of the Galton-Watson process: It is a weak limit of rescaled Galton-Watson processes, in the same sense that Brownian Motion is a weak limit of rescaled random walks.

Exercise 1. Scaling law for the Feller diffusion: Let Y_t be a solution of the integral equation (17) with volatility parameter $\sigma > 0$ and initial value $Y_0 = 1$.

(A) Show that for any $\alpha > 0$ the process

$$\tilde{Y}_t := \alpha^{-1} Y_{\alpha t} \tag{18}$$

is a Feller diffusion with initial value α^{-1} and volatility parameter σ.

(B) Use this to deduce a simple relationship between the distributions of the hitting time τ for the Feller diffusion under the different initial conditions $Y_0 = 1$ and $Y_0 = \alpha^{-1}$, respectively.

Exercise 2. Superposition law for the Feller diffusion: Let Y_t^A and Y_t^B be independent Feller diffusion processes with initial values $Y_0^A = \alpha$ and $Y_0^B = \beta$: In particular, assume that Y^A and Y^B satisfy stochastic integral equations (17) with respect to independent Brownian motions W^A and W^B. Define the *superposition* of Y^A and Y^B to be the process

$$Y_t^C := Y_t^A + Y_t^B.$$

(A) Show that Y_t^C is a Feller diffusion with initial condition $Y_0^C = \alpha + \beta$.

(B) Use this to deduce a simple relationship among the hitting time distributions for the three processes.

Exercise 3. Zero is not an entrance boundary: The stochastic differential equation (16) has a *singularity* at the endpoint 0 of the state space, in the sense that the volatility $\sigma\sqrt{y}$ becomes 0 in a non-smooth manner as $y \to 0$. Equation (16) has solutions up to time τ for any initial value $Y_0 = y > 0$; however, it is unclear whether or not there are solutions Y_t of (16) such that $\lim_{t\to 0} Y_t = 0$. Use the scaling law to prove that there are no such solutions. HINT: Consider the time that it would take to get from 2^{-k-m} to 2^{-m}.

4 The Differential Operator Associated to an SDE

We have seen that the Laplacian differential operator $\frac{1}{2}\Delta$ plays a central role in the study of Brownian motion:

(A) The transition probabilities $p_t(x, y)$ obey the backward and forward heat equation:

$$\frac{\partial}{\partial t} p_t(x, y) = \frac{1}{2}\Delta_x p_t(x, y) = \frac{1}{2}\Delta_y p_t(x, y).$$

(B) For any C^∞ function f with compact support, the *Dynkin formula* holds:

$$E^x f(W_t) = f(x) + \frac{1}{2} E^x \int_0^t \Delta f(W_s)\, ds.$$

(C) For any C^∞ function f with compact support, the *Itô formula* holds:

$$f(W_t) - f(W_0) = \int_0^t \nabla f \cdot dW_t + \frac{1}{2}\int_0^t \Delta f(W_s)\, ds.$$

Associated to any autonomous, time-homogeneous stochastic differential equation of the form (1) is a corresponding second-order differential operator, the so-called *generator*, which figures into the study of solutions X_t in much the same way as does the Laplacian for Brownian motion. The generator is defined as follows:

$$\mathcal{G} = \mu(x)\frac{d}{dx} + \frac{1}{2}\sigma(x)^2\frac{d^2}{dx^2}. \tag{19}$$

Assume henceforth that the coefficients $\mu(x)$ and $\sigma(x)$ are continuous.

Proposition 1. *Let X_t be any weak solution to the stochastic differential equation* (1) *and let $f : R \to \mathbb{R}$ be a C^∞ function with compact support. Then*

$$f(X_t) - f(X_0) = \int_0^t \mu(X_s)f'(X_s)\,ds + \int_0^t f'(X_s)\sigma(X_s)\,dW_s + \frac{1}{2}\int_0^t \sigma(X_s)^2 f''(X_s)\,ds. \tag{20}$$

Consequently, the process

$$M_t^f := f(X_t) - \int_0^t \mathcal{G}f(X_s)\,ds \tag{21}$$

is a martingale.

Proof. Formula (20) is just a restatement of the Itô formula for Itô processes – see the Lecture Notes on Itô Calculus, sec. 2.2. The second assertion is a direct consequence of formula (20), because

$$M_t^f = \int_0^t f'(X_s)\sigma(X_s)\,dW_s$$

and the integrand $f'(X_s)\sigma(X_s)$ is uniformly bounded, since f has compact support. □

Equations (20) and (21) are the analogues of the Itô and Dynkin formulas, respectively. There is also an analogue of Property (A) (the heat equation), but this is, unfortunately, much more difficult to establish. In particular, it is by no means obvious that transition densities even exist (i.e., that for a given initial value $X_0 = x$ the random variable X_t will have an absolutely continuous distribution). Following is a weak analogue of Property (A).

Proposition 2. *Let X_t be any weak solution to the stochastic differential equation* (1) *and let $u : [0,T]\times\mathbb{R} \to R$ be a continuous, bounded function that satisfies the diffusion equation*

$$\frac{\partial}{\partial t}u(t,x) = \mathcal{G}u(t,x) \quad \textit{for } 0 < t < T. \tag{22}$$

Then the process $\{u(T-t, X_t)\}_{0\le t\le T}$ is a martingale, and so

$$u(T,x) = E^x u(0, X_T). \tag{23}$$

Proof. It is implicit in the hypotheses that the function u is of class $C^{1\times 2}$ in $(0,T)\times\mathbb{R}$, and so the Itô formula applies in this range: in particular,

$$\begin{aligned} du(T-t, X_t) &= -\frac{\partial}{\partial t}u(T-t,X_t)\,dt + \frac{\partial}{\partial x}u(T-t,X_t)\,dX_t + \frac{\partial^2}{\partial x^2}u(T-t,X_t)\,d[X]_t \\ &= \left(-\frac{\partial}{\partial t} + \mathcal{G}\right)u(T-t,X_t)\,dt + \frac{\partial}{\partial x}u(T-t,X_t)\,dW_t \\ &= \frac{\partial}{\partial x}u(T-t,X_t)\,dW_t. \end{aligned}$$

This implies that the process $u(T-t, X_t)$ is a *local* martingale. But since the function u is bounded, it then follows by the dominated convergence theorem that $u(T-t), X_t$ is a proper martingale. □

5 Distributional Uniqueness and the Strong Markov Property

Recall that a *weak* solution to the stochastic differential equation (1) need not be adapted to the minimal filtration for the driving Brownian motion, so the random variables X_t might somehow incorporate some "exogenous randomness". Thus, one might wonder if the distribution of the process $\{X_t\}_{t\geq 0}$ gotten from a weak solution could depend in a nontrivial way on the particular filtration.

Definition 3. Say that *distributional uniqueness* holds for the stochastic differential equation (1) if any two weak solutions X_t and $\tilde{X}_t$ with the same initial value x have the same finite-dimensional distributions.

When does distributional uniqueness hold? The answer, it turns out, depends on properties of the generator $\mathcal{G}$ associated with the stochastic differential equation.

Definition 4. Let $f : \mathbb{R} \to \mathbb{R}$ be a bounded, continuous function. The *Cauchy problem* (more properly, the *Cauchy initial value problem*) for the data $\mathcal{G}, f$ is the partial differential equation

$$\frac{\partial}{\partial t} u(t, x) = \mathcal{G} u(t, x) \tag{24}$$

with initial condition

$$u(0, x) = f(x). \tag{25}$$

Theorem 2. *If for every C^∞ function $f : \mathbb{R} \to \mathbb{R}$ with compact support the Cauchy problem has a unique, bounded solution $u : [0, \infty) \times \mathbb{R} \to \mathbb{R}$ then distributional uniqueness holds for the stochastic differential equation* (1).

Thus, the problem of distributional uniqueness devolves to a problem in the theory of partial differential equations. What do the PDE folks have to tell us about this problem? Quite a lot, it seems (cf. Avner Friedman, *Partial differential equations of parabolic type* for an entire book on the subject). Following is a simple sufficient condition.

Theorem 3. *If the coefficients $\mu(x)$ and $\sigma(x)$ are bounded and continuous, then for every C^∞ function $f : \mathbb{R} \to \mathbb{R}$ with compact support the Cauchy problem has a unique, bounded solution.*

I will not prove this, but I will remark that probability theory gives an explicit form for the solution: if X_t^x is any weak solution of the stochastic differential equation (1) with initial value x, then

$$u(t, x) = Ef(X_t^x) \tag{26}$$

is the unique solution to the Cauchy problem for the initial data f. Distributional uniqueness implies that the expectation in (26) does not depend on the particular weak solution used, and a simple application of the Itô formula (essentially the same calculation as in the proof of Proposition 2) would show that (26) is a solution of the Cauchy problem if we could show that u is $C^{1\times 2}$. Unfortunately, there seems to be no easy way to prove this directly.

Proof of Theorem 2. For any C^∞ function $f : \mathbb{R} \to \mathbb{R}$ with compact support, let $u_f(t, x)$ be the unique solution to the Cauchy problem (24)–(25) with initial data f. To prove Theorem 2, we must show that for any weak solution X_t to the stochastic differential equation (1) with initial value $X_0 = x$ and any choice of times $0 =< t_1 < \cdots < t_m < \infty$ the joint distribution of $X_{t_1}, X_{t_2}, \ldots, X_{t_m}$ is completely determined by the functions $\mu(x)$ and $\sigma(x)$. By a routine induction (EXERCISE), it suffices to show that for any $s, t > 0$ and any C^∞ function f with compact support,

$$E(f(X_{t+s}) \mid \mathcal{F}_s) = u_f(t, X_s) \quad \text{almost surely.} \tag{27}$$

But this follows easily from Proposition 2: since u_f satisfies the partial differential equation (24),the process $\{u_f(t-r, X_{s+r})\}_{0\leq r\leq t}$ is a martingale relative to the filtration $\{\mathcal{F}_{s+r}\}_{0\leq r\leq t}$, and so

$$E(f(X_{t+s})\,|\,\mathcal{F}_s) = u_f(t, X_s) \quad \text{almost surely.}$$

□

Finally, we shall show that distributional uniqueness implies that any weak solution X_t to the stochastic differential equation (1) is a strong Markov process. Because the laws that govern the evolution of the process X_t are not spatially homogeneous, the strong Markov property cannot be formulated in quite the same way as for Brownian motion.

Theorem 4. *Assume that distributional uniqueness holds for the stochastic differential equation* (1)*and that the coefficients $\mu(x), \sigma(x)$ satisfy the hypotheses of Ito's Existence/Uniqueness Theorem. Then any weak solution $(X_t)_{t\geq 0}$ with non-random initial point $X_0 = x$ has the strong Markov property. More precisely:*

For any $x \in \mathbb{R}$, denote by μ_x the distribution[3] *of any weak solution $(X_t)_{t\geq 0}$ of* (1) *with initial point $X_0 = x$. Denote by $(\mathcal{F}_t)_{t\geq 0}$ the filtration for the weak solution $(X_t)_{t\geq 0}$, and let τ be any finite stopping time for this filtration. Then the conditional distribution of the post-τ process $(X_{\tau+t})_{t\geq 0}$ given the $\sigma-$algebra $\mathcal{F}_\tau$ is μ_{X_τ}.*

Proof. By an induction argument similar to that in the proof of Theorem 2, it suffices to show that for any C^∞ function f with compact support and any $t > 0$,

$$E^x(f(X_{\tau+t})\,|\,\mathcal{F}_\tau) = u_f(t, X_\tau)$$

where u_f is defined by equation (26). But the process $(\tilde{X}_t)_{t\geq 0} = (X_{\tau+t})_{t\geq 0}$ solves the stochastic differential equation

$$d\tilde{X}_t = \mu(\tilde{X}_t)\,dt + \sigma(\tilde{X}_t)\,d\tilde{W}_t \tag{28}$$

where $\tilde{W}_t = W_{\tau+t} - W_\tau$ and the relevant filtration is $(\mathcal{F}_{\tau+t})_{t\geq 0}$. By Itô's theorem, for any $y \in \mathbb{R}$ the equation (28) has a unique strong solution with initial value $\tilde{X}_0 = y$. By hypothesis, this solution has the same distribution as does the process X_t with initial value $X_0 = y$; consequently,

$$E^x(f(X_{\tau+t})\,|\,\mathcal{F}_\tau) = (E^y f(X_t))_{y=X_\tau} = u_f(t, X_\tau).$$

□

[3]that is, the measure on the space $C[0,\infty)$ of continuous paths $x(t)$ such that for any Borel set $B \subset C[0,\infty)$,

$$P((X_t)_{t\geq 0} \in B) = \mu_x(B).$$