Chapter 26
Stochastic Differential Equations

Stochastic differential equations describe the time evolution of certain continuous Markov processes with values in $\mathbb{R}^n$. In contrast with classical differential equations, in addition to the derivative of the function, there is a term that describes the random fluctuations that are coded as an Itô integral with respect to a Brownian motion. Depending on how seriously we take the concrete Brownian motion as the driving force of the noise, we speak of strong and weak solutions. In the first section, we develop the theory of strong solutions under Lipschitz conditions for the coefficients. In the second section, we develop the so-called (local) martingale problem as a method of establishing weak solutions. In the third section, we present some examples in which the method of duality can be used to prove weak uniqueness.

As stochastic differential equations are a very broad subject, and since things quickly become very technical, we only excursively touch some of the most important results, partly without proofs, and illustrate them with examples.

26.1 Strong Solutions

Consider a stochastic differential equation (SDE) of the type

$$\begin{aligned} X_0 &= \xi, \\ dX_t &= \sigma(t, X_t)\, dW_t + b(t, X_t)\, dt. \end{aligned} \tag{26.1}$$

Here $W = (W^1, \ldots, W^m)$ is an m-dimensional Brownian motion, ξ is an $\mathbb{R}^n$-valued random variable with distribution μ that is independent of W, $\sigma(t,x) = (\sigma_{ij}(t,x))_{\substack{i=1,\ldots,n \\ j=1,\ldots,m}}$ is a real $n \times m$ matrix and $b(t,x) = (b_i(t,x))_{i=1,\ldots,n}$ is an n-dimensional vector. Assume the maps $(t,x) \mapsto \sigma_{ij}(t,x)$ and $(t,x) \mapsto b_i(t,x)$ are measurable.

A. Klenke, *Probability Theory*, Universitext, DOI 10.1007/978-1-4471-5361-0_26,

By a solution of (26.1) we understand a continuous adapted stochastic process X with values in $\mathbb{R}^n$ that satisfies the integral equation

$$X_t = \xi + \int_0^t \sigma(s, X_s)\, dW_s + \int_0^t b(s, X_s)\, ds \quad \mathbf{P}\text{-a.s. for all } t \geq 0. \tag{26.2}$$

Written in full, this is

$$X_t^i = \xi^i + \sum_{j=1}^m \int_0^t \sigma_{ij}(s, X_s)\, dW_s^j + \int_0^t b_i(s, X_s)\, ds \quad \text{for all } i = 1, \ldots, n.$$

Now the following problem arises: To which filtration $\mathbb{F}$ do we wish X to be adapted? Should it be the filtration that is generated by ξ and W, or do we allow $\mathbb{F}$ to be larger? Already for ordinary differential equations, depending on the equation, uniqueness of the solution may fail (although existence is usually not a problem); for example, for $f' = |f|^{1/3}$. If $\mathbb{F}$ is larger than the filtration generated by W, then we can define further random variables that select one out of a variety of possible solutions. We thus have more possibilities for solutions than if $\mathbb{F} = \sigma(W)$. Indeed, it will turn out that in some situations for the existence of a solution, it is necessary to allow a larger filtration. Roughly speaking, X is a strong solution of (26.1) if (26.2) holds and if X is adapted to $\mathbb{F} = \sigma(W)$. On the other hand, X is a weak solution if X is adapted to a larger filtration $\mathbb{F}$ with respect to which W is still a martingale. Weak solutions will be dealt with in Section 26.2.

Definition 26.1 (Strong solution) We say that the stochastic differential equation (SDE) (26.1) has a *strong solution* X if there exists a map $F : \mathbb{R}^n \times C([0, \infty); \mathbb{R}^m) \to C([0, \infty); \mathbb{R}^n)$ with the following properties:

(i) For every $t \geq 0$, the map $(x, w) \mapsto F(x, w)$ is measurable with respect to $\mathcal{B}(\mathbb{R}^n) \otimes \mathcal{G}_t^m - \mathcal{G}_t^n$, where (for $k = m$ or $k = n$) $\mathcal{G}_t^k := \sigma(\pi_s : s \in [0, t])$ is the σ-algebra generated by the coordinate maps $\pi_s : C([0, \infty); \mathbb{R}^k) \to \mathbb{R}$, $w \mapsto w(s)$.
(ii) The process $X = F(\xi, W)$ satisfies (26.2).

Condition (i) says that the path $(X_s)_{s\in[0,t]}$ depends only on ξ and $(W_s)_{s\in[0,t]}$ but not on further information. In particular, X is adapted to $\mathcal{F}_t = \sigma(\xi, W_s : s \in [0, t])$ and is progressively measurable; hence the Itô integral in (26.2) is well-defined if σ and b do not grow too quickly for large x.

Remark 26.2 Clearly, a strong solution of an SDE is a generalized n-dimensional diffusion. If the coefficients σ and b are independent of t, then the solution is an n-dimensional diffusion. ◇

Remark 26.3 Let X be a strong solution and let F be as in Definition 26.1. If W' is an m-dimensional Brownian motion on a space $(\Omega', \mathcal{F}', \mathcal{P}')$ with filtration $\mathbb{F}'$, and if ξ' is independent of W' and is $\mathcal{F}_0'$-measurable, then $X' = F(\xi', W')$ satisfies the integral equation (26.2). Hence, it is a strong solution of (26.1) with W replaced by W'. Thus the existence of a strong solution does not depend on the actual realization of the Brownian motion or on the filtration $\mathbb{F}$. ◊

Definition 26.4 We say that the SDE (26.1) has a unique strong solution if there exists an F as in Definition 26.1 such that:

(i) If W is an m-dimensional Brownian motion on some probability space $(\Omega, \mathcal{F}, \mathbf{P})$ with filtration $\mathbb{F}$ and if ξ is an $\mathcal{F}_0$-measurable random variable that is independent of W and such that $\mathbf{P} \circ \xi^{-1} = \mu$, then $X := F(\xi, W)$ is a solution of (26.2).
(ii) For every solution (X, W) of (26.2), we have $X = F(\xi, W)$.

Example 26.5 Let $m = n = 1$, $b \in \mathbb{R}$ and $\sigma > 0$. The *Ornstein–Uhlenbeck process*

$$X_t := e^{bt}\xi + \sigma \int_0^t e^{(t-s)b}\, dW_s, \quad t \ge 0, \tag{26.3}$$

is a strong solution of the SDE $X_0 = \xi$ and

$$dX_t = \sigma\, dW_t + bX_t\, dt.$$

In the language of Definition 26.1, we have (in the sense of the pathwise Itô integral with respect to w)

$$F(x, w) = \left(t \mapsto e^{bt}x + \int_0^t e^{(t-s)b}\, dw(s) \right)$$

for all $w \in \mathcal{C}_{\mathrm{qv}}$ (that is, with continuous square variation). Since $\mathbf{P}[W \in \mathcal{C}_{\mathrm{qv}}] = 1$, we can define $F(x, w) = 0$ for $w \in C([0, \infty); \mathbb{R}) \setminus \mathcal{C}_{\mathrm{qv}}$.

Indeed, by Fubini's theorem for Itô integrals, we have (Exercise 25.3.1)

$$\begin{aligned}
\xi &+ \int_0^t \sigma\, dW_s + \int_0^t bX_s\, ds \\
&= \xi + \sigma W_t + \int_0^t be^{bs}\xi\, ds + \int_0^t \sigma b \left(\int_0^s e^{b(s-r)}\, dW_r \right) ds \\
&= \xi + \sigma W_t + \left(e^{bt} - 1\right)\xi + \int_0^t \sigma \left(\int_r^t be^{b(s-r)}\, ds \right) dW_r \\
&= e^{bt}\xi + \int_0^t \left(\sigma + \left(e^{b(t-r)} - 1\right)\sigma\right) dW_r \\
&= X_t.
\end{aligned}$$

It can be shown (see Theorem 26.8) that the solution is also (strongly) unique. ◊

Example 26.6 Let $\alpha, \beta \in \mathbb{R}$. The one-dimensional SDE $X_0 = \xi$ and

$$dX_t = \alpha X_t\, dW_t + \beta X_t\, dt \tag{26.4}$$

has the strong solution

$$X_t = \xi \exp\left(\alpha W_t + \left(\beta - \frac{\alpha^2}{2}\right)t\right).$$

In the language of Definition 26.1, we have $\sigma(t,x) = \alpha x$, $b(t,x) = \beta x$ and

$$F(x,w) = \left(t \mapsto x \exp\left(\alpha w(t) + \left(\beta - \frac{\alpha^2}{2}\right)t\right)\right)$$

for all $w \in C([0,\infty);\mathbb{R})$ and $x \in \mathbb{R}$. Indeed, by the time-dependent Itô formula (Corollary 25.35),

$$X_t = \xi + \int_0^t \alpha X_s \, dW_s + \int_0^t \left(\left(\beta - \frac{\alpha^2}{2}\right) + \frac{1}{2}\alpha^2\right) X_s \, ds.$$

Also in this case, we have strong uniqueness of the solution (see Theorem 26.8). The process X is called a *geometric Brownian motion* and, for example, serves in the so-called *Black–Scholes model* as the process of stock prices. ◊

We give a simple criterion for existence and uniqueness of strong solutions. For an $n \times m$ matrix A, define the *Hilbert–Schmidt norm*

$$\|A\| = \sqrt{\operatorname{trace}\left(AA^T\right)} = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{m} A_{i,j}^2}. \tag{26.5}$$

For $b \in \mathbb{R}^n$, we use the Euclidean norm $\|b\|$. Since all norms on finite-dimensional vector spaces are equivalent, it is not important exactly which norm we use. However, the Hilbert–Schmidt norm simplifies the computations, as the following lemma shows.

Lemma 26.7 *Let $t \mapsto H(t) = (H_{ij}(t))_{i=1,\dots,n,\,j=1,\dots,m}$ be progressively measurable and $\mathbf{E}[\int_0^T H_{ij}^2(t)\,dt] < \infty$ for all i, j. Then*

$$\mathbf{E}\left[\left\| \int_0^T H(t)\,dW_t \right\|^2\right] = \mathbf{E}\left[\int_0^T \|H(t)\|^2 \, dt\right], \tag{26.6}$$

where $\|H\|$ is the Hilbert–Schmidt norm from (26.5).

Proof For $i = 1, \dots, n$, the process $I_i(t) := \sum_{j=1}^m \int_0^t H_{ij}(s)\,dW_s^j$ is a continuous martingale with square variation process $\langle I_i \rangle_t = \int_0^t \sum_{j=1}^m H_{ij}^2(s)\,ds$. Hence

$$\mathbf{E}\left[\left(I_i(T)\right)^2\right] = \mathbf{E}\left[\int_0^T \sum_{j=1}^m H_{ij}^2(s)\,ds\right].$$

The left-hand side in (26.6) equals

$$\sum_{i=1}^{n} \mathbf{E}\big[\big(I_i(T)\big)^2\big] = \mathbf{E}\left[\int_0^T \sum_{i=1}^{n}\sum_{j=1}^{m} H_{ij}^2(s)\,ds\right].$$

Hence the claim follows by the definition of $\|H(s)\|^2$. □

Theorem 26.8 *Let b and σ be Lipschitz continuous in the first coordinate. That is, we assume that there exists a $K > 0$ such that, for all $x, x' \in \mathbb{R}^n$ and $t \geq 0$,*

$$\big\|\sigma(x,t) - \sigma\big(x',t\big)\big\| + \big\|b(x,t) - b\big(x',t\big)\big\| \leq K\,\big\|x - x'\big\|. \tag{26.7}$$

Further, assume the growth condition

$$\big\|\sigma(t,x)\big\|^2 + \big\|b(t,x)\big\|^2 \leq K^2\big(1 + \|x\|^2\big) \quad \text{for all } x \in \mathbb{R}^n, t \geq 0. \tag{26.8}$$

Then, for every initial point $X_0 = x \in \mathbb{R}^n$, there exists a unique strong solution X of the SDE (26.1). This solution is a Markov process and in the case where σ and b do not depend on t, it is a strong Markov process.

As the main tool, we need the following lemma.

Lemma 26.9 (Gronwall) *Let $f, g : [0,T] \to \mathbb{R}$ be integrable and let $C > 0$ such that*

$$f(t) \leq g(t) + C\int_0^t f(s)\,ds \quad \text{for all } t \in [0,T]. \tag{26.9}$$

Then

$$f(t) \leq g(t) + C\int_0^t e^{C(t-s)} g(s)\,ds \quad \text{for all } t \in [0,T].$$

In particular, if $g(t) \equiv G$ is constant, then $f(t) \leq Ge^{Ct}$ for all $t \in [0,T]$.

Proof Let $F(t) = \int_0^t f(s)\,ds$ and $h(t) = F(t)e^{-Ct}$. Then, by (26.9),

$$\frac{d}{dt}h(t) = f(t)e^{-Ct} - CF(t)e^{-Ct} \leq g(t)e^{-Ct}.$$

Integration yields

$$F(t) = e^{Ct}h(t) \leq \int_0^t e^{C(t-s)} g(s)\,ds.$$

Substituting this into (26.9) gives

$$f(t) \leq g(t) + CF(t) \leq g(t) + C\int_0^t g(s)e^{C(t-s)}\,ds.$$

□

Proof of Theorem 26.8 It is enough to show that, for every $T < \infty$, there exists a unique strong solution up to time T.

Uniqueness. We first show uniqueness of the solution. Let X and X' be two solutions of (26.2). Then

$$X_t - X'_t = \int_0^t \big(b(s, X_s) - b\big(s, X'_s\big)\big)\, ds + \int_0^t \big(\sigma(s, X_s) - \sigma\big(s, X'_s\big)\big)\, dW_s.$$

Hence

$$\|X_t - X'_t\|^2 \le 2\left\|\int_0^t \big(b(s, X_s) - b\big(s, X'_s\big)\big)\, ds\right\|^2 + 2\left\|\int_0^t \big(\sigma(s, X_s) - \sigma\big(s, X'_s\big)\big)\, dW_s\right\|^2. \tag{26.10}$$

For the first summand in (26.10), use the Cauchy–Schwarz inequality, and for the second one use Lemma 26.7 to obtain

$$\mathbf{E}\big[\|X_t - X'_t\|^2\big] \le 2t \int_0^t \mathbf{E}\big[\big\|b(s, X_s) - b\big(s, X'_s\big)\big\|^2\big]\, ds + 2\int_0^t \mathbf{E}\big[\big\|\sigma(s, X_s) - \sigma\big(s, X'_s\big)\big\|^2\big]\, ds.$$

Write $f(t) = \mathbf{E}[\|X_t - X'_t\|^2]$ and $C := 2(T+1)K^2$. Then $f(t) \le C\int_0^t f(s)\, ds$. Hence Gronwall's lemma (with $g \equiv 0$) yields $f \equiv 0$.

Existence. We use a version of the Picard iteration scheme. For $N \in \mathbb{N}_0$, recursively define processes X^N by $X_t^0 \equiv x$ and

$$X_t^N := x + \int_0^t b\big(s, X_s^{N-1}\big)\, ds + \int_0^t \sigma\big(s, X_s^{N-1}\big)\, dW_s \quad \text{for } N \in \mathbb{N}. \tag{26.11}$$

Using the growth condition (26.8), it can be shown inductively that

$$\int_0^T \mathbf{E}\big[\|X_t^N\|^2\big]dt \le 2(T+1)K^2\left(T + \int_0^T \mathbf{E}\big[\|X_t^{N-1}\|^2\big]\, dt\right) \le \big(2T(T+1)K^2\big)^N\big(1 + \|x\|^2\big) < \infty, \quad N \in \mathbb{N}.$$

Hence, at each step, the Itô integral is well-defined.

Consider now the differences

$$X_t^{N+1} - X_t^N = I_t + J_t,$$

where

$$I_t := \int_0^t \big(\sigma\big(s, X_s^N\big) - \sigma\big(s, X_s^{N-1}\big)\big)\, dW_s$$

and

$$J_t := \int_0^t \big(b\big(s, X_s^N\big) - b\big(s, X_s^{N-1}\big)\big)\, ds.$$

By applying Doob's L^2-inequality to the nonnegative submartingale $(\|I_t\|^2)_{t\ge 0}$, using Lemma 26.7 and (26.7), we obtain

$$\begin{aligned}\mathbf{E}\Big[\sup_{s\le t}\|I_s\|^2\Big] &\le 4\mathbf{E}\big[\|I_t\|^2\big]\\ &= 4\mathbf{E}\bigg[\int_0^t \big\|\sigma\big(s, X_s^N\big) - \sigma\big(s, X_s^{N-1}\big)\big\|^2\, ds\bigg]\\ &\le 4K^2\int_0^t \mathbf{E}\big[\big\|X_s^N - X_s^{N-1}\big\|^2\big]\, ds. \end{aligned} \tag{26.12}$$

For J_t, by the Cauchy–Schwarz inequality, we get

$$\|J_t\|^2 \le t\int_0^t \big\|b\big(s, X_s^N\big) - b\big(s, X_s^{N-1}\big)\big\|^2\, ds.$$

Hence

$$\begin{aligned}\mathbf{E}\Big[\sup_{s\le t}\|J_s\|^2\Big] &\le t\mathbf{E}\bigg[\int_0^t \big\|b\big(s, X_s^N\big) - b\big(s, X_s^{N-1}\big)\big\|^2\, ds\bigg]\\ &\le tK^2\int_0^t \mathbf{E}\big[\big\|X_s^N - X_s^{N-1}\big\|^2\big]\, ds. \end{aligned} \tag{26.13}$$

Defining

$$\Delta^N(t) := \mathbf{E}\Big[\sup_{s\le t}\big\|X_s^N - X_s^{N-1}\big\|^2\Big],$$

and $C := 2K^2(4+T) \vee 2(T+1)K^2(1+\|x\|^2)$, we obtain (using the growth condition (26.8))

$$\Delta^{N+1}(t) \le C\int_0^t \Delta^N(s)\, ds \quad \text{for } N \ge 1$$

and

$$\begin{aligned}\Delta^1(t) &\le 2t\int_0^t \big\|b(s,x)\big\|^2\, ds + 2\int_0^t \big\|\sigma(s,x)\big\|^2\, ds\\ &\le 2(T+1)K^2\big(1+\|x\|^2\big)\cdot t \le Ct.\end{aligned}$$

Inductively, we get $\Delta^N(t) \le \frac{(Ct)^N}{N!}$. Thus, by Markov's inequality,

$$\sum_{N=1}^{\infty} \mathbf{P}\Big[\sup_{s\le t}\big\|X_s^N - X_s^{N-1}\big\|^2 > 2^{-N}\Big] \le \sum_{N=1}^{\infty} 2^N \Delta^N(t)$$

$$\le \sum_{N=1}^{\infty} \frac{(2Ct)^N}{N!} \le e^{2Ct} < \infty.$$

Using the Borel–Cantelli lemma, we infer $\sup_{s\le t} \|X_s^N - X_s^{N-1}\|^2 \stackrel{N\to\infty}{\longrightarrow} 0$ a.s. Hence a.s. $(X^N)_{N\in\mathbb{N}}$ is a Cauchy sequence in the Banach space $(C([0,T]), \|\cdot\|_\infty)$. Therefore, X^N converges a.s. uniformly to some X. As uniform convergence implies convergence of the integrals, X is a strong solution of (26.2).

Markov property. The strong Markov property follows from the strong Markov property of the Brownian motion that drives the SDE. □

We have already seen some important examples of this theorem. Many interesting problems, however, lead to stochastic differential equations with coefficients that are not Lipschitz continuous. In the one-dimensional case, using special comparison methods, one can show that it is sufficient that σ is Hölder-continuous of order $\frac{1}{2}$ in the space variable.

Theorem 26.10 (Yamada–Watanabe) *Consider the one-dimensional situation where $m = n = 1$. Assume that there exist $K < \infty$ and $\alpha \in [\frac{1}{2}, 1]$ such that, for all $t \ge 0$ and $x, x' \in \mathbb{R}$, we have*

$$\big|b(t,x) - b\big(t,x'\big)\big| \le K\big|x - x'\big| \quad \text{and} \quad \big|\sigma(t,x) - \sigma\big(t,x'\big)\big| \le \big|x - x'\big|^\alpha.$$

Then, for every $X_0 \in \mathbb{R}$, the SDE (26.1) has a unique strong solution X and X is a strong Markov process.

Proof See [172] or [85, Proposition 5.2.13] and [49, Theorem 5.3.11] for existence and uniqueness. The strong Markov property follows from Theorem 26.26. □

Example 26.11 Consider the one-dimensional SDE

$$dX_t = \sqrt{\gamma X_t^+}\, dW_t + a\big(b - X_t^+\big)\, dt \tag{26.14}$$

with initial point $X_0 = x \ge 0$, where $\gamma > 0$ and $a, b \ge 0$ are parameters. The conditions of Theorem 26.10 are fulfilled with $\alpha = \frac{1}{2}$ and $K = \sqrt{\gamma} + a$. Obviously, the unique strong solution X remains nonnegative if $X_0 \ge 0$. (In fact, it can be shown that $X_t > 0$ for all $t > 0$ if $2ab/\gamma \ge 1$, and that X_t hits zero arbitrarily often with probability 1 if $2ab/\gamma < 1$. See, e.g., [78, Example IV.8.2, p. 237]. Compare Example 26.16. See Figs. 26.1 and 26.2 for computer simulations.)

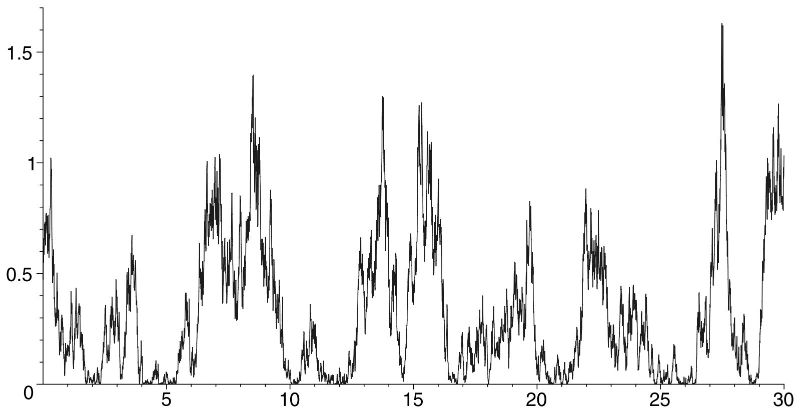

Fig. 26.1 Cox–Ingersoll–Ross diffusion with parameters $\gamma = 1$, $b = 1$ and $a = 0.3$. The path hits zero again and again since $2ab/\gamma = 0.6 < 1$

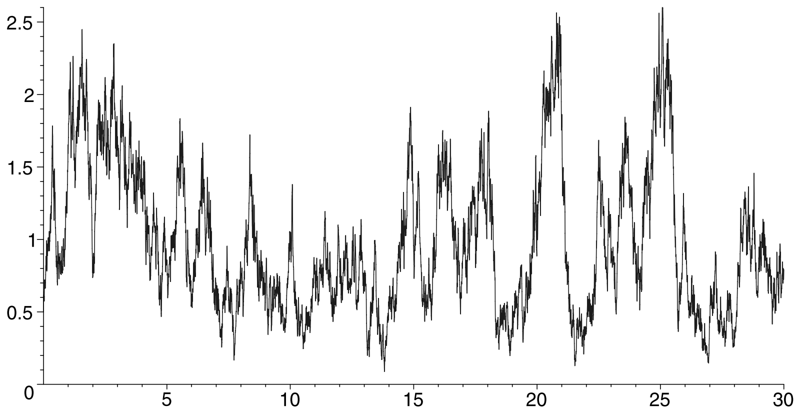

Fig. 26.2 Cox–Ingersoll–Ross diffusion with parameters $\gamma = 1$, $b = 1$ and $a = 2$. The path never hits zero since $2ab/\gamma = 4 \geq 1$

Depending on the context, this process is sometimes called *Feller's branching diffusion with immigration* or the *Cox–Ingersoll–Ross model* for the time evolution of interest rates.

For the case $a = b = 0$, use the Itô formula to compute that

$$e^{-\lambda X_t} - e^{-\lambda x} - \gamma \frac{\lambda^2}{2} \int_0^t e^{-\lambda X_s} X_s \, ds = \lambda \int_0^t e^{-\lambda X_s} \sqrt{\gamma X_s} \, dW_s$$

is a martingale. Take expectations for the Laplace transform $\varphi(t, \lambda, x) = \mathbf{E}_x[e^{-\lambda X_t}]$ to get the differential equation

$$\frac{d}{dt}\varphi(t, \lambda, x) = \gamma \frac{\lambda^2}{2}\mathbf{E}\big[X_t e^{-\lambda X_t}\big] = -\frac{\gamma\lambda^2}{2}\frac{d}{d\lambda}\varphi(t, \lambda, x).$$

With initial value $\varphi(0, \lambda, x) = e^{-\lambda x}$, the unique solution is

$$\varphi(t, \lambda, x) = \exp\left(-\frac{\lambda}{(\gamma/2)\lambda t + 1}x\right).$$

However (for $\gamma = 2$), this is exactly the Laplace transform of the transition probabilities of the Markov process that we defined in Theorem 21.48 and that in Lindvall's theorem (Theorem 21.51) we encountered as the limit of rescaled Galton–Watson branching processes. ◊

Exercise 26.1.1 Let $a, b \in \mathbb{R}$. Show that the stochastic differential equation

$$dX_t = \frac{b - X_t}{1 - t}\,dt + dW_t$$

with initial value $X_0 = a$ has a unique strong solution for $t \in [0, 1)$ and that $X_1 := \lim_{t\uparrow 1} X_1 = b$ almost surely. Furthermore, show that the process $Y = (X_t - a - t(b-a))_{t\in[0,1]}$ can be described by the Itô integral

$$Y_t = (1-t)\int_0^t (1-s)^{-1}\,dW_s, \qquad t \in [0, 1),$$

and is hence a Brownian bridge (compare Exercise 21.5.3).

26.2 Weak Solutions and the Martingale Problem

In the last section, we studied strong solutions of the stochastic differential equation

$$dX_t = \sigma(t, X_t)\,dW_t + b(t, X_t)\,dt. \tag{26.15}$$

A strong solution is a solution where any path of the Brownian motion W gets mapped onto a path of the solution X. In this section, we will study the notion of a weak solution where additional information (or additional noise) can be used to construct the solution.

Definition 26.12 (Weak solution of an SDE) A *weak solution* of (26.15) with initial distribution $\mu \in \mathcal{M}_1(\mathbb{R}^n)$ is a triple

$$L = \big((X, W), (\Omega, \mathcal{F}, \mathbf{P}), \mathbb{F}\big),$$

where

- $(\Omega, \mathcal{F}, \mathbf{P})$ is a probability space,
- $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ is a filtration on $(\Omega, \mathcal{F}, \mathbf{P})$ that satisfies the usual conditions,
- W is a Brownian motion on $(\Omega, \mathcal{F}, \mathbf{P})$ and is a martingale with respect to $\mathbb{F}$,
- X is continuous and adapted (hence progressively measurable),
- $\mathbf{P} \circ (X_0)^{-1} = \mu$, and
- (X, W) satisfies

$$X_t = X_0 + \int_0^t \sigma(s, X_s)\, dW_s + \int_0^t b(s, X_s)\, ds \quad \mathbf{P}\text{-a.s.} \tag{26.16}$$

A weak solution L is called (weakly) unique if, for any further solution L' with initial distribution μ, we have $\mathbf{P}' \circ (X')^{-1} = \mathbf{P} \circ X^{-1}$.

Remark 26.13 Clearly, a weak solution of an SDE is a generalized n-dimensional diffusion. If the coefficients σ and b do not depend on t, then the solution is an n-dimensional diffusion. ♢

Remark 26.14 Clearly, every strong solution of (26.15) is a weak solution. The converse is false, as the following example shows. ♢

Example 26.15 Consider the SDE (with initial value $X_0 = 0$)

$$dX_t = \operatorname{sign}(X_t)\, dW_t, \tag{26.17}$$

where $\operatorname{sign} = \mathbb{1}_{(0,\infty)} - \mathbb{1}_{(-\infty,0)}$ is the sign function. Then

$$X_t = X_0 + \int_0^t \operatorname{sign}(X_s)\, dW_s \quad \text{for all } t \geq 0 \tag{26.18}$$

if and only if

$$W_t = \int_0^t dW_s = \int_0^t \operatorname{sign}(X_s)\, dX_s \quad \text{for all } t \geq 0. \tag{26.19}$$

A weak solution of (26.17) is obtained as follows. Let X be a Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and $\mathbb{F} = \sigma(X)$. If we define W by (26.19), then W is a continuous $\mathbb{F}$-martingale with square variation

$$\langle W \rangle_t = \int_0^1 \big(\operatorname{sign}(X_s)\big)^2\, ds = t.$$

Thus, by Lévy's characterization (Theorem 25.28), W is a Brownian motion. Hence $((X, W), (\Omega, \mathcal{F}, \mathbf{P}), \mathbb{F})$ is a weak solution of (26.17).

In order to show that a strong solution does not exist, take any weak solution and show that X is not adapted to $\sigma(W)$. Since, by (26.18), X is a continuous martingale with square variation $\langle X \rangle_t = t$, X is a Brownian motion.

Let $F_n \in C^2(\mathbb{R})$ be a convex even function with derivatives F_n' and F_n'' such that

$$\sup_{x\in\mathbb{R}}\big|F_n(x)-|x|\big| \overset{n\to\infty}{\longrightarrow} 0,$$

$|F_n'(x)| \le 1$ for all $x \in \mathbb{R}$ and $F_n'(x) = \operatorname{sign}(x)$ for $|x| > \frac{1}{n}$. In particular, we have

$$\int_0^t \big(F_n'(X_s) - \operatorname{sign}(X_s)\big)^2\, ds \overset{n\to\infty}{\longrightarrow} 0 \quad \text{a.s.}$$

and thus

$$\int_0^t F_n'(X_s)\, dX_s \overset{n\to\infty}{\longrightarrow} \int_0^t \operatorname{sign}(X_s)\, dX_s \quad \text{in } L^2. \tag{26.20}$$

By passing to a subsequence, if necessary, we may assume that almost sure convergence holds in (26.20).

Since F_n'' is even, we have

$$\begin{aligned} W_t &= \int_0^t \operatorname{sign}(X_s)\, dX_s = \lim_{n\to\infty}\int_0^t F_n'(X_s)\, dX_s \\ &= \lim_{n\to\infty}\left(F_n(X_t) - F_n(0) - \frac{1}{2}\int_0^t F_n''(X_s)\, ds\right) \\ &= |X_t| - \lim_{n\to\infty}\frac{1}{2}\int_0^t F_n''\big(|X_s|\big)\, ds. \end{aligned}$$

As the right-hand side depends only on $|X_s|$, $s \in [0,t]$, W is adapted to $\mathbb{G} := (\sigma(|X_s| : s \in [0,t]))$. Hence $\sigma(W) \subset \mathbb{G} \subsetneqq \sigma(X)$, and thus X is not adapted to $\sigma(W)$. ◊

Example 26.16 Let $B = (B^1, \ldots, B^n)$ be an n-dimensional Brownian motion started at $y \in \mathbb{R}^n$. Let $x := \|y\|^2$, $X_t := \|B_t\|^2 = (B_t^1)^2 + \ldots + (B_t^n)^2$ and

$$W_t := \sum_{i=1}^n \int_0^t \frac{1}{\sqrt{X_s}} B_s^i\, dB_s^i.$$

Then W is a continuous local martingale with $\langle W\rangle_t = t$ for every $t \ge 0$ and

$$X_t = x + nt + \int_0^t \sqrt{X_s}\, dW_s.$$

That is, (X, W) is a weak solution of the SDE $dX_t = \sqrt{2X_t}\, dW_t + n\, dt$. X is called an n-dimensional *Bessel process*. By Theorem 25.42, B (and thus X) hits the origin for some $t > 0$ if and only if $n = 1$. Clearly, we can define X also for noninteger $n \ge 0$. One can show that X hits zero if and only if $n \le 1$. Compare Example 26.11. ◊

For the connection between existence and uniqueness of weak solutions and strong solutions, we only quote here the theorem of Yamada and Watanabe.

Definition 26.17 (Pathwise uniqueness) A solution of the SDE (26.15) with initial distribution μ is said to be *pathwise unique* if, for every $\mu \in \mathcal{M}_1(\mathbb{R}^n)$ and for any two weak solutions (X, W) and (X', W) on the same space $(\Omega, \mathcal{F}, \mathbf{P})$ with the same filtration $\mathbb{F}$, we have $\mathbf{P}[X_t = X'_t \text{ for all } t \geq 0] = 1$.

Theorem 26.18 (Yamada and Watanabe) *The following are equivalent.*

(i) *The SDE* (26.15) *has a unique strong solution.*
(ii) *For any* $\mu \in \mathcal{M}_1(\mathbb{R}^n)$, (26.15) *has a weak solution, and pathwise uniqueness holds.*

If (i) *and* (ii) *hold, then the solution is weakly unique.*

Proof See [172], [147, pp. 151ff] or [78, pp. 163ff]. □

Example 26.19 Let X be a weak solution of (26.17). Then $-X$ is also a weak solution; that is, pathwise uniqueness does not hold (although it can be shown that the solution is weakly unique; see Theorem 26.25). ◊

Consider the one-dimensional case $m = n = 1$. If X is a solution (strong or weak) of (26.15), then

$$M_t := X_t - \int_0^t b(s, X_s)\, ds$$

is a continuous local martingale with square variation

$$\langle M \rangle_t = \int_0^t \sigma^2(s, X_s)\, ds.$$

We will see that this characterizes a weak solution of (26.15) (under some mild growth conditions on σ and b).

Now assume that, for all $t \geq 0$ and $x \in \mathbb{R}^n$, the $n \times n$ matrix $a(t, x)$ is symmetric and nonnegative definite, and let $(t, x) \mapsto a(t, x)$ be measurable.

Definition 26.20 An n-dimensional continuous process X is called a solution of the *local martingale problem* for a and b with initial condition $\mu \in \mathcal{M}_1(\mathbb{R}^n)$ (briefly, LMP(a, b, μ)) if $\mathbf{P} \circ X_0^{-1} = \mu$ and if, for every $i = 1, \ldots, n$,

$$M_t^i := X_t^i - \int_0^t b_i(s, X_s)\, ds, \qquad t \geq 0,$$

is a continuous local martingale with quadratic covariation

$$\left\langle M^i, M^j \right\rangle_t = \int_0^t a_{ij}(s, X_s)\, ds \quad \text{for all } t \geq 0,\, i, j = 1, \ldots, n.$$

We say that the solution of *LMP*(a, b, μ) is unique if, for any two solutions X and X', we have $\mathbf{P} \circ X^{-1} = \mathbf{P} \circ (X')^{-1}$.

Denote by σ^T the transposed matrix of σ. Clearly, $a = \sigma\sigma^T$ is a nonnegative semidefinite symmetric $n \times n$ matrix.

Theorem 26.21 *X is a solution of $LMP(\sigma\sigma^T, b, \mu)$ if and only if (on a suitable extension of the probability space) there exists a Brownian motion W such that (X, W) is a weak solution of* (26.15).

In particular, there exists a unique weak solution of the SDE (26.15) *with initial distribution μ if $LMP(\sigma\sigma^T, b, \mu)$ is uniquely solvable.*

Proof We show the statement only for the case $m = n = 1$. The general case needs some consideration on the roots of nonnegative semidefinite symmetric matrices, which, however, do not yield any further insight into the stochastics of the problem. For this we refer to [85, Proposition 5.4.6].

"$\Longleftarrow$" If (X, W) is a weak solution, then, by Corollary 25.19, X solves the local martingale problem.

"$\Longrightarrow$" Let X be a solution of $LMP(\sigma^2, b, \mu)$. By Theorem 25.29, on an extension of the probability space there exists a Brownian motion $\tilde{W}$ such that $M_t = \int_0^t |\sigma(s, X_s)| \, d\tilde{W}_s$. If we define

$$W_t := \int_0^t \operatorname{sign}\big(\sigma(s, X_s)\big) \, d\tilde{W}_s,$$

then $M_t = \int_0^t \sigma(s, X_s) \, dW_s$ and hence (X, W) is a weak solution of (26.15). □

In some sense, a local martingale problem is a very natural way of writing a stochastic differential equation; that is:

> X locally has derivative (drift) b and additionally has random normally distributed fluctuations of size σ.

Here, a concrete Brownian motion does not appear. In fact, in most problems its occurrence is rather artificial. Just as Markov chains are described by their transition probabilities and not by a concrete realization of the random transitions (as in Theorem 17.17), many continuous (space and time) processes are most naturally described by the drift and the *size* of the fluctuations but not by the concrete realization of the random fluctuations.

From a technical point of view, the formulation of a stochastic differential equation as a local martingale problem is very convenient since it makes SDEs accessible to techniques such as martingale inequalities and approximation theorems that can be used to establish existence and uniqueness of solutions. Here we simply quote two important results.

Theorem 26.22 (Existence of solutions) *Let $(t, x) \mapsto b(t, x)$ and $(t, x) \mapsto a(t, x)$ be continuous and bounded. Then, for every $\mu \in \mathcal{M}_1(\mathbb{R}^n)$, there exists a solution X of the $LMP(a, b, \mu)$.*

Proof See [147, Theorem V.23.5]. □

Definition 26.23 The *LMP*(a, b) is said to be *well-posed* if, for every $x \in \mathbb{R}^n$, there exists a unique solution X of *LMP*(a, b, δ_x).

Remark 26.24 If σ and b satisfy the Lipschitz conditions of Theorem 26.8, then the LMP$(\sigma\sigma^T, b)$ is well-posed. This follows by Theorem 26.8, Theorem 26.18 and Theorem 26.21. ◊

In the following, we assume

$$(t, x) \mapsto \sigma(t, x) \text{ resp. } (t, x) \mapsto a(t, x) \text{ is bounded on compact sets.} \tag{26.21}$$

This condition ensures the equivalence of the local martingale problems to the somewhat more common martingale problem (see [85, Proposition 5.4.11]).

Theorem 26.25 (Uniqueness in the martingale problem) *Assume* (26.21) *and that, for any* $x \in \mathbb{R}^n$, *there exists a solution* X^x *of LMP*(a, b, δ_x). *The distribution of* X^x *will be denoted by* $\mathbf{P}_x := \mathbf{P} \circ (X^x)^{-1}$.

Assume that, for any two solutions X^x *and* Y^x *of LMP*(a, b, δ_x), *we have*

$$\mathbf{P} \circ \left(X_T^x\right)^{-1} = \mathbf{P} \circ \left(Y_T^x\right)^{-1} \quad \textit{for any } T \geq 0. \tag{26.22}$$

Then LMP(a, b) *is well-posed, and the canonical process* X *is a strong Markov process with respect to* $(\mathbf{P}_x, x \in \mathbb{R}^n)$. *If* $a = \sigma\sigma^T$, *then under* $\mathbf{P}_x$, *the process* X *is the unique weak solution of the SDE* (26.15).

Proof See [49, Theorem 4.4.2 and Problem 49] and [85, Proposition 5.4.11]. □

A fundamental strength of this theorem is that we do not need to check the uniqueness of the whole process but only have to check in (26.22) the one-dimensional marginal distributions. We will use this in Section 26.3 in some examples.

The existence of solutions of a stochastic differential equation (or equivalently of a local martingale problem) is often easier to show than the uniqueness of solutions. We know already that Lipschitz conditions for the coefficients b and σ (not $\sigma\sigma^T$!) ensure uniqueness (Theorem 26.8 and Theorem 26.18), as here strong uniqueness of the solution holds.

At first glance, it might seem confusing that random fluctuations have a stabilising effect on the solution. That is, there are deterministic differential equations whose solution is unique only after adding random noise terms. For example, consider the following equation:

$$dX_t = \operatorname{sign}(X_t)|X_t|^{1/3}\, dt + \sigma\, dW_t, \qquad X_0 = 0. \tag{26.23}$$

If $\sigma = 0$, then the deterministic differential equation has a continuum of solutions that can be parameterized by $v \in \{-1, +1\}$ and $T \geq 0$, namely $X_t = v2\sqrt{2}(t - T)^{3/2}\mathbb{1}_{\{t>T\}}$. If $\sigma > 0$, then the noise eliminates the instability of (26.23) at $x = 0$. We quote the following theorem for the time-independent case from [147, Theorem V.24.1] (see also [161, Chapter 10]).

Theorem 26.26 (Stroock–Varadhan) *Let $a_{ij} : \mathbb{R}^n \to \mathbb{R}$ be continuous and let $b_i : \mathbb{R}^n \to \mathbb{R}$ be measurable for $i, j = 1, \ldots, n$. Assume*

(i) *$a(x) = (a_{ij}(x))$ is symmetric and strictly positive definite for every $x \in \mathbb{R}^n$,*
(ii) *there exists a $C < \infty$ such that, for all $x \in \mathbb{R}^n$ and $i, j = 1, \ldots, n$, we have*

$$\left|a_{ij}(x)\right| \le C\left(1 + \|x\|^2\right) \quad \text{and} \quad \left|b_i(x)\right| \le C\left(1 + \|x\|\right).$$

Then the LMP(a, b) is well-posed and the SDE (26.15) has a unique strong solution that is a strong Markov process. The solution X has the Feller property: For every $t > 0$ and every bounded measurable $f : \mathbb{R}^n \to \mathbb{R}$, the map $x \mapsto \mathbf{E}_x[f(X_t)]$ is continuous.

We will present explicit examples in Section 26.3. Here we just remark that we have developed a particular method in order to construct Markov processes, namely as the solution of a stochastic differential equation or of a local martingale problem. In the framework of models in discrete time, in Section 17.2 and especially in Exercise 17.2.1, we characterized certain Markov chains as solutions of martingale problems. In order for drift and square variation to be sufficient for uniqueness of the Markov chain described by the martingale problem, it was essential that, for any step of the chain, we only allowed three possibilities. Here, however, the decisive restriction is the continuity of the processes.

Exercise 26.2.1 Consider the time-homogeneous one-dimensional case ($m = n = 1$). Let σ and b be such that, for every $X_0 \in \mathbb{R}$, there exists a unique weak solution of

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

that is a strong Markov process. Further, assume that there exists an $x_0 \in \mathbb{R}$ with

$$C := \int_{-\infty}^{\infty} \frac{1}{\sigma^2(x)} \exp\left(\int_{x_0}^{x} \frac{2b(r)}{\sigma^2(r)}\,dr\right) dr < \infty.$$

(i) Show that the measure $\pi \in \mathcal{M}_1(\mathbb{R})$ with density

$$\frac{\pi(dx)}{dx} = C^{-1} \frac{1}{\sigma^2(x)} \exp\left(\int_{x_0}^{x} \frac{2b(r)}{\sigma^2(r)}\,dr\right)$$

is an invariant distribution for X.
(ii) For which values of b does the Ornstein–Uhlenbeck process $dX_t = \sigma\,dW_t + bX_t\,dt$ have an invariant distribution? Determine this distribution and compare the result with what could be expected by an explicit computation using the representation in (26.3).
(iii) Compute the invariant distribution of the Cox–Ingersoll–Ross SDE (26.14) (i.e., Feller's branching diffusion).

(iv) Let $\gamma, c > 0$ and $\theta \in (0,1)$. Show that the invariant distribution of the solution X of the SDE on $[0,1]$,

$$dX_t = \sqrt{\gamma X_t(1-X_t)}\, dW_t + c(\theta - X_t)\, dt$$

is the Beta distribution $\beta_{2c\gamma/\theta, 2c\gamma/(1-\theta)}$.

Exercise 26.2.2 Let $\gamma > 0$. Let X^1 and X^2 be solutions of $dX_t^i = \sqrt{\gamma X_t^i} dW_t^i$, where W^1 and W^2 are two independent Brownian motions with initial values $X_0^1 = x_0^1 > 0$ and $X_0^2 = x_0^2 > 0$. Show that $Z := X^1 + X^2$ is a weak solution of $Z_0 = 0$ and $dZ_t = \sqrt{\gamma Z_t} dW_t$.

26.3 Weak Uniqueness via Duality

The Stroock–Varadhan theorem provides a strong criterion for existence and uniqueness of solutions of stochastic differential equations. However, in many cases, the condition of locally uniform ellipticity of a (Condition (i) in Theorem 26.26) is not fulfilled. This is the case, in particular, if the solutions are defined only on subsets of $\mathbb{R}^n$.

Here we will study a powerful tool that in many special cases can yield weak uniqueness of solutions.

Definition 26.27 (Duality) Let $X = (X^x, x \in E)$ and $Y = (Y^y, y \in E')$ be families of stochastic processes with values in the spaces E and E', respectively, and such that $X_0^x = x$ a.s. and $Y_0^y = y$ a.s. for all $x \in E$ and $y \in E'$. We say that X and Y are *dual* to each other with *duality function* $H : E \times E' \to \mathbb{C}$ if, for all $x \in E$, $y \in E'$ and $t \geq 0$, the expectations $\mathbf{E}[H(X_t^x, y)]$ and $\mathbf{E}[H(x, Y_t^y)]$ exist and are equal:

$$\mathbf{E}\big[H\big(X_t^x, y\big)\big] = \mathbf{E}\big[H\big(x, Y_t^y\big)\big].$$

In the following, we assume that $\sigma_{ij} : \mathbb{R}^n \to \mathbb{R}$ and $b_i : \mathbb{R}^n \to \mathbb{R}$ are bounded on compact sets for all $i = 1, \ldots, n$, $j = 1, \ldots, m$. Consider the time-homogeneous stochastic differential equation

$$dX_t = \sigma(X_t)\, dW_t + b(X_t)\, dt. \tag{26.24}$$

Theorem 26.28 (Uniqueness via duality) *Assume that, for every $x \in \mathbb{R}^n$, there exists a solution of the local martingale problem for $(\sigma\sigma^T, b, \delta_x)$. Further, assume that there exists a family $(Y^y, y \in E')$ of Markov processes with values in the measurable space $(E', \mathcal{E}')$ and a measurable map $H : \mathbb{R}^n \times E' \to \mathbb{C}$ such that, for every $y \in E'$, $x \in \mathbb{R}^n$ and $t \geq 0$, the expectation $\mathbf{E}[H(x, Y_t^y)]$ exists and is finite. Further, let $(H(\cdot, y), y \in E')$ be a separating class of functions for $\mathcal{M}_1(\mathbb{R}^n)$ (see Definition* 13.9).

For every $x \in \mathbb{R}^n$ and every solution X^x of LMP$(\sigma\sigma^T, b, \delta_x)$, assume that the duality equation holds:

$$\mathbf{E}\big[H\big(X_t^x, y\big)\big] = \mathbf{E}\big[H\big(x, Y_t^y\big)\big] \quad \textit{for all } y \in \mathbf{E}', t \geq 0. \tag{26.25}$$

Then the local martingale problem of $(\sigma\sigma^T, b)$ is well-posed and hence (26.24) has a unique weak solution that is a strong Markov process.

Proof By Theorem 26.25, it is enough to check that, for every $x \in \mathbb{R}^n$, every solution X^x of LMP$(\sigma\sigma^T, b, \delta_x)$ and every $t \geq 0$, the distribution $\mathbf{P} \circ (X_t^x)^{-1}$ is unique. Since $(H(\cdot, y), y \in E')$ is a separating class of functions, this follows from (26.16). □

Example 26.29 (Wright–Fisher diffusion) Consider the Wright–Fisher SDE

$$dX_t = \mathbb{1}_{[0,1]}(X_t)\sqrt{\gamma X_t(1 - X_t)}\, dW_t, \tag{26.26}$$

where $\gamma > 0$ is a parameter. See Fig. 26.3 for a computer simulation. By Theorem 26.22, for every $x \in \mathbb{R}$, there exists a weak solution $(\tilde{X}, W)$ of (26.26). $\tilde{X}$ is a continuous local martingale with square variation

$$\langle \tilde{X} \rangle_t = \int_0^t \gamma \tilde{X}_s(1 - \tilde{X}_s)\mathbb{1}_{[0,1]}(\tilde{X}_s)\, ds.$$

Let $\tau := \inf\{t > 0 : \tilde{X}_t \notin [0, 1]\}$ and let $X := \tilde{X}^\tau$ be the process stopped at τ. Then X is a continuous bounded martingale with

$$\langle X \rangle_t = \int_0^t \gamma X_s(1 - X_s)\mathbb{1}_{[0,1]}(X_s)\, ds.$$

Hence, (X, W) is a solution of (26.26). By construction, $X_t \in [0, 1]$ for all $t \geq 0$ if $X_0 = \tilde{X}_0 \in [0, 1]$.

Let $\tau' := \inf\{t > 0 : \tilde{X}_t \in [0, 1]\}$. If $\tilde{X}_0 \notin [0, 1]$, then $\tau' > 0$ since $\tilde{X}$ is continuous. Since $\tilde{X}^{\tau'}$ is a continuous local martingale with $\langle \tilde{X}^{\tau'} \rangle \equiv 0$, we have $\tilde{X}_t^{\tau'} = \tilde{X}_0$ for all $t \geq 0$. However, this implies $\tilde{X}_t = \tilde{X}_0$ for all $t < \tau'$. Again, by continuity of $\tilde{X}$, we get $\tau' = \infty$ and $\tilde{X}_t = \tilde{X}_0$ for all $t \geq 0$.

Hence, it is enough to show uniqueness of the solution for $\tilde{X}_0 = x \in [0, 1]$. To this end, let $Y = (Y_t)_{t \geq 0}$ be the Markov process on $\mathbb{N}$ with Q-matrix

$$q(m, n) = \begin{cases} \gamma\binom{m}{2}, & \text{if } n = m - 1, \\ -\gamma\binom{m}{2}, & \text{if } n = m, \\ 0, & \text{else.} \end{cases}$$

We show duality of X and Y with respect to $H(x, n) = x^n$:

$$\mathbf{E}_x\big[X_t^n\big] = \mathbf{E}_n\big[x^{N_t}\big] \quad \text{for all } t \geq 0, x \in [0, 1], n \in \mathbb{N}. \tag{26.27}$$

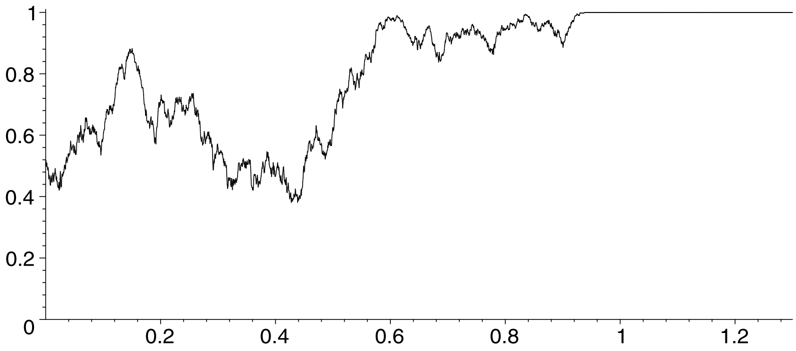

Fig. 26.3 Simulation of a Wright–Fisher diffusion with parameter $\gamma = 1$

Define $m^{x,n}(t) = \mathbf{E}_x[X_t^n]$ and $g^{x,n}(t) = \mathbf{E}_n[x^{N_t}]$. By the Itô formula,

$$X_t^n - x^n - \int_0^t \gamma \binom{n}{2} X_s^{n-1}(1 - X_s)\, ds = \int_0^t n X_s^{n-1} \sqrt{\gamma X_s (1 - X_s)}\, dW_s$$

is a martingale.

Taking expectations, we obtain the following recursive equations for the moments of X:

$$\begin{aligned} m^{x,1}(t) &= x, \\ m^{x,n}(t) &= x^n + \gamma \binom{n}{2} \int_0^t \big(m^{x,n-1}(s) - m^{x,n}(s)\big)\, ds \quad \text{for } n \geq 2. \end{aligned} \tag{26.28}$$

Clearly, this system of linear differential equations can be uniquely solved recursively in n.

Due to the Markov property of Y, for $h > 0$ and $t \geq 0$, we have

$$\begin{aligned} g^{x,n}(t+h) = \mathbf{E}_n\big[x^{Y_{t+h}}\big] &= \mathbf{E}_n\big[\mathbf{E}_{Y_h}\big[x^{Y_t}\big]\big] \\ &= \sum_{m=1}^{n} \mathbf{P}_n[Y_h = m] \mathbf{E}_m\big[x^{Y_t}\big] \\ &= \sum_{m=1}^{n} \mathbf{P}_n[Y_h = m] g^{x,m}(t). \end{aligned}$$

This implies

$$\begin{aligned} \frac{d}{dt} g^{x,n}(t) &= \lim_{h \downarrow 0} h^{-1} \big[g^{x,n}(t+h) - g^{x,n}(t)\big] \\ &= \lim_{h \downarrow 0} h^{-1} \sum_{m=1}^{n} \mathbf{P}_n[Y_h = m] \big(g^{x,m}(t) - g^{x,n}(t)\big) \end{aligned}$$

$$= \sum_{m=1}^{n} q(n,m) g^{x,m}(t)$$
$$= \gamma \binom{n}{2} \big(g^{x,n-1}(t) - g^{x,n}(t)\big). \tag{26.29}$$

Evidently, $g^{x,1}(t) = x$ for all $x \in [0,1]$ and $t \geq 0$ and $g^{x,n}(0) = x^n$. That is, $g^{x,n}$ solves (26.28), and thus (26.27) holds.

By Theorem 15.4, the family $(H(\cdot, n), n \in \mathbb{N}) \subset C([0,1])$ is separating for $\mathcal{M}_1([0,1])$; hence the conditions of Theorem 26.28 are fulfilled. Therefore, X is the unique weak solution of (26.26) and is a strong Markov process. ◊

Remark 26.30 The martingale problem for the Wright–Fisher diffusion is almost identical to the martingale problem for the Moran model (see Example 17.22) $M^N = (M_n^N)_{n \in \mathbb{N}_0}$ with population size N: M^N is a martingale with values in the set $\{0, 1/N, \ldots, (N-1)/N, 1\}$ and with square variation process

$$\langle M^N \rangle_n = \frac{2}{N^2} \sum_{k=0}^{n-1} M_k^N \big(1 - M_k^N\big).$$

At each step, M^N can either stay put or increase or decrease by $1/N$. In Exercise 17.2.1, we saw that this determines the process M^N uniquely. Similarly as in Theorem 21.51 for branching processes, it can be shown that the time-rescaled Moran processes $\tilde{M}_t^N = M_{\lfloor N^2 t \rfloor}^N$ converge to the Wright–Fisher diffusion with $\gamma = 2$. The Wright–Fisher diffusion thus occurs as the limiting model of a genealogical model and describes the gene frequency (that is, the fraction) of a certain allele in a population that fluctuates randomly due to resampling. ◊

Example 26.31 (Feller's branching diffusion) Let $(Z_n^N)_{n \in \mathbb{N}_0}$ be a Galton–Watson branching process with critical geometric offspring distribution $p_k = 2^{-k-1}$, $k \in \mathbb{N}_0$ and $Z_0^N = N$ for any $N \in \mathbb{N}$. Then Z^N is a discrete martingale and we have

$$\mathbf{E}\big[\big(Z_n^N - Z_{n-1}^N\big)^2 \,\big|\, Z_{n-1}^N\big] = Z_{n-1}^N \left(\sum_{k=0}^{\infty} p_k k^2 - 1 \right) = 2 Z_{n-1}^N.$$

Hence Z^N has square variation

$$\langle Z^N \rangle_n = \sum_{k=0}^{n-1} 2 Z_k^N.$$

Define the linearly interpolated version

$$Z_t^N := \big(t - N^{-1} \lfloor tN \rfloor\big)\big(Z_{\lfloor tN \rfloor + 1}^N - Z_{\lfloor tN \rfloor}^N\big) + \frac{1}{n} Z_{\lfloor tN \rfloor}^N$$

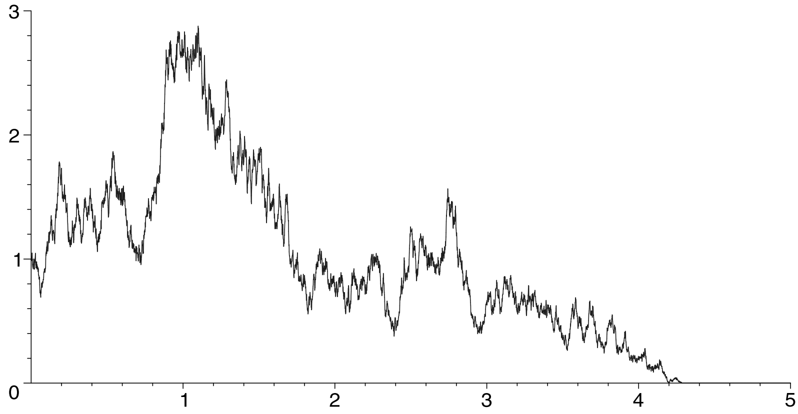

Fig. 26.4 Simulation of Feller's branching diffusion with parameter $\gamma = 1$

of $N^{-1}Z^N_{\lfloor tN \rfloor}$. By Lindvall's theorem (Theorem 21.51), there is a continuous Markov process Z such that $Z^N \overset{N\to\infty}{\longrightarrow} Z$ in distribution. See Fig. 26.4 for a computer simulation of Z. Since it can be shown that the moments also converge, we have that Z is a continuous martingale with square variation

$$\langle Z \rangle_t = \int_0^t 2Z_s \, ds.$$

In fact, in Example 26.11, we have already shown that Z is the unique solution of the SDE

$$dZ_t = \sqrt{2Z_t}\, dW_t \tag{26.30}$$

with initial value $Z_0 = 1$. There we also showed that Z is dual to $Y^y_t = \left(\frac{t\gamma}{2} + \frac{1}{y}\right)^{-1}$ with $H(x, y) = e^{-xy}$. This implies uniqueness of the solution of (26.30) and the strong Markov property of Z. ◊

It could be objected that in Examples 26.29 and 26.31, we considered only one-dimensional problems for which the Yamada–Watanabe theorem (Theorem 26.10) yields uniqueness (indeed of a strong solution) anyway. The full strength of the method of duality is displayed only in higher-dimensional problems. As an example, we consider an extension of Example 26.29.

Example 26.32 (Interacting Wright–Fisher diffusions) The Wright–Fisher diffusion from Example 26.29 describes the fluctuations of the gene frequency of an allele in *one* large population. Now we consider more populations, which live at the points $i \in S := \{1, \dots, N\}$ and interact with each other by a migration that is quantified by migration rates $r(i, j) \geq 0$. As a model for the gene frequencies $X_t(i)$ at site

i at time t we use the following N-dimensional SDE for $X = (X(1), \ldots, X(N))$:

$$dX_t(i) = \sqrt{\gamma X_t(i)\big(1 - X_t(i)\big)}\, dW_t^i + \sum_{j=1}^{N} r(i,j)\big(X_t(j) - X_t(i)\big)\, dt. \qquad (26.31)$$

Here $W = (W^1, \ldots, W^N)$ is an N-dimensional Brownian motion. By Theorem 26.22, this SDE has weak solutions; however, none of our general criteria for weak uniqueness apply. We will thus show weak uniqueness by virtue of duality.

As in Example 26.29, it is not hard to show that solutions of (26.31), started at $X_0 = x \in E := [0,1]^S$, remain in $[0,1]^S$. The diagonal terms $r(i,i)$ do not appear in (26.31). We use our freedom and define these terms as $r(i,i) = -\sum_{j \neq i} r(i,j)$. Let $Y = (Y_t)_{t \geq 0}$ be the Markov process on $E' := (\mathbb{N}_0)^S$ with the following Q-matrix:

$$q(\varphi,\eta) = \begin{cases} \varphi(i) r(i,j), & \text{if } \eta = \varphi - \mathbb{1}_{\{i\}} + \mathbb{1}_{\{j\}} \text{ for} \\ & \text{some } i, j \in S, i \neq j, \\ \gamma \binom{\varphi(i)}{2}, & \text{if } \eta = \varphi - \mathbb{1}_{\{i\}} \text{ for some } i \in S, \\ \sum_{i \in S} \big(\varphi(i) r(i,i) - \gamma \binom{\varphi(i)}{2}\big), & \text{if } \eta = \varphi, \\ 0, & \text{else.} \end{cases}$$

Here $\varphi \in E'$ denotes a generic state with $\varphi(i)$ particles at site $i \in S$, and $\mathbb{1}_{\{i\}} \in E'$ denotes the state with exactly one particle at site i. The process Y describes a system of particles that independently with rate $r(i,j)$ jump from site i to site j. If there is more than one particle at the same site i, then any of the $\binom{\varphi(i)}{2}$ pairs of particles coalesce with the same rate γ to one particle. The common genealogical interpretation of this process is that (in reversed time) it describes the lines of descent of samples of $Y_0(i)$ individuals at each site $i \in S$. By migration, the lines change sites. If two individuals have the same common ancestor, then their lines coalesce. Clearly, for two particles to have the same ancestor at a given time, it is necessary but not sufficient for them to be at the same site.

For $x \in \mathbb{R}^n$ and $\varphi \in E'$, we denote $x^\varphi := \prod_{i \in S} x(i)^{\varphi(i)}$. We show that X and Y are dual to each other with the duality function $H(x,\varphi) = x^\varphi$:

$$\mathbf{E}_x\big[X_t^\varphi\big] = \mathbf{E}_\varphi\big[x^{Y_t}\big] \quad \text{for all } \varphi \in S^{\mathbb{N}_0}, x \in [0,1]^S, t \geq 0. \qquad (26.32)$$

Let $m^{x,\varphi}(t) := \mathbf{E}_x[X_t^\varphi]$ and $g^{x,\varphi}(t) := \mathbf{E}_\varphi[x^{Y_t}]$. Clearly, H has the derivatives $\partial_i H(\cdot,\varphi)(x) = \varphi(i) x^{\varphi - \mathbb{1}_{\{i\}}}$ and $\partial_i \partial_i H(\cdot,\varphi)(x) = 2\binom{\varphi(i)}{2} x^{\varphi - 2\mathbb{1}_{\{i\}}}$.

By the Itô formula,

$$X_t^\varphi - X_0^\varphi - \int_0^t \sum_{i,j \in S} \varphi(i) r(i,j)\big(X_s(j) - X_s(i)\big) X_t^{\varphi - \mathbb{1}_{\{i\}}}\, ds$$
$$- \sum_{i \in S} \int_0^t \gamma \binom{\varphi(i)}{2} \big(X_s(i)\big(1 - X_s(i)\big)\big) X_s^{\varphi - 2\mathbb{1}_{\{i\}}}\, ds$$

is a martingale. Taking expectations, we get a system of linear integral equations

$$
\begin{aligned}
m^{x,0}(t) &= 1, \\
m^{x,\varphi}(t) &= x^{\varphi} + \int_0^t \sum_{i,j\in S} \varphi(i) r(i,j)\big(m^{x,\varphi+\mathbb{1}_{\{j\}}-\mathbb{1}_{\{i\}}}(s) - m^{x,\varphi}(s)\big)\,ds \\
&\quad + \int_0^t \gamma \sum_{i\in S} \binom{\varphi(i)}{2} \big(m^{x,\varphi-\mathbb{1}_{\{i\}}}(s) - m^{x,\varphi}(s)\big)\,ds.
\end{aligned}
\tag{26.33}
$$

This system of equations can be solved uniquely by induction on $n = \sum_{i\in I} \varphi(i)$. However, we do not intend to compute this solution explicitly. We show only that it coincides with $g^{x,\varphi}(t)$ by showing that g solves an equivalent system of differential equations.

For g as in (26.29), we obtain

$$
\begin{aligned}
\frac{d}{dt} g^{x,\varphi}(t) &= \sum_{\eta\in E'} q(\varphi,\eta) g^{x,\varphi}(t) \\
&= \sum_{i,j\in S} r(i,j)\big(g^{x,\varphi+\mathbb{1}_{\{j\}}-\mathbb{1}_{\{i\}}}(t) - g^{x,\varphi}(t)\big) \\
&\quad + \sum_{i\in S} \gamma \binom{\varphi(i)}{2} \big(g^{x,\varphi-\mathbb{1}_{\{i\}}}(t) - g^{x,\varphi}(t)\big).
\end{aligned}
\tag{26.34}
$$

Together with the initial values $g^{x,0}(t) = 1$ and $g^{x,\varphi}(0) = x^{\varphi}$, the system (26.34) of differential equations is equivalent to (26.33). Hence the duality (26.32) holds, and thus the SDE (26.31) has a unique weak solution. (In fact, it can be shown that there exists a unique strong solution, even if S is countably infinite, as long as r then satisfies certain regularity conditions such as if it is the Q-matrix of a random walk on $S = \mathbb{Z}^d$; see [153].) ◊

Exercise 26.3.1 (Extinction probability of Feller's branching diffusion) Let $\gamma > 0$ and let Z be the solution of $dZ_t := \sqrt{\gamma Z_t}\,dW_t$ with initial value $Z_0 = z > 0$. Use the duality to show

$$
\mathbf{P}_z[Z_t = 0] = \exp\left(-\frac{2z}{\gamma t}\right). \tag{26.35}
$$

Use Lemma 21.44 to compute the probability that a Galton–Watson branching process X with critical geometric offspring distribution and with $X_0 = N \in \mathbb{N}$ is extinct by time $n \in \mathbb{N}$. Compare the result with (26.35).