Chapter 6

APPLICATIONS

This chapter is devoted to some applications and extensions of the theory developed earlier.

6.1. STOPPING TIMES

6.1.1. Definitions, basic properties. Let $(\Omega, \mathcal{U}, P)$ be a probability space and $\mathcal{F}(\cdot)$ a filtration of σ-algebras, as in Chapters 4 and 5. We introduce now some random times that are well-behaved with respect to $\mathcal{F}(\cdot)$:

DEFINITION. A random variable $\tau : \Omega \to [0, \infty]$ is called a *stopping time* with respect to $\mathcal{F}(\cdot)$ provided

$$\{\tau \leq t\} \in \mathcal{F}(t) \quad \text{for all } t \geq 0. \tag{1}$$

This says that the set of all $\omega \in \Omega$ such that $\tau(\omega) \leq t$ is an $\mathcal{F}(t)$-measurable set. Note that τ is allowed to take on the value $+\infty$ and also that any constant $\tau \equiv t_0$ is a stopping time.

THEOREM (Properties of stopping times). *Let τ_1 and τ_2 be stopping times with respect to $\mathcal{F}(\cdot)$. Then*

(i) *$\{\tau < t\} \in \mathcal{F}(t)$, and so $\{\tau = t\} \in \mathcal{F}(t)$, for all times $t \geq 0$.*

(ii) *$\tau_1 \wedge \tau_2 := \min(\tau_1, \tau_2)$ and $\tau_1 \vee \tau_2 := \max(\tau_1, \tau_2)$ are stopping times.*

Proof. Observe that

$$\{\tau < t\} = \bigcup_{k=1}^{\infty} \underbrace{\{\tau \leq t - 1/k\}}_{\in \mathcal{F}(t-1/k) \subseteq \mathcal{F}(t)}.$$

Also, we have $\{\tau_1 \wedge \tau_2 \le t\} = \{\tau_1 \le t\} \cup \{\tau_2 \le t\} \in \mathcal{F}(t)$, and furthermore $\{\tau_1 \vee \tau_2 \le t\} = \{\tau_1 \le t\} \cap \{\tau_2 \le t\} \in \mathcal{F}(t)$. □

The notion of stopping times comes up naturally in the study of stochastic differential equations, as it allows us to investigate phenomena occurring over "random time intervals". An example will make this clearer:

EXAMPLE (Hitting a set). Consider the solution $\mathbf{X}(\cdot)$ of the SDE

$$\begin{cases} d\mathbf{X}(t) = \mathbf{b}(t,\mathbf{X})dt + \mathbf{B}(t,\mathbf{X})d\mathbf{W} \\ \mathbf{X}(0) = \mathbf{X}_0, \end{cases}$$

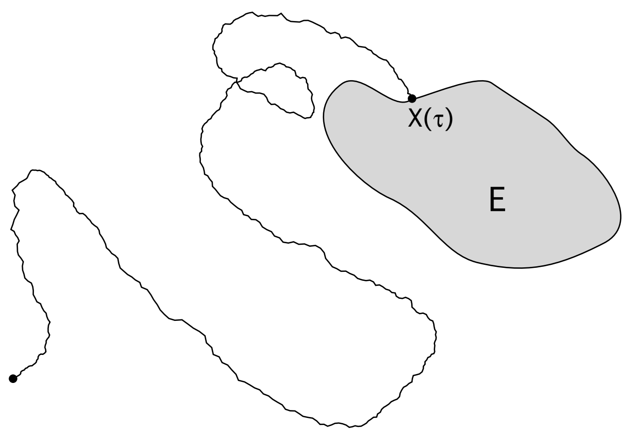
where $\mathbf{b}$, $\mathbf{B}$ and $\mathbf{X}_0$ satisfy the hypotheses of the Existence and Uniqueness Theorem.

THEOREM. *Let E be either a nonempty closed subset or a nonempty open subset of $\mathbb{R}^n$. Then*

$$\tau := \inf\{t \ge 0 \,|\, \mathbf{X}(t) \in E\}$$

is a stopping time.

(We put $\tau = +\infty$ for those sample paths of $\mathbf{X}(\cdot)$ that never hit E.)



Proof. Fix $t \ge 0$; we must show that $\{\tau \le t\} \in \mathcal{F}(t)$. Take $\{t_i\}_{i=1}^\infty$ to be a countable dense subset of $[0,\infty)$. First we assume that $E = U$ is an open set. Then the event

$$\{\tau \le t\} = \bigcup_{t_i \le t} \underbrace{\{\mathbf{X}(t_i) \in U\}}_{\in \mathcal{F}(t_i) \subseteq \mathcal{F}(t)}$$

belongs to $\mathcal{F}(t)$.

Next we assume that $E = C$ is a closed set. Set $d(x,C) := \operatorname{dist}(x,C)$ and define the open sets

$$U_n = \{x : d(x,C) < \frac{1}{n}\}.$$

The event

$$\{\tau \le t\} = \bigcap_{n=1}^{\infty} \bigcup_{t_i \le t} \underbrace{\{\mathbf{X}(t_i) \in U_n\}}_{\in \mathcal{F}(t_i) \subseteq \mathcal{F}(t)}$$

also belongs to $\mathcal{F}(t)$. □

REMARKS. (i) The random variable

$$\sigma := \sup\{t \ge 0 \,|\, \mathbf{X}(t) \in E\},$$

the last time that $\mathbf{X}(t)$ hits E, is *not* a stopping time. The reason is that the event $\{\sigma \le t\}$ depends upon the entire future history of the process and thus is not in general $\mathcal{F}(t)$-measurable. (In applications $\mathcal{F}(t)$ "contains the history of $\mathbf{X}(\cdot)$ up to and including time t but does not contain information about the future".)

(ii) The name "stopping time" comes from the example, where we sometimes think of halting the sample path $\mathbf{X}(\cdot)$ at the first time τ that it hits a set E. But there are many examples where we do not really stop the process at time τ. Thus "stopping time" is not a particularly good name. □

6.1.2. Stochastic integrals and stopping times. Our next task is to consider stochastic integrals with random limits of integration and to work out an Itô formula for these.

DEFINITION. If $G \in \mathbb{L}^2(0,T)$ and τ is a stopping time with $0 \le \tau \le T$, we define

$$\int_0^\tau G\, dW := \int_0^T \chi_{\{t \le \tau\}} G\, dW.$$

LEMMA. *If $G \in \mathbb{L}^2(0,T)$ and $0 \le \tau \le T$ is a stopping time, then*

(i)
$$E\left(\int_0^\tau G\, dW\right) = 0,$$

(ii)
$$E\left(\left(\int_0^\tau G\, dW\right)^2\right) = E\left(\int_0^\tau G^2\, dt\right).$$

Proof. We have

$$E\left(\int_0^\tau G\, dW\right) = E\left(\int_0^T \underbrace{\chi_{\{t \le \tau\}} G}_{\in \mathbb{L}^2(0,T)}\, dW\right) = 0$$

and

$$E\left(\left(\int_0^\tau G\,dW\right)^2\right) = E\left(\left(\int_0^T \chi_{\{t\le\tau\}}G\,dW\right)^2\right)$$
$$= E\left(\int_0^T (\chi_{\{t\le\tau\}}G)^2\,dt\right)$$
$$= E\left(\int_0^\tau G^2\,dt\right). \qquad \square$$

Similar formulas hold for vector-valued processes.

6.1.3. Itô's chain rule with stopping times. As usual, let $\mathbf{W}(\cdot)$ denote m-dimensional Brownian motion. Recall next from Chapter 4 that if

$$d\mathbf{X} = \mathbf{b}(\mathbf{X},t)dt + \mathbf{B}(\mathbf{X},t)d\mathbf{W},$$

then for each C^2 function u,

$$du(\mathbf{X},t) = u_t dt + \sum_{i=1}^n u_{x_i} d\mathbf{X}^i + \frac{1}{2}\sum_{i,j=1}^n u_{x_i x_j}\sum_{k=1}^m b^{ik}b^{jk}dt. \tag{2}$$

DEFINITION. The *generator* L associated with the process X is the partial differential operator

$$Lu := \sum_{i,j=1}^n a^{ij}u_{x_i x_j} + \sum_{i=1}^n b^i u_{x_i}, \tag{3}$$

where

$$a^{ij} = \frac{1}{2}\sum_{k=1}^m b^{ik}b^{jk}. \tag{4}$$

We can write the integral form of Itô's chain rule in terms of the generator as

$$u(\mathbf{X}(t),t) - u(\mathbf{X}(0),0) = \int_0^t u_t + Lu\,ds + \int_0^t Du\cdot\mathbf{B}\,d\mathbf{W}, \tag{5}$$

for

$$Du\cdot\mathbf{B}\,d\mathbf{W} = \sum_{k=1}^m\sum_{i=1}^n u_{x_i}b^{ik}dW^k.$$

The argument of u in these integrals is $(\mathbf{X}(s),s)$.

REMARK. The generator L is a (possibly degenerate) *elliptic* second-order linear partial differential operator. To see this, observe that if $\xi =$

$(\xi_1, \dots, \xi_n)$, then

$$\sum_{i,j=1}^{n} a^{ij}\xi_i\xi_j = \frac{1}{2}\sum_{i,j=1}^{n}\sum_{k=1}^{m} b^{ik}b^{jk}\xi_i\xi_j = \frac{1}{2}\sum_{k=1}^{m}\left(\sum_{i=1}^{n} b^{ik}\xi_i\right)^2 \geq 0.$$

We call L *uniformly elliptic* if there exists a constant $\theta > 0$ such that

$$\sum_{i,j=1}^{n} a^{ij}\xi_i\xi_j \geq \theta|\xi|^2$$

for all $\xi \in \mathbb{R}^n$. Depending upon the noise term $\mathbf{B}$, the generator L may or may not be uniformly elliptic. In the deterministic case that $\mathbf{B} \equiv 0$, L is a first-order partial differential operator. □

WARNING ABOUT NOTATION. The above expression for the generator L differs from the notation for elliptic operators in my PDE book [**E**]. Here there is no minus sign in front of the second-order term, as there is in [**E**, Chapter 6]. □

For a fixed $\omega \in \Omega$, formula (5) holds for all $0 \leq t \leq T$. Thus we may set $t = \tau$, where τ is a stopping time, $0 \leq \tau \leq T$:

$$u(\mathbf{X}(\tau), \tau) - u(\mathbf{X}(0), 0) = \int_0^\tau u_t + Lu\, ds + \int_0^\tau Du \cdot \mathbf{B}\, d\mathbf{W}.$$

Take the expected value:

$$E(u(\mathbf{X}(\tau), \tau)) - E(u(\mathbf{X}(0), 0)) = E\left(\int_0^\tau u_t + Lu\, ds\right). \tag{6}$$

We will see in the next section that this formula provides a very important link between stochastic differential equations and (nonrandom) partial differential equations.

6.1.4. Brownian motion and the Laplacian. The most important case is $\mathbf{X}(\cdot) = \mathbf{W}(\cdot)$, n-dimensional Brownian motion, the generator of which is

$$Lu = \frac{1}{2}\sum_{i=1}^{n} u_{x_i x_i} =: \frac{1}{2}\Delta u.$$

The expression Δu is called the *Laplacian* of u and occurs throughout mathematics and physics; see for instance [**E**, Section 2.2].

We demonstrate in the next section some important links with Brownian motion.

6.2. APPLICATIONS TO PDE, FEYNMAN–KAC FORMULA

6.2.1. Probabilistic formulas for solutions of PDE.

EXAMPLE 1 (Expected hitting time to a boundary). Let $U \subset \mathbb{R}^n$ be a bounded open set, with smooth boundary ∂U. According to standard PDE theory, there exists a smooth solution u of the equation

$$\begin{cases} -\frac{1}{2}\Delta u = 1 & \text{in } U \\ u = 0 & \text{on } \partial U. \end{cases} \tag{7}$$

Our goal is to find a probabilistic representation formula for u. For this, fix any point $x \in U$ and then consider an n-dimensional Brownian motion $\mathbf{W}(\cdot)$. Then $\mathbf{X}(\cdot) := \mathbf{W}(\cdot) + x$ represents a "Brownian motion starting at x". Define

$$\tau_x := \text{the first time } \mathbf{X}(\cdot) \text{ hits } \partial U.$$

THEOREM. *We have*

$$u(x) = E(\tau_x) \tag{8}$$

for each point $x \in U$.

Proof. We employ formula (6), with $Lu = \frac{1}{2}\Delta u$. We have for each $n = 1, 2, \ldots$

$$E(u(\mathbf{X}(\tau_x \wedge n))) - E(u(\mathbf{X}(0))) = E\left(\int_0^{\tau_x \wedge n} \frac{1}{2}\Delta u(\mathbf{X})\, ds\right).$$

Since $\frac{1}{2}\Delta u = -1$ according to (7) and since u is bounded,

$$\lim_{n\to\infty} E(\tau_x \wedge n) < \infty.$$

Thus τ_x is integrable. So if we let $n \to \infty$ above, we get

$$u(x) - E(u(\mathbf{X}(\tau_x))) = E\left(\int_0^{\tau_x} 1\, ds\right) = E(\tau_x).$$

But $u = 0$ on ∂U, and consequently $u(\mathbf{X}(\tau_x)) \equiv 0$. Formula (8) follows. □

REMARK. Since u is bounded on U, we see that

$$E(\tau_x) < \infty; \text{ and so } \tau_x < \infty \text{ a.s., for all } x \in U.$$

Therefore *Brownian sample paths starting at any point $x \in U$ will with probability 1 eventually hit ∂U.* □

EXAMPLE 2 (Probabilistic representation of harmonic functions). Let $U \subset \mathbb{R}^n$ be a smooth, bounded domain and $g : \partial U \to \mathbb{R}$ a given continuous function. It is known from classical PDE theory that there exists a function $u \in C^2(U) \cap C(\bar{U})$ satisfying the boundary-value problem:

$$\begin{cases} \Delta u = 0 & \text{in } U \\ u = g & \text{on } \partial U. \end{cases} \tag{9}$$

We call u a *harmonic* function and the PDE $\Delta u = 0$ is *Laplace's equation.*

THEOREM. *We have for each point $x \in U$*

$$u(x) = E(g(\mathbf{X}(\tau_x))), \tag{10}$$

for

$$\mathbf{X}(\cdot) := \mathbf{W}(\cdot) + x, \tag{11}$$

Brownian motion starting at x.

Proof. As shown above,

$$E(u(\mathbf{X}(\tau_x))) = E(u(\mathbf{X}(0))) + E\left(\int_0^{\tau_x} \tfrac{1}{2}\Delta u(\mathbf{X})\, ds\right) = E(u(\mathbf{X}(0))) = u(x),$$

the second equality valid since $\Delta u = 0$ in U. Since $u = g$ on ∂U, formula (10) follows. □

INTERPRETATION ("Random characteristics"). As explained in [**E**, Section 3.2] solutions of even strongly nonlinear first-order PDE can be constructed locally by solving an appropriate system of ODE, called the *characteristic equations*. But this is not possible for second-order PDE, such as Laplace's equation.

However we can understand the formula (10) as giving the solution of the boundary-value problem (9) for Laplace's equation in terms of "random characteristics", namely the trajectories of the shifted Brownian motion (11). In this interpretation we do not build the solution by integrating along a single characteristic curve, but rather we look at the ensemble of all the sample paths of a random process and then average. □

REMARK. In particular, if $\Delta u = 0$ in some open set containing the ball $B(x, r)$, then

$$u(x) = E(u(\mathbf{X}(\tau_x))),$$

where τ_x now denotes the hitting time of Brownian motion starting at x to $\partial B(x, r)$.

Since Brownian motion is isotropic in space, we may reasonably guess that the term on the right-hand side is just the average of u over the sphere

$\partial B(x,r)$, with respect to surface measure. That is, we have the identity

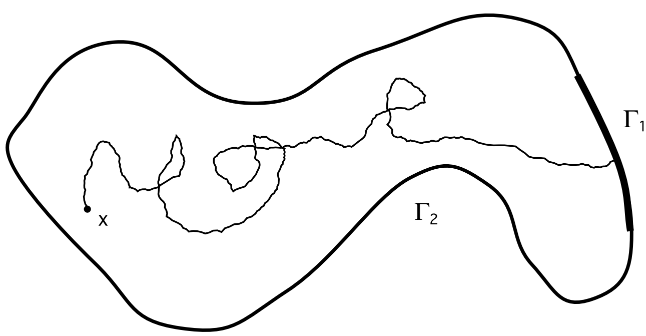
$$u(x) = \frac{1}{\text{area of } \partial B(x,r)} \int_{\partial B(x,r)} u \, dS.$$

This is the *mean-value formula* for harmonic functions. See [**E**, Section 2.2.2] for a nonprobabilistic derivation. □

EXAMPLE 3 (Hitting one part of a boundary first). Assume next that we can write ∂U as the union of two disjoint parts Γ_1, Γ_2. Let u solve the PDE

$$\begin{cases} \Delta u = 0 & \text{in } U \\ u = 1 & \text{on } \Gamma_1 \\ u = 0 & \text{on } \Gamma_2. \end{cases} \tag{12}$$

THEOREM. *For each point $x \in U$, $u(x)$ is the probability that a Brownian motion starting at x hits Γ_1 before hitting Γ_2:*



Hitting one part of the boundary first

Proof. Apply (10) for

$$g = \begin{cases} 1 & \text{on } \Gamma_1 \\ 0 & \text{on } \Gamma_2. \end{cases}$$

Then

$$u(x) = E(g(\mathbf{X}(\tau_x))) = \text{probability of hitting } \Gamma_1 \text{ before } \Gamma_2.$$ □

6.2.2. Feynman–Kac formula. Now we extend Example 2 above to obtain a probabilistic representation for the unique solution of the PDE

$$\begin{cases} -\frac{1}{2}\Delta u + cu = f & \text{in } U \\ u = 0 & \text{on } \partial U. \end{cases} \tag{13}$$

We assume c, f are smooth functions, with $c \geq 0$ in U.

THEOREM (Feynman–Kac formula). *For each* $x \in U$,

$$u(x) = E\left(\int_0^{\tau_x} f(\mathbf{X}(t)) e^{-\int_0^t ic(\mathbf{X})\,ds} dt \right) \tag{14}$$

where, as before,

$$\mathbf{X}(\cdot) := \mathbf{W}(\cdot) + x$$

is a Brownian motion starting at x *and* τ_x *denotes the first hitting time of* ∂U.

Proof. We know that $E(\tau_x) < \infty$. Since $c \geq 0$, the integral above converges.

First look at the process

$$Y(t) := e^{Z(t)},$$

for $Z(t) := -\int_0^t c(\mathbf{X})\,ds$. Then

$$dZ = -c(\mathbf{X})dt,$$

and so Itô's chain rule yields

$$dY = -c(\mathbf{X})Y dt.$$

Hence the Itô product rule implies

$$\begin{aligned} d\left(u(\mathbf{X})e^{-\int_0^t c(\mathbf{X})\,ds}\right) &= (du(\mathbf{X}))e^{-\int_0^t c(\mathbf{X})\,ds} \\ &\quad + u(\mathbf{X})d\left(e^{-\int_0^t c(\mathbf{X})ds}\right) \\ &= \left(\tfrac{1}{2}\Delta u(\mathbf{X})dt + \textstyle\sum_{i=1}^n u_{x_i}(\mathbf{X})dW^i\right) e^{-\int_0^t c(\mathbf{X})ds} \\ &\quad + u(\mathbf{X})(-c(\mathbf{X})dt)e^{-\int_0^t c(\mathbf{X})\,ds}. \end{aligned}$$

We use formula (6) with $\tau = \tau_x$ and take the expected value, obtaining

$$\begin{aligned} E\left(u(\mathbf{X}(\tau_x))e^{-\int_0^{\tau_x} c(\mathbf{X})\,ds}\right) &- E(u(\mathbf{X}(0))) \\ &= E\left(\int_0^{\tau_x} \left[\tfrac{1}{2}\Delta u(\mathbf{X}) - c(\mathbf{X})u(\mathbf{X})\right] e^{-\int_0^t c(\mathbf{X})\,ds} dt\right). \end{aligned}$$

Since u solves the PDE (13), this simplifies to give

$$u(x) = E(u(\mathbf{X}(0))) = E\left(\int_0^{\tau_x} f(\mathbf{X})e^{-\int_0^t ic(\mathbf{X})\,ds} dt\right),$$

as claimed. □

INTERPRETATION. We can explain this formula as describing a Brownian motion with "killing", as follows.

Suppose that the Brownian particles may disappear at a random killing time, for example by being absorbed into the medium within which it is moving. Assume further that the probability of its being killed in a short time interval $[t, t+h]$ is

$$c(\mathbf{X}(t))h + o(h).$$

Then the probability of the particle surviving until time t is approximately equal to

$$(1 - c(\mathbf{X}(t_1))h)(1 - c(\mathbf{X}(t_2))h) \cdots (1 - c(\mathbf{X}(t_n))h),$$

where $0 = t_0 < t_1 < \cdots < t_n = t$, $h = t_{k+1} - t_k$. As $h \to 0$, this converges to $e^{-\int_0^t c(\mathbf{X})\,ds}$.

The Feynman–Kac formula says that $u(x)$ equals the average of $f(\mathbf{X}(\cdot))$ integrated along sample paths, weighted by their survival probabilities. □

REMARK. If we consider in these examples the solution of the SDE

$$\begin{cases} d\mathbf{X} = \mathbf{b}(\mathbf{X})dt + \mathbf{B}(\mathbf{X})d\mathbf{W} \\ \mathbf{X}(0) = x, \end{cases}$$

we can obtain similar formulas, where now

$$\tau_x = \text{the hitting time of } \partial U \text{ for } \mathbf{X}(\cdot)$$

and $\frac{1}{2}\Delta u$ is replaced by the generator

$$Lu := \sum_{i,j=1}^{n} a^{ij} u_{x_i x_j} + \sum_{i=1}^{n} b^i u_{x_i}, \quad a^{ij} = \frac{1}{2}\sum_{k=1}^{m} b^{ik} b^{jk}.$$

Note, however, that we need to know that the various PDE have smooth solutions. This need not always be the case for degenerate elliptic operators L. □

6.3. OPTIMAL STOPPING

The general mathematical setting for many control theory problems is this. We are given some "system" whose state evolves in time according to a differential equation (deterministic or stochastic). Given also are certain *controls* which somehow affect the behavior of the system: these controls typically either modify some parameters in the dynamics or else stop the process, or both. Finally we are given a *cost criterion*, depending upon our choice of control and the corresponding state of the system.

> *The goal is to discover an optimal choice of controls, to minimize the cost criterion.*

The easiest stochastic control problem of the general type outlined above occurs when we cannot directly affect the SDE controlling the evolution of $\mathbf{X}(\cdot)$ and can only decide at each instance whether or not to stop. An important problem of this type follows.

6.3.1. Stopping a stochastic differential equation. Let $U \subset \mathbb{R}^m$ be a bounded, smooth domain. Suppose $\mathbf{b} : \mathbb{R}^n \to \mathbb{R}^n$, $\mathbf{B} : \mathbb{R}^n \to M^{n\times m}$ satisfy the usual assumptions.

Then for each $x \in U$ the stochastic differential equation

$$\begin{cases} d\mathbf{X} = \mathbf{b}(\mathbf{X})dt + \mathbf{B}(\mathbf{X})d\mathbf{W} \\ X_0 = x \end{cases}$$

has a unique solution. Let $\tau = \tau_x$ denote the hitting time of ∂U. Let θ be any stopping time with respect to $\mathcal{F}(\cdot)$, and for each such θ define the *expected cost* of stopping $\mathbf{X}(\cdot)$ at time $\theta \wedge \tau$ to be

$$J_x(\theta) := E\left(\int_0^{\theta\wedge\tau} f(\mathbf{X}(s))\,ds + g(\mathbf{X}(\theta\wedge\tau))\right). \tag{15}$$

The idea is that if we stop at the possibly random time $\theta < \tau$, then the cost is a given function g of the current state of $\mathbf{X}(\theta)$. If instead we do not stop the process before it hits ∂U, that is, if $\theta \geq \tau$, the cost is $g(\mathbf{X}(\tau))$. In addition there is a running cost per unit time f of keeping the system in operation until time $\theta \wedge \tau$.

6.3.2. Optimal stopping. The main question is this: does there exist an optimal stopping time $\theta^* = \theta^*_x$ for which

$$J_x(\theta^*) = \min_{\substack{\theta \text{ stopping} \\ \text{time}}} J_x(\theta)?$$

And if so, how can we find θ^*? It turns out to be very difficult to try to design θ^* directly. A much better idea is to turn our attention to the *value function*

$$u(x) := \inf_\theta J_x(\theta) \tag{16}$$

and to try to determine u as a function of $x \in U$. Note that $u(x)$ is the minimum expected cost, given that we start the process at x. It turns out that once we know u, we will then be able to construct an optimal θ^*. This approach is called *dynamic programming.*

Optimality conditions. So assume u is defined as above and suppose u is smooth enough to justify the following calculations. We wish to determine the properties of this function.

First of all, notice that we can take $\theta \equiv 0$ in the definition (16). That is, we could just stop immediately and incur the cost $g(\mathbf{X}(0)) = g(x)$. Hence

$$u(x) \leq g(x) \quad \text{for each point } x \in U. \tag{17}$$

Furthermore, $\tau \equiv 0$ if $x \in \partial U$, and so

$$u(x) = g(x) \quad \text{for each point } x \in \partial U. \tag{18}$$

Next take any point $x \in U$ and fix some small number $\delta > 0$. Now if we do not stop the system for time δ, then according to (SDE) the new state of the system at time δ will be $\mathbf{X}(\delta)$. Then, given that we are at the point $\mathbf{X}(\delta)$, the best we can achieve in minimizing the cost thereafter must be

$$u(\mathbf{X}(\delta)).$$

So if we choose not to stop the system for time δ, then, assuming we do not hit ∂U, our cost is at least

$$E\left(\int_0^\delta f(\mathbf{X})\,ds + u(\mathbf{X}(\delta))\right).$$

Since $u(x)$ is the infimum of costs over all stopping times, we therefore have

$$u(x) \leq E\left(\int_0^\delta f(\mathbf{X})\,ds + u(\mathbf{X}(\delta))\right).$$

Now by Itô's chain rule

$$E(u(\mathbf{X}(\delta))) = u(x) + E\left(\int_0^\delta Lu(\mathbf{X})\,ds\right),$$

for

$$Lu = \sum_{i,j=1}^n a^{ij} u_{x_i x_j} + \sum_{i=1}^n b^i u_{x_i}, \quad a^{ij} = \frac{1}{2}\sum_{k=1}^m b^{ik} b^{jk}.$$

Hence

$$0 \leq E\left(\int_0^\delta f(\mathbf{X}) + Lu(\mathbf{X})\,ds\right).$$

Divide by $\delta > 0$, and then let $\delta \to 0$:

$$0 \leq f(x) + Lu(x).$$

Equivalently, we have

$$Mu \leq f \quad \text{in } U, \tag{19}$$

where

$$Mu := -Lu.$$

Finally we observe that if in (19) a strict inequality holds, that is, if

$$u(x) < g(x) \text{ at some point } x \in U,$$

then it is optimal not to stop the process at once. Thus it is plausible to think that we should leave the system going, for at least some very small time δ. In this circumstance we then would have an equality in the formula above, and so

$$Mu = f \quad \text{at those points where } u < g. \tag{20}$$

In summary, we combine (17)–(20) to find that *if* the formal reasoning above is valid, then the value function u satisfies

$$\begin{cases} \max(Mu - f, u - g) = 0 & \text{in } U \\ u = g & \text{on } \partial U. \end{cases} \tag{21}$$

This is called an *obstacle problem* (introduced in a different context in [**E**, Section 8.4.2]).

6.3.3. Solving for the value function. Our rigorous study of the stopping time problem now begins by first showing that there exists a unique solution u of (21) and second that this u is in fact $\min_\theta J_x(\theta)$. Then we will use u to design θ^*, an optimal stopping time.

THEOREM. *Suppose f, g are given smooth functions. There exists a unique function u, with bounded second derivatives, such that*

(i) $u \leq g$ *in* U,

(ii) $Mu \leq f$ *almost everywhere in* U,

(iii) $\max(Mu - f, u - g) = 0$ *almost everywhere in* U,

(iv) $u = g$ *on* ∂U.

In general $u \notin C^2(U)$.

The idea of the proof is to approximate the obstacle problem (21) by a *penalized problem* of this form:

$$\begin{cases} Mu^\varepsilon + \beta_\varepsilon(u^\varepsilon - g) = f & \text{in } U \\ u^\varepsilon = g & \text{on } \partial U, \end{cases}$$

where $\beta_\varepsilon : \mathbb{R} \to \mathbb{R}$ is a smooth, convex function, $\beta'_\varepsilon \geq 0$, and $\beta_\varepsilon \equiv 0$ for $x \leq 0$, $\lim_{\epsilon \to 0} \beta_\varepsilon(x) = \infty$ for $x > 0$. Then $u^\varepsilon \to u$. In practice it will be difficult to find a precise formula for u, but computers can provide accurate numerical approximations.

6.3.4. Designing an optimal stopping policy. Now we show that our solution of (21) is in fact the value function, and along the way we will learn how to design an optimal stopping strategy θ^*.

First note that the *stopping set*

$$S := \{x \in U \mid u(x) = g(x)\}$$

is closed. For each $x \in \bar{U}$, define

$$\theta^* = \text{the first hitting time of } S.$$

THEOREM. *Let u be the solution of* (21)*. Then*

$$u(x) = J_x(\theta^*) = \inf_{\theta} J_x(\theta) \tag{22}$$

for all $x \in \bar{U}$.

This says that we should first compute the solution to (21) to find the stopping set S, define θ^* as above, and then we should run $\mathbf{X}(\cdot)$ until it hits S (or else exits from U).

Proof. 1. Define the *continuation set*

$$C := U - S = \{x \in U \mid u(x) < g(x)\}.$$

On this set $Lu = f$, and furthermore $u = g$ on ∂C. Since $\tau \wedge \theta^*$ is the exit time from C, we have for $x \in C$

$$u(x) = E\left(\int_0^{\tau\wedge\theta^*} f(\mathbf{X}(s))\,ds + g(\mathbf{X}(\theta^* \wedge \tau))\right) = J_x(\theta^*).$$

On the other hand, if $x \in S$, $\tau \wedge \theta^* = 0$, and so

$$u(x) = g(x) = J_x(\theta^*).$$

Thus for all $x \in \bar{U}$, we have $u(x) = J_x(\theta^*)$.

2. Now let θ be any other stopping time. We need to show that

$$u(x) = J_x(\theta^*) \leq J_x(\theta).$$

Now by Itô's chain rule

$$u(x) = E\left(\int_0^{\tau\wedge\theta} Mu(\mathbf{X})\,ds + u(\mathbf{X}(\tau \wedge \theta))\right).$$

But $Mu \leq f$ and $u \leq g$ in $\bar{U}$. Hence

$$u(x) \leq E\left(\int_0^{\tau\wedge\theta} f(\mathbf{X})\,ds + g(\mathbf{X}(\tau \wedge \theta))\right) = J_x(\theta).$$

But since $u(x) = J_x(\theta^*)$, we consequently have

$$u(x) = J_x(\theta^*) = \min_{\theta} J_x(\theta),$$

as asserted. □

6.4. OPTIONS PRICING

In this section we outline an application of the Itô stochastic calculus to mathematical finance.

6.4.1. The basic problem. Let us consider a given *security*, say a stock, whose price at time t is $S(t)$. We suppose that S evolves according to the SDE discussed in Chapter 5:

$$\begin{cases} dS = \mu S dt + \sigma S dW \\ S(0) = s_0, \end{cases} \tag{23}$$

where $\mu > 0$ is the drift and $\sigma \neq 0$ is the volatility. The initial price s_0 is known.

A *derivative* is a financial instrument whose payoff depends upon (i.e., is derived from) the behavior of $S(\cdot)$. We will investigate a *European call option*, which is the right to buy one share of the stock S, at the price p at time T. The number p is called the *strike price* and $T > 0$ the *strike* (or *expiration*) *time*. The basic question is this:

What is the "proper price" at time $t = 0$ of this option?

In other words, if you run a financial firm and wish to sell your customers this call option, how much should you charge? (We are looking for the "break-even" price, the price for which the firm neither makes nor loses money.)

6.4.2. Arbitrage and hedging. To simplify, we assume hereafter that the prevailing, no-risk interest rate is the constant $r > 0$. This means that \$1 put in a bank at time $t = 0$ becomes \$$e^{rT}$ at time $t = T$. Equivalently, \$1 at time $t = T$ is worth only \$$e^{-rT}$ at time $t = 0$.

As for the problem of pricing our call option, a first guess might be that the proper price should be

$$e^{-rT} E((S(T) - p)^+), \tag{24}$$

for $x^+ := \max(x, 0)$. The reasoning behind this guess is that if $S(T) < p$, then the option is worthless. If $S(T) > p$, we can buy a share for the price p, immediately sell at price $S(T)$, and thereby make a profit of $(S(T) - p)^+$. We average this over all sample paths and multiply by the discount factor e^{-rT}, to arrive at (24).

As reasonable as this may all seem, (24) is in fact *not* the proper price. Other forces are at work in financial markets. Indeed the fundamental factor in options pricing is *arbitrage*, meaning the possibility of risk-free profits.

> *We must price our option so as to create no arbitrage opportunities for others.*

To convert this principle into mathematics, we introduce also the notion of *hedging*. This means somehow eliminating our risk as the seller of the call option. The exact details appear below, but the basic idea is that we can in effect "duplicate" our option by a portfolio consisting of (continually changing) holdings of a risk-free bond and of the stock on which the call is written.

6.4.3. A partial differential equation. We next demonstrate how to use these principles to convert our pricing problem into a PDE. We introduce for $s \geq 0$ and $0 \leq t \leq T$, the unknown *price function*

$$u(s,t) = \text{the proper price of the option at time } t, \text{ given } S(t) = s. \tag{25}$$

Then $u(s_0, 0)$ is the price we are seeking.

Terminal and boundary conditions. We want to calculate u as a function of s and t. For this, notice first that at the expiration time T, we have

$$u(s,T) = (s-p)^+ \qquad (s \geq 0). \tag{26}$$

Furthermore, if $s = 0$, then $S(t) = 0$ for all $0 \leq t \leq T$, and so

$$u(0,t) = 0 \qquad (0 \leq t \leq T). \tag{27}$$

Duplicating an option, self-financing. We now seek a PDE that u solves for $s > 0$, $0 \leq t \leq T$.

To do so, define the process

$$C(t) := u(S(t),t) \qquad (0 \leq t \leq T). \tag{28}$$

Thus $C(t)$ is the current price of the option at time t and is random since the stock price $S(t)$ is random. According to Itô's chain rule and (23),

$$\begin{aligned} dC &= u_t dt + u_s dS + \frac{1}{2} u_{ss} (dS)^2 \\ &= \left(u_t + \mu S u_s + \frac{\sigma^2}{2} S^2 u_{ss} \right) dt + \sigma S u_s dW. \end{aligned} \tag{29}$$

Now comes the key idea: we propose to "duplicate" C by a portfolio consisting of shares of S and of a *bond* B. More precisely, assume that B is a risk-free investment, which therefore grows at the prevailing interest rate r:

$$\begin{cases} dB = rBdt \\ B(0) = 1. \end{cases} \tag{30}$$

This just means $B(t) = e^{rt}$, of course. We will try to find processes ϕ and ψ so that

$$C = \phi S + \psi B \qquad (0 \le t \le T). \tag{31}$$

Discussion. The point is that if we can construct ϕ, ψ so that (31) holds, we can eliminate all risk. To see this more clearly, imagine that your financial firm sells a call option, as above. The firm thereby incurs the risk that at time T, the stock price $S(T)$ will exceed p, and so the buyer will exercise the option. But if in the meantime the firm has constructed the portfolio (31), the profits from it will exactly equal the funds needed to pay the customer. Conversely, if the option is worthless at time T, the portfolio will have no profit.

But to make this work, the financial firm should not have to inject any new money into the hedging scheme, beyond the initial investment to set it up. We ensure this by requiring that the portfolio represented on the right-hand side of (31) be *self-financing*. This means that the changes in the value of the portfolio should depend only upon the changes in S, B. We therefore require that

$$dC = \phi dS + \psi dB \qquad (0 \le t \le T). \tag{32}$$

Combining formulas (29), (30) and (32) provides the identity

$$\begin{aligned} &\left(u_t + \mu S u_s + \frac{\sigma^2}{2} S^2 u_{ss}\right) dt + \sigma S u_s dW \\ &\qquad = \phi(\mu S dt + \sigma S dW) + \psi r B dt. \end{aligned} \tag{33}$$

So if (31) holds, (33) must be valid, and we are trying to select ϕ, ψ to make all of this so. We observe in particular that the terms multiplying dW on each side of (33) will match provided we take

$$\phi(t) := u_s(S(t), t) \qquad (0 \le t \le T). \tag{34}$$

Then (33) simplifies, to read

$$\left(u_t + \frac{\sigma^2}{2} S^2 u_{ss}\right) dt = r\psi B dt.$$

But $\psi B = C - \phi S = u - u_s S$, according to (32), (34). Hence

$$\left(u_t + rSu_s + \frac{\sigma^2}{2} S^2 u_{ss} - ru\right) dt = 0.$$

The argument of u and its partial derivatives is $(S(t), t)$.

Consequently, to make sure that (31) is valid, we ask that the function $u = u(s,t)$ solve the *Black–Scholes–Merton* equation

$$u_t + rsu_s + \frac{\sigma^2}{2}s^2u_{ss} - ru = 0 \qquad (0 \le t \le T). \tag{35}$$

The main outcome of our financial reasoning is the derivation of this partial differential equation. Observe that the parameter μ does not appear.

More on self-financing. Before going on, we return to the self-financing condition (32). The Itô product rule and (31) imply

$$dC = \phi dS + \psi dB + Sd\phi + Bd\psi + d\phi\, dS.$$

To have (32), we consequently must make sure that

$$Sd\phi + Bd\psi + d\phi\, dS = 0, \tag{36}$$

where we recall that $\phi = u_s(S(t),t)$. Now

$$d\phi\, dS = \sigma^2 S^2 u_{ss} dt.$$

Thus (36) is valid provided

$$d\psi = -B^{-1}\left(Sd\phi + \sigma^2 S^2 u_{ss} dt\right). \tag{37}$$

We can confirm this by noting that (31), (34) imply

$$\psi = B^{-1}(C - \phi S) = e^{-rt}\,(u(S,t) - u_s(S,t)S).$$

We can now check (37) by a direct calculation.

Summary. To price our call option, we solve the boundary/initial-value problem

$$\begin{cases} u_t + rsu_s + \frac{\sigma^2}{2}s^2u_{ss} - ru = 0 & (s > 0, 0 \le t \le T) \\ u = (s-p)^+ & (s > 0,\, t = T) \\ u = 0 & (s = 0, 0 \le t \le T) \end{cases} \tag{38}$$

for the function $u = u(s,t)$. Then $u(s_0, 0)$ is the option price we have been trying to find.

It turns out that this problem can be solved explicitly, although we omit the details here; see Baxter–Rennie [**B-R**].

6.5. THE STRATONOVICH INTEGRAL

We next discuss the *Stratonovich stochastic calculus*, which is an alternative to Itô's. See Arnold [**A**, Chapter 10] for more discussion and for proofs omitted here.

6.5.1. Motivation. Let us consider first of all the *formal* random differential equation

$$\begin{cases} \dot{X} = d(t)X + f(t)X\xi \\ X(0) = X_0, \end{cases} \tag{39}$$

where $m = n = 1$ and $\xi(\cdot)$ is one-dimensional "white noise". If we interpret this rigorously as the stochastic differential equation

$$\begin{cases} dX = d(t)Xdt + f(t)XdW \\ X(0) = X_0, \end{cases} \tag{40}$$

we then recall from Chapter 5 that the unique solution is

$$X(t) = X_0 e^{\int_0^t d(s) - \frac{1}{2}f^2(s)\,ds + \int_0^t f(s)\,dW}. \tag{41}$$

On the other hand, perhaps (39) is a proposed mathematical model of some physical process and we are not really sure whether $\xi(\cdot)$ is "really" white noise. It could perhaps be instead some process with smooth (but highly complicated) sample paths. How would this possibility change the solution?

6.5.2. Approximating white noise. More precisely, suppose $\{\xi^k(\cdot)\}_{k=1}^\infty$ is a sequence of stochastic processes satisfying

(a) $E(\xi^k(t)) = 0$,
(b) $E(\xi^k(t)\xi^k(s)) =: d^k(t-s)$,
(c) $\xi^k(t)$ is Gaussian for all $t \geq 0$,
(d) $t \mapsto \xi^k(t)$ is smooth for all ω,

where we suppose that the functions $d^k(\cdot)$ converge as $k \to \infty$ to δ_0, the Dirac point mass at 0.

In light of the formal interpretation of white noise $\xi(\cdot)$ as a Gaussian process with $E\xi(t) = 0$, $E(\xi(t)\xi(s)) = \delta_0(t-s)$ (see page 44), the $\xi^k(\cdot)$ are thus presumably smooth approximations of $\xi(\cdot)$.

6.5.3. Limits of solutions. Now consider the problem

$$\begin{cases} \dot{X}^k = d(t)X^k + f(t)X^k\xi^k \\ X^k(0) = X_0. \end{cases} \tag{42}$$

For each ω this is just a regular ODE, whose solution is

$$X^k(t) := X_0 e^{\int_0^t d(s)\,ds + \int_0^t f(s)\xi^k(s)\,ds}.$$

Next look at

$$Z^k(t) := \int_0^t f(s)\xi^k(s)\,ds.$$

For each time $t \geq 0$, this is a Gaussian random variable, with

$$E(Z^k(t)) = 0.$$

Furthermore,

$$\begin{aligned} E(Z^k(t)Z^k(s)) &= \int_0^t \int_0^s f(\tau)f(\sigma)d_k(\tau - \sigma)\,d\sigma d\tau \\ &\to \int_0^t \int_0^s f(\tau)f(\sigma)\delta_0(\tau - \sigma)\,d\sigma d\tau \\ &= \int_0^{t\wedge s} f^2(\tau)\,d\tau. \end{aligned}$$

Hence as $k \to \infty$, $Z^k(t)$ converges in L^2 to a process whose distributions agree with those of $\int_0^t f(s)\,dW$. And therefore $X^k(t)$ converges to a process whose distributions agree with

$$\hat{X}(t) := X_0 e^{\int_0^t d(s)\,ds + \int_0^t f(s)\,dW}. \tag{43}$$

This is not the solution (41) *we found earlier*!

INTERPRETATION. Thus if we regard (39) as an Itô SDE with $\xi(\cdot)$ a "true" white noise, (41) is our solution. But if we approximate $\xi(\cdot)$ by smooth processes $\xi^k(\cdot)$, solve the approximate problems (42), and pass to limits with the approximate solutions $X^k(\cdot)$, we get the different solution (43). This means that (39) is *unstable* with respect to changes in the random term $\xi(\cdot)$. This conclusion has important consequences in questions of modeling, since it may be unclear experimentally whether we really have $\xi(\cdot)$ or instead $\xi^k(\cdot)$ in (39) and similar problems. □

In view of all this, it is appropriate to ask if there is some way to redefine the stochastic integral so that these difficulties do not come up. One answer is the *Stratonovich integral*.

6.5.4. Definition of the Stratonovich integral.

EXAMPLE. Recall that in Chapter 4 we defined for one-dimensional Brownian motion

$$\int_0^T W\,dW := \lim_{|P^n|\to 0} \sum_{k=0}^{m_n-1} W(t_k^n)(W(t_{k+1}^n) - W(t_k^n))$$
$$= \frac{W^2(T) - T}{2},$$

where $P^n := \{0 = t_0^n < t_1^n < \cdots < t_{m_n}^n = T\}$ is a partition of $[0,T]$. This corresponds to a sequence of Riemann sum approximations, where the integrand is evaluated at the *left-hand endpoint* of each subinterval $[t_k^n, t_{k+1}^n]$.

The corresponding *Stratonovich integral* is instead defined this way:

$$\int_0^T W \circ dW := \lim_{|P^n|\to 0} \sum_{k=0}^{m_n-1} \left(\tfrac{W(t_{k+1}^n)+W(t_k^n)}{2}\right)(W(t_{k+1}^n) - W(t_k^n))$$
$$= \frac{W^2(T)}{2}.$$

According to calculations on page 63, we also have

$$\int_0^T W \circ dW = \lim_{|P^n|\to 0} \sum_{k=0}^{m_n-1} W\left(\tfrac{t_{k+1}^n+t_k^n}{2}\right)(W(t_{k+1}^n) - W(t_k^n)).$$

Therefore for this case the Stratonovich integral corresponds to a Riemann sum approximation where we evaluate the integrand at the *midpoint* of each subinterval $[t_k^n, t_{k+1}^n]$. □

NOTATION. Observe that we hereafter write a small circle $\circ$ before the dW to signify the Stratonovich integral.

We generalize this example and so introduce the

DEFINITION. Let $\mathbf{W}(\cdot)$ be an n-dimensional Brownian motion and let $\mathbf{B} : \mathbb{R}^n \times [0,T] \to \mathbb{M}^{n\times n}$ be a C^1 function such that

$$E\left(\int_0^T |\mathbf{B}(\mathbf{W},t)|^2\,dt\right) < \infty.$$

Then we define

$$\int_0^T \mathbf{B}(\mathbf{W},t)\circ d\mathbf{W} := \lim_{|P^n|\to 0} \sum_{k=0}^{m_n-1} \mathbf{B}\left(\tfrac{\mathbf{W}(t_{k+1}^n)+\mathbf{W}(t_k^n)}{2}, t_k^n\right)(\mathbf{W}(t_{k+1}^n) - \mathbf{W}(t_k^n)).$$

It can be shown that this limit exists in $L^2(\Omega)$.

REMARK. Remember that Itô's integral can be computed this way:

$$\int_0^T \mathbf{B}(\mathbf{W},t)\,d\mathbf{W} = \lim_{|P^n|\to 0} \sum_{k=0}^{m_n-1} \mathbf{B}(\mathbf{W}(t_k^n),t_k^n)(\mathbf{W}(t_{k+1}^n)-\mathbf{W}(t_k^n)).$$

This is in general not equal to the Stratonovich integral, but there is a *conversion formula*

$$\left[\int_0^T \mathbf{B}(\mathbf{W},t)\circ d\mathbf{W}\right]^i = \left[\int_0^T \mathbf{B}(\mathbf{W},t)\,d\mathbf{W}\right]^i + \frac{1}{2}\int_0^T \sum_{j=1}^n b_{x_j}^{ij}(\mathbf{W},t)\,dt \tag{44}$$

for $i = 1,\dots,n$. Here v^i means the i-th component of the vector function $\mathbf{v}$. This formula is proved by noting that

$$\int_0^T \mathbf{B}(\mathbf{W},t)\circ d\mathbf{W} - \int_0^T \mathbf{B}(\mathbf{W},t)\,d\mathbf{W} = \lim_{|P^n|\to 0}\sum_{k=0}^{m_n-1}\left[\mathbf{B}\left(\tfrac{\mathbf{W}(t_{k+1}^n)+\mathbf{W}(t_k^n)}{2},t_k^n\right)-\mathbf{B}(\mathbf{W}(t_k^n),t_k^n)\right]\cdot(\mathbf{W}(t_{k+1}^n)-\mathbf{W}(t_k^n))$$

and using the Mean Value Theorem plus some usual methods for evaluating the limit. We omit the details.

If $n = 1$, the conversion formula reads

$$\int_0^T b(W,t)\circ dW = \int_0^T b(W,t)\,dW + \frac{1}{2}\int_0^T b_x(W,t)\,dt. \tag{45}$$ □

Assume now that $\mathbf{B} : \mathbb{R}^n \times [0,T] \to \mathbb{M}^{n\times m}$ and that $\mathbf{W}(\cdot)$ is an m-dimensional Brownian motion. We make this informal

DEFINITION. If $\mathbf{X}(\cdot)$ is a stochastic process with values in $\mathbb{R}^n$, we define

$$\int_0^T \mathbf{B}(\mathbf{X},t)\circ d\mathbf{W} := \lim_{|P^n|\to 0}\sum_{k=0}^{m_n-1}\mathbf{B}\left(\tfrac{\mathbf{X}(t_{k+1}^n)+\mathbf{X}(t_k^n)}{2},t_k^n\right)(\mathbf{W}(t_{k+1}^n)-\mathbf{W}(t_k^n))$$

provided this limit exists in $L^2(\Omega)$ for all sequences of partitions P^n, with $|P^n|\to 0$.

6.5.5. Stratonovich chain rule.

DEFINITION. Suppose that the process $\mathbf{X}(\cdot)$ solves the Stratonovich integral equation

$$\mathbf{X}(t) = \mathbf{X}(0) + \int_0^t \mathbf{b}(\mathbf{X},s)\,ds + \int_0^t \mathbf{B}(\mathbf{X},s)\circ d\mathbf{W} \quad (0\le t\le T)$$

for $\mathbf{b} : \mathbb{R}^n \times [0,T] \to \mathbb{R}^n$ and $\mathbf{B} : \mathbb{R}^n \times [0,T] \to \mathbb{M}^{n\times m}$. We then write

$$d\mathbf{X} = \mathbf{b}(\mathbf{X},t)dt + \mathbf{B}(\mathbf{X},t) \circ d\mathbf{W}, \tag{46}$$

the second term on the right being the *Stratonovich stochastic differential.*

THEOREM (Stratonovich chain rule). *Assume*

$$d\mathbf{X} = \mathbf{b}(\mathbf{X},t)dt + \mathbf{B}(\mathbf{X},t) \circ d\mathbf{W}$$

and suppose $u : \mathbb{R}^n \times [0,T] \to \mathbb{R}$ is smooth.

Then

$$\begin{aligned} du(\mathbf{X},t) &= u_t dt + \sum_{i=1}^n u_{x_i} \circ d\mathbf{X}^i \\ &= \left(u_t + \sum_{i=1}^n u_{x_i} b^i \right) dt + \sum_{i=1}^n \sum_{k=1}^m u_{x_i} b^{ik} \circ d\mathbf{W}^k. \end{aligned} \tag{47}$$

Thus the ordinary chain rule holds for Stratonovich stochastic differentials, and there is no additional term involving $u_{x_i x_j}$ as there is for Itô's chain rule.

We omit the proof, which is similar to that for the Itô rule. The main difference is that we make use of the formula $\int_0^T W \circ dW = \frac{1}{2}W^2(T)$ in the approximations.

REMARK. Next let us return to the motivational example we began with. We have seen that if the differential equation (39) is interpreted to mean

$$\begin{cases} dX = d(t)X dt + f(t)X dW & \text{(Itô's sense)} \\ X(0) = X_0, \end{cases}$$

then

$$X(t) = X_0 e^{\int_0^t d(s) - \frac{1}{2} f^2(s)\, ds + \int_0^t f(s)\, dW}.$$

However, if we understand (39) to mean

$$\begin{cases} dX = d(t)X dt + f(t)X \circ dW & \text{(Stratonovich's sense)} \\ X(0) = X_0, \end{cases}$$

the solution is

$$\tilde{X}(t) = X_0 e^{\int_0^t d(s)\, ds + \int_0^t f(s)\, dW},$$

as is easily checked using the Stratonovich calculus described above.

This solution $\tilde{X}(\cdot)$ is also the solution obtained by approximating the "white noise" $\xi(\cdot)$ by smooth processes $\xi^k(\cdot)$ and passing to limits. This suggests that interpreting (39) and similar formal random differential equations in the Stratonovich sense will provide solutions which are stable with respect to perturbations in the random terms. This is indeed the case; see the articles [**S1**], [**S2**] by Sussmann.

Note also that these considerations clarify a bit the problems of interpreting mathematically the *formal* random differential equation (39) but do not say which interpretation is physically correct. This is a question of modeling and is not, strictly speaking, a mathematical issue. □

6.5.6. Conversion rules for SDE. Let $\mathbf{W}(\cdot)$ be an m-dimensional Brownian motion and suppose $\mathbf{b} : \mathbb{R}^n \times [0,T] \to \mathbb{R}^n$, $\mathbf{B} : \mathbb{R}^n \times [0,T] \to \mathbb{M}^{n\times m}$ satisfy the hypotheses of the basic existence and uniqueness theorem. Then $\mathbf{X}(\cdot)$ solves the *Itô* stochastic differential equation

$$\begin{cases} d\mathbf{X} = \mathbf{b}(\mathbf{X},t)dt + \mathbf{B}(\mathbf{X},t)d\mathbf{W} \\ \mathbf{X}(0) = \mathbf{X}_0 \end{cases}$$

if and only if $\mathbf{X}(\cdot)$ solves the *Stratonovich* stochastic differential equation

$$\begin{cases} d\mathbf{X} = \left[\mathbf{b}(\mathbf{X},t) - \tfrac{1}{2}\mathbf{c}(\mathbf{X},t)\right] dt + \mathbf{B}(\mathbf{X},t) \circ d\mathbf{W} \\ \mathbf{X}(0) = \mathbf{X}_0, \end{cases}$$

for

$$c^i(x,t) := \sum_{k=1}^{m}\sum_{j=1}^{n} b^{ik}_{x_j}(x,t)b^{jk}(x,t) \qquad (i = 1,\dots,n).$$

For $m = n = 1$, this says

$$dX = b(X)dt + \sigma(X)dW$$

if and only if

$$dX = \left(b(X) - \frac{1}{2}\sigma'(X)\sigma(X)\right) dt + \sigma(X) \circ dW.$$

6.5.7. Summary. We conclude by summarizing the advantages of each definition of the stochastic integral:

- **Advantages of Itô integral**
  - (i) Simple formulas:

$$E\left(\int_0^t G\,dW\right) = 0, \quad E\left(\left(\int_0^t G\,dW\right)^2\right) = E\left(\int_0^t G^2\,dt\right).$$

  - (ii) $I(t) = \int_0^t G\,dW$ is a martingale.

- **Advantages of Stratonovich integral**
  - (i) Ordinary chain rule holds.
  - (ii) Solutions of stochastic differential equations interpreted in the Stratonovich sense are stable with respect to changes in random terms.

Exercises

1. Show, using the formal manipulations for Itô's chain rule discussed in Chapter 1, that

$$Y(t) := e^{W(t)-\frac{t}{2}}$$

solves the stochastic differential equation

$$\begin{cases} dY = Y dW \\ Y(0) = 1. \end{cases}$$

(Hint: If $X(t) := W(t) - \frac{t}{2}$, then $dX = -\frac{dt}{2} + dW$.)

2. Show that

$$S(t) = s_0 e^{\sigma W(t) + \left(\mu - \frac{\sigma^2}{2}\right)t}$$

solves

$$\begin{cases} dS = \mu S dt + \sigma S dW \\ S(0) = s_0. \end{cases}$$

3. (i) Let $(\Omega, \mathcal{U}, P)$ be a probability space and let $A_1 \subseteq A_2 \subseteq \cdots \subseteq A_n \subseteq \ldots$ be events. Show that

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{m\to\infty} P(A_m).$$

(Hint: Look at the disjoint events $B_n := A_{n+1} - A_n$.)

(ii) Likewise, show that if $A_1 \supseteq A_2 \supseteq \cdots \supseteq A_n \supseteq \cdots$, then

$$P\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{m\to\infty} P(A_m).$$

4. Let Ω be any set and $\mathcal{A}$ any collection of subsets of Ω. Show that there exists a unique smallest σ-algebra $\mathcal{U}$ of subsets of Ω containing $\mathcal{A}$. We call $\mathcal{U}$ the σ-algebra *generated* by $\mathcal{A}$.

   (Hint: Take the intersection of all the σ-algebras containing $\mathcal{A}$.)

5. Show that if $A_1, \dots, A_n$ are events, then

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{1 \le i < j \le n} P(A_i \cap A_j) + \sum_{1 \le i < j < k \le n} P(A_i \cap A_j \cap A_k) - \cdots + (-1)^{n-1} P(A_1 \cap A_2 \cap \cdots \cap A_n).$$

   (Hint: Do the case $n = 2$ first and then the general case by induction.)

6. Let $X = \sum_{i=1}^{k} a_i \chi_{A_i}$ be a simple random variable, where the real numbers a_i are distinct, the events A_i are pairwise disjoint, and $\Omega = \bigcup_{i=1}^{k} A_i$. Let $\mathcal{U}(X)$ be the σ-algebra generated by X.

   (i) Describe precisely which sets are in $\mathcal{U}(X)$.

   (ii) Suppose the random variable Y is $\mathcal{U}(X)$-measurable. Show that Y is constant on each set A_i.

   (iii) Show that therefore Y can be written as a function of X.

7. Verify:

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}, \qquad \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x e^{-\frac{(x-m)^2}{2\sigma^2}} dx = m,$$

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (x-m)^2 e^{-\frac{(x-m)^2}{2\sigma^2}} dx = \sigma^2.$$

8. Suppose A and B are independent events in some probability space. Show that A^c and B are independent. Likewise, show that A^c and B^c are independent.

9. Suppose we have three cards: one is red on both sides, one is red on one side and white on the other side, and one is white on both sides.

   (i) Pick a card and then one of its sides at random. What is the probability it is red?

   (ii) Given that the side of the card is red, what is the probability that the other side is red?

10. Suppose that $A_1, A_2, \dots, A_m$ are disjoint events, each of positive probability, such that $\Omega = \bigcup_{j=1}^m A_j$. Prove *Bayes' formula*:

$$P(A_k \mid B) = \frac{P(B \mid A_k)P(A_k)}{\sum_{j=1}^m P(B \mid A_j)P(A_j)}$$

for $k = 1, \dots, m$ provided $P(B) > 0$.

11. During one fall semester 105 women applied to Miskatonic University, of whom 76 were accepted, and 400 men applied, of whom 230 were accepted.

During the subsequent spring semester, 300 women applied, of whom 100 were accepted, and 112 men applied, of whom 21 were accepted.

Calculate numerically

(a) the probability of a female applicant being accepted during the fall,
(b) the probability of a male applicant being accepted during the fall,
(c) the probability of a female applicant being accepted during the spring,
(d) the probability of a male applicant being accepted during the spring.

Consider now the total applicant pool for both semesters together and calculate

(e) the probability of a female applicant being accepted,
(f) the probability of a male applicant being accepted.

Are the university's admission policies biased towards females or towards males?

12. Let X be a real-valued, $N(0,1)$ random variable, and set $Y := X^2$. Calculate the density g of the distribution function for Y.

(Hint: You must find g so that $P(-\infty < Y \leq a) = \int_{-\infty}^a g\,dy$ for all a.)

13. Take $\Omega = [0,1] \times [0,1]$, with $\mathcal{U}$ the Borel sets and P Lebesgue measure. Let $g : [0,1] \to \mathbb{R}$ be a continuous function.

Define the random variables

$$X_1(\omega) := g(x_1), \quad X_2(\omega) := g(x_2) \quad \text{for } \omega = (x_1, x_2) \in \Omega.$$

Show that X_1 and X_2 are independent and identically distributed.

14. Let $f : [0,1] \to \mathbb{R}$ be continuous and define the *Bernstein polynomial*

$$b_n(x) := \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k} x^k (1-x)^{n-k}.$$

Prove that $b_n \to f$ uniformly on $[0,1]$ as $n \to \infty$, by providing the details for the following steps.

(i) Since f is uniformly continuous, for each $\epsilon > 0$ there exists $\delta(\epsilon) > 0$ such that $|f(x) - f(y)| \leq \epsilon$ if $|x - y| \leq \delta(\epsilon)$.

(ii) Given $x \in [0,1]$, take a sequence of independent random variables X_k such that $P(X_k = 1) = x, P(X_k = 0) = 1 - x$. Write $S_n = X_1 + \cdots + X_n$. Then $b_n(x) = E(f(\frac{S_n}{n}))$.

(iii) Therefore

$$\begin{aligned}|b_n(x) - f(x)| &\leq E(|f(\tfrac{S_n}{n}) - f(x)|)\\ &= \int_A |f(\tfrac{S_n}{n}) - f(x)|\,dP + \textstyle\int_{A^c} |f(\tfrac{S_n}{n}) - f(x)|\,dP,\end{aligned}$$

for $A := \{\omega \in \Omega \mid |\frac{S_n}{n} - x| \leq \delta(\epsilon)\}$.

(iv) Then show that

$$|b_n(x) - f(x)| \leq \epsilon + \tfrac{2M}{\delta(\epsilon)^2} V(\tfrac{S_n}{n}) = \epsilon + \tfrac{2M}{n\delta(\epsilon)^2} V(X_1),$$

for $M = \max|f|$. Conclude that $b_n \to f$ uniformly.

15. Let X and Y be independent random variables, and suppose that f_X and f_Y are the density functions for X, Y. Show that the density function for $X + Y$ is

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y)\,dy.$$

(Hint: If $g : \mathbb{R} \to \mathbb{R}$, we have

$$E(g(X+Y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y) g(x+y)\,dxdy,$$

where $f_{X,Y}$ is the joint density function of X, Y.)

16. Let X and Y be two independent positive random variables, each with density

$$f(x) = \begin{cases} e^{-x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0. \end{cases}$$

Find the density of $X + Y$.

17. Show that
$$\lim_{n\to\infty}\int_0^1\int_0^1\cdots\int_0^1 f(\tfrac{x_1+\cdots+x_n}{n})\,dx_1dx_2\ldots dx_n = f(\tfrac{1}{2})$$
for each continuous function f.

(Hint: Let $X_1,\ldots,X_n,\ldots$ be independent random variables, each of which has density function $f_i(x)=1$ if $0\le x\le 1$ and $=0$ otherwise. Then $P(|\frac{X_1+\cdots+X_n}{n}-\frac{1}{2}|>\epsilon)\le\frac{1}{\epsilon^2}V(\frac{X_1+\cdots+X_n}{n})=\frac{1}{12\epsilon^2 n}$.)

18. Prove that

(i) $E(E(X\,|\,\mathcal{V}))=E(X)$,

(ii) $E(X)=E(X\,|\,\mathcal{W})$, where $\mathcal{W}=\{\emptyset,\Omega\}$ is the trivial σ-algebra.

19. Let X,Y be two real-valued random variables and suppose their joint distribution function has the density $f(x,y)$. Show that
$$E(X|Y)=\Phi(Y)\quad\text{a.s.}$$
for
$$\Phi(y)=\frac{\int_{-\infty}^{\infty}xf(x,y)\,dx}{\int_{-\infty}^{\infty}f(x,y)\,dx}.$$

(Hints: $\Phi(Y)$ is a function of Y and so it is $\mathcal{U}(Y)$-measurable. Therefore we must show that

$$(*)\qquad \int_A X\,dP=\int_A\Phi(Y)\,dP\quad\text{for all } A\in\mathcal{U}(Y).$$

Now $A=Y^{-1}(B)$ for some Borel subset of $\mathbb{R}$. So the left-hand side of $(*)$ is

$$(**)\qquad \int_A X\,dP=\int_\Omega\chi_B(Y)X\,dP=\int_{-\infty}^{\infty}\int_B xf(x,y)\,dydx.$$

The right-hand side of $(*)$ is
$$\int_A\Phi(Y)\,dP=\int_{-\infty}^{\infty}\int_B\Phi(y)f(x,y)\,dydx,$$
which equals the right-hand side of $(**)$. Fill in the details.)

20. A smooth function $\Phi:\mathbb{R}\to\mathbb{R}$ is called *convex* if $\Phi''(x)\ge 0$ for all $x\in\mathbb{R}$.

(i) Show that if Φ is convex, then
$$\Phi(y)\ge\Phi(x)+\Phi'(x)(y-x)\quad\text{for all } x,y\in\mathbb{R}.$$

(ii) Show that
$$\Phi\left(\frac{x+y}{2}\right)\le\frac{1}{2}\Phi(x)+\frac{1}{2}\Phi(y)\quad\text{for all } x,y\in\mathbb{R}.$$

(iii) A smooth function $\Phi : \mathbb{R}^n \to \mathbb{R}$ is called *convex* if the matrix $D^2\Phi = ((\Phi_{x_i x_j}))$ is nonnegative definite for all $x \in \mathbb{R}^n$. (This means that $\sum_{i,j=1}^n \Phi_{x_i x_j}\xi_i\xi_j \geq 0$ for all $\xi \in \mathbb{R}^n$.) Prove that

$$\Phi(y) \geq \Phi(x) + D\Phi(x) \cdot (y - x),$$
$$\Phi\left(\frac{x+y}{2}\right) \leq \frac{1}{2}\Phi(x) + \frac{1}{2}\Phi(y)$$

for all $x, y \in \mathbb{R}^n$. (Here "D" denotes the gradient.)

21. (i) Prove *Jensen's inequality*:

$$\Phi(E(X)) \leq E(\Phi(X))$$

for a random variable $X : \Omega \to \mathbb{R}$, where Φ is convex. (Hint: Use assertion (iii) from the previous exercise.)

(ii) Prove the *conditional Jensen inequality*:

$$\Phi(E(X|\mathcal{V})) \leq E(\Phi(X)|\mathcal{V}).$$

22. Let $W(\cdot)$ be a one-dimensional Brownian motion. Show that

$$E(W^{2k}(t)) = \frac{(2k)!t^k}{2^k k!} \quad (t > 0).$$

23. Show that if $\mathbf{W}(\cdot)$ is an n-dimensional Brownian motion, then so are

(i) $\mathbf{W}(t+s) - \mathbf{W}(s)$ for all $s \geq 0$,
(ii) $c\mathbf{W}(t/c^2)$ for all $c > 0$ ("Brownian scaling").

24. Let $W(\cdot)$ be a one-dimensional Brownian motion, and define

$$\bar{W}(t) := \begin{cases} tW(\frac{1}{t}) & \text{for } t > 0 \\ 0 & \text{for } t = 0. \end{cases}$$

Show that $\bar{W}(t) - \bar{W}(s)$ is $N(0, t-s)$ for times $0 \leq s \leq t$. ($\bar{W}(\cdot)$ also has independent increments and so is a one-dimensional Brownian motion. You do not need to show this.)

25. Define $X(t) := \int_0^t W(s)\,ds$, where $W(\cdot)$ is a one-dimensional Brownian motion. Show that

$$E(X^2(t)) = \frac{t^3}{3} \quad \text{for each } t > 0.$$

26. Define $X(t)$ as in the previous exercise. Show that

$$E(e^{\lambda X(t)}) = e^{\frac{\lambda^2 t^3}{6}} \quad \text{for each } t > 0.$$

(Hint: $X(t)$ is a Gaussian random variable, the variance of which we know from the previous exercise.)

27. Define $U(t) := e^{-t}W(e^{2t})$, where $W(\cdot)$ is a one-dimensional Brownian motion. Show that
$$E(U(t)U(s)) = e^{-|t-s|} \quad \text{for all } -\infty < s, t < \infty.$$

28. Let $W(\cdot)$ be a one-dimensional Brownian motion. Show that
$$\lim_{m\to\infty} \frac{W(m)}{m} = 0 \quad \text{almost surely.}$$
(Hint: Fix $\epsilon > 0$ and define the event $A_m := \{|\frac{W(m)}{m}| \geq \epsilon\}$. Then $A_m = \{|X| \geq \sqrt{m}\epsilon\}$ for the $N(0,1)$ random variable $X = \frac{W(m)}{\sqrt{m}}$. Apply the Borel–Cantelli Lemma.)

29. (i) Let $0 < \gamma \leq 1$. Show that if $f : [0,T] \to \mathbb{R}^n$ is uniformly Hölder continuous with exponent γ, it is also uniformly Hölder continuous with each exponent $0 < \delta < \gamma$.
(ii) Show that $f(t) = t^\gamma$ is uniformly Hölder continuous with exponent γ on the interval $[0,1]$.

30. Let $0 < \gamma < \frac{1}{2}$. We showed in Chapter 3 that if $W(\cdot)$ is a one-dimensional Brownian motion, then for almost every ω there exists a constant K, *depending on* ω, such that
$$(*) \qquad |W(t,\omega) - W(s,\omega)| \leq K|t-s|^\gamma \quad \text{for all } 0 \leq s, t \leq 1.$$
Show that there does not exist a constant K such that $(*)$ holds for almost all ω.

31. Prove that if $G, H \in \mathbb{L}^2(0,T)$, then
$$E\left(\int_0^T G\,dW \int_0^T H\,dW\right) = E\left(\int_0^T GH\,dt\right).$$
(Hint: $2ab = (a+b)^2 - a^2 - b^2$.)

32. Let $(\Omega, \mathcal{U}, P)$ be a probability space, and take $\mathcal{F}(\cdot)$ to be a filtration of σ-algebras. Assume X to be an integrable random variable, and define $X(t) := E(X|\mathcal{F}(t))$ for times $t \geq 0$.
Show that $X(\cdot)$ is a martingale.

33. Show directly that $I(t) := W^2(t) - t$ is a martingale.
(Hint: $W^2(t) = (W(t) - W(s))^2 - W^2(s) + 2W(t)W(s)$. Take the conditional expectation with respect to $\mathcal{W}(s)$, the history of $W(\cdot)$, and then condition with respect to the history of $I(\cdot)$.)

34. Suppose $X(\cdot)$ is a real-valued martingale and $\Phi : \mathbb{R} \to \mathbb{R}$ is convex. Assume also that $E(|\Phi(X(t))|) < \infty$ for all $t \geq 0$. Show that
$$\Phi(X(\cdot)) \text{ is a submartingale.}$$
(Hint: Use the conditional Jensen inequality.)

35. Use the Itô chain rule to show that $Y(t) := e^{\frac{t}{2}} \cos(W(t))$ is a martingale.

36. Let $\mathbf{W}(\cdot) = (W^1, \ldots, W^n)$ be an n-dimensional Brownian motion, and write $Y(t) := |\mathbf{W}(t)|^2 - nt$ for times $t \geq 0$. Show that $Y(\cdot)$ is a martingale.
    (Hint: Compute dY.)

37. Show that
$$\int_0^T W^2\, dW = \frac{1}{3} W^3(T) - \int_0^T W\, dt$$
and
$$\int_0^T W^3\, dW = \frac{1}{4} W^4(T) - \frac{3}{2} \int_0^T W^2\, dt.$$

38. Recall from the text that
$$Y := e^{\int_0^t g\, dW - \frac{1}{2}\int_0^t g^2\, ds}$$
satisfies
$$dY = gY\, dW.$$
Use this to prove
$$E(e^{\int_0^T g\, dW}) = e^{\frac{1}{2}\int_0^T g^2\, ds}.$$

39. Let $u = u(x,t)$ be a smooth solution of the *backwards diffusion equation*
$$u_t + \frac{1}{2} u_{xx} = 0,$$
and suppose $W(\cdot)$ is a one-dimensional Brownian motion.
    Show that for each time $t > 0$
$$E(u(W(t), t)) = u(0, 0).$$

40. Calculate $E(B^2(t))$ for the Brownian bridge $B(\cdot)$, and show in particular that $E(B^2(t)) \to 0$ as $t \to 1^-$.

41. Let X solve the Langevin equation, and suppose that X_0 is an $N(0, \frac{\sigma^2}{2b})$ random variable. Show that
$$E(X(s)X(t)) = \frac{\sigma^2}{2b} e^{-b|t-s|}.$$

42. (i) Consider the ODE
$$\begin{cases} \dot{x} = x^2 \quad (t > 0) \\ x(0) = x_0. \end{cases}$$
Show that if $x_0 > 0$, the solution "blows up to infinity" in finite time.

(ii) Next, look at the ODE

$$\begin{cases} \dot{x} = x^{\frac{1}{2}} \quad (t > 0) \\ x(0) = 0. \end{cases}$$

Show that this problem has infinitely many nonnegative solutions.

(Hint: $x \equiv 0$ is a solution. Find also a solution which is positive for times $t > 0$, and then combine these solutions to find ones which are zero for some time and then become positive.)

43. (i) Use the substitution $X = u(W)$ to solve the SDE

$$\begin{cases} dX = -\frac{1}{2}e^{-2X}dt + e^{-X}dW \\ X(0) = x_0. \end{cases}$$

(ii) Show that the solution blows up at a finite, random time.

44. Solve the SDE $dX = -X\,dt + e^{-t}dW$.

45. Let $\mathbf{W} = (W^1, W^2, \ldots, W^n)$ be an n-dimensional Brownian motion and write

$$R := |\mathbf{W}| = \left(\sum_{i=1}^{n}(W^i)^2\right)^{\frac{1}{2}}.$$

Show that R solves the *stochastic Bessel equation*

$$dR = \frac{n-1}{2R}dt + \sum_{i=1}^{n}\frac{W^i}{R}dW^i.$$

46. (i) Show that $\mathbf{X} = (\cos(W), \sin(W))$ solves the system of SDE

$$\begin{cases} dX^1 = -\frac{1}{2}X^1dt - X^2dW \\ dX^2 = -\frac{1}{2}X^2dt + X^1dW. \end{cases}$$

(ii) Show also that if $\mathbf{X} = (X^1, X^2)$ is any other solution, then $|\mathbf{X}|$ is constant in time.

47. Solve the system

$$\begin{cases} dX^1 = dt + dW^1 \\ dX^2 = X^1dW^2, \end{cases}$$

where $\mathbf{W} = (W^1, W^2)$ is a Brownian motion.

48. Solve

$$\begin{cases} dX^1 = X^2dt + dW^1 \\ dX^2 = X^1dt + dW^2. \end{cases}$$

49. Solve

$$\begin{cases} dX = \frac{1}{2}\sigma'(X)\sigma(X)dt + \sigma(X)dW \\ X(0) = 0 \end{cases}$$

where W is a one-dimensional Brownian motion and σ is a smooth, positive function.

(Hint: Let $f(x) := \int_0^x \frac{dy}{\sigma(y)}$ and set $g := f^{-1}$, the inverse function of f. Show that $X = g(W)$.)

50. Let τ be the first time a one-dimensional Brownian motion hits the half-open interval $(a, b]$. Show that τ is a stopping time.

51. Let $\mathbf{W}$ denote an n-dimensional Brownian motion for $n \geq 3$. Write $\mathbf{X} = \mathbf{W} + x_0$, where the point x_0 lies in the region $U = \{0 < R_1 < |x| < R_2\}$. Calculate explicitly the probability that $\mathbf{X}$ will hit the outer sphere $\{|x| = R_2\}$ before hitting the inner sphere $\{|x| = R_1\}$.

(Hint: Check that

$$\Phi(x) = \frac{1}{|x|^{n-2}}$$

satisfies $\Delta\Phi = 0$ for $x \neq 0$. Modify Φ to build a function u which equals 0 on the inner sphere and 1 on the outer sphere.)