Chapter 13

Stochastic Optimal Control

In previous chapters we assumed that the state variables of the system were known with certainty. If this were not the case, the state of the system over time would be a stochastic process. We are then faced with a *stochastic optimal control problem* where the state of the system is represented by a controlled stochastic process. We shall only consider the case when the state equation is perturbed by a Wiener process, which gives rise to the state as a Markov diffusion process. In Appendix D.2 we have defined the Wiener process, also known as Brownian motion. In Section 13.1, we will formulate a stochastic optimal control problem governed by stochastic differential equations involving a Wiener process, known as Itô equations. Our goal will be to synthesize optimal feedback controls for systems subject to Itô equations in a way that maximizes the expected value of a given objective function.

In this chapter, we also assume that the state is (fully) observed. On the other hand, when the system is subject to noisy measurements, we face partially observed optimal control problems. In some important special cases, it is possible to separate the problem into two problems: optimal estimation and optimal control. We discuss one such case in Appendix D.4.1. In general, these problems are very difficult and are beyond the scope of this book. Interested readers can consult some references listed in Section 13.5.

In Section 13.2, we will extend the production planning model of Chapter 6 to allow for some uncertain disturbances. We will obtain an optimal production policy for the stochastic production planning problem thus formulated. In Section 13.3, we solve an optimal stochastic advertising problem explicitly. The problem is a modification as well as

a stochastic extension of the optimal control problem of the Vidale-Wolfe advertising model treated in Section 7.2.4. In Section 13.4, we will introduce investment decisions in the consumption model of Example 1.3. We will consider both risk-free and risky investments. Our goal will be to find optimal consumption and investment policies in order to maximize the discounted value of the utility of consumption over time.

In Section 13.5, we will conclude the chapter by mentioning other types of stochastic optimal control problems that arise in practice.

13.1 Stochastic Optimal Control

In Appendix D.1 on the Kalman filter, we obtain optimal state estimation for linear systems with noise and noisy measurements. In Section D.4.1, we see that for stochastic linear-quadratic optimal control problems, the separation principle allows us to solve the problem in two steps: to obtain the optimal estimate of the state and to use it in the optimal feedback control formula for deterministic linear-quadratic problems.

In this section we will introduce the possibility of controlling a system governed by Itô stochastic differential equations. In other words, we will introduce control variables into equation (D.20). This produces the formulation of a stochastic optimal control problem.

It should be noted that for such problems, the separation principle does not hold in general. Therefore, to simplify the treatment, it is often assumed that the state variables are observable, in the sense that they can be directly measured. Furthermore, most of the literature on these problems use dynamic programming or the Hamilton-Jacobi-Bellman framework rather than stochastic maximum principles. In what follows, therefore, we will formulate the stochastic optimal control problem under consideration, and provide a brief, informal development of the Hamilton-Jacobi-Bellman equation for the problem. A detailed analysis of the problem is available in Fleming and Rishel (1975). For problems involving jump disturbances, see Davis (1993) for the methodology of optimal control of piecewise deterministic processes. For stochastic optimal control in discrete time, see Bertsekas and Shreve (1996).

Let us consider the problem of maximizing

$$E\left[\int_0^T F(X_t, U_t, t)dt + S(X_T, T)\right], \tag{13.1}$$

where X_t is the state variable, U_t is the closed-loop control variable, Z_t is a standard Wiener process, and together they are required to satisfy

the Itô stochastic differential equation

$$dX_t = f(X_t, U_t, t)dt + G(X_t, U_t, t)dZ_t, \quad X_0 = x_0. \tag{13.2}$$

For convenience in exposition we assume the drift coefficient function $F : E^1 \times E^1 \times E^1 \to E^1$, $S : E^1 \times E^1 \to E^1$, $f : E^1 \times E^1 \times E^1 \to E^1$ and the diffusion coefficient function $G : E^1 \times E^1 \times E^1 \to E^1$, so that (13.2) is a scalar equation. We also assume that the functions F and S are continuous in their arguments and the functions f and G are continuously differentiable in their arguments. For multidimensional extensions of this problem, see Fleming and Rishel (1975).

Since (13.2) is a scalar equation, the subscript t here means only time t. Thus, writing X_t, in place of writing $X(t)$, will not cause any confusion and, at the same time, will eliminate the need of writing many parentheses. Thus, dZ_t in (13.2) is the same as $dZ(t)$ in (**??**), except that in (13.2), dZ_t is a scalar.

To solve the problem defined by (13.1) and (13.2), let $V(x, t)$, known as the *value function*, be the expected value of the objective function (13.1) from t to T, when an optimal policy is followed from t to T, given $X_t = x$. Then, by the principle of optimality,

$$V(x, t) = \max_u E[F(x, u, t)dt + V(x + dX_t, t + dt)]. \tag{13.3}$$

By Taylor's expansion, we have

$$\begin{aligned} V(x + dX_t, t + dt) = V(x, t) \quad & + V_t dt + V_x dX_t + \tfrac{1}{2} V_{xx}(dX_t)^2 \\ & + \tfrac{1}{2} V_{tt}(dt)^2 + \tfrac{1}{2} V_{xt} dX_t dt \\ & + \text{higher-order terms.} \end{aligned} \tag{13.4}$$

From (13.2), we can formally write

$$(dX_t)^2 = f^2(dt)^2 + G^2(dZ_t)^2 + 2fGdZ_t dt, \tag{13.5}$$

$$dX_t dt = f(dt)^2 + GdZ_t dt. \tag{13.6}$$

The exact meaning of these expressions comes from the theory of stochastic calculus; see Arnold (1974, Chapter 5), Durrett (1996) or Karatzas and Shreve (1997). For our purposes, it is sufficient to know the multiplication rules of the stochastic calculus:

$$(dZ_t)^2 = dt, \; dZ_t.dt = 0, \; dt^2 = 0. \tag{13.7}$$

Substitute (13.4) into (13.3) and use (13.5), (13.6), (13.7), and the property that $E[Z_t] = 0$ to obtain

$$V = \max_u E\left[Fdt + V + V_t dt + V_x f dt + \frac{1}{2}V_{xx}G^2 dt + o(dt)\right]. \quad (13.8)$$

Note that we have suppressed the arguments of the functions involved in (13.8).

Cancelling the the term V on both sides of (13.8), dividing the remainder by dt, and letting $dt \to 0$, we obtain the Hamilton-Jacobi-Bellman (HJB) equation

$$0 = \max_u [F + V_t + V_x f + \frac{1}{2}V_{xx}G^2] \quad (13.9)$$

for the value function $V(t, x)$ with the boundary condition

$$V(x, T) = S(x, T). \quad (13.10)$$

In the next section we will apply this theory of stochastic optimal control to a simple stochastic production inventory problem treated by Sethi and Thompson (1981).

13.2 A Stochastic Production Inventory Model

In Section 6.1.1, we had formulated a deterministic production-inventory model. In this section, we extend a simplified version of that model by including a random process. Let us define the following quantities:

$$\begin{aligned}
I_t &= \text{the inventory level at time } t \text{ (state variable)},\\
P_t &= \text{the production rate at time } t \text{ (control variable)},\\
S &= \text{the constant demand rate at time } t;\ S > 0,\\
T &= \text{the length of planning period},\\
I_0 &= \text{the initial inventory level},\\
B &= \text{the salvage value per unit of inventory at time } T,\\
Z_t &= \text{the standard Wiener process},\\
\sigma &= \text{the constant diffusion coefficient}.
\end{aligned}$$

The inventory process evolves according to the stock-flow equation stated as the Itô stochastic differential equation

$$dI_t = (P_t - S)dt + \sigma dZ_t, \quad I_0 \text{ given}, \quad (13.11)$$

where I_0 denotes the initial inventory level. As in (D.22) and (**??**), we note that the process dZ_t can be formally expressed as $w(t)dt$, where $w(t)$ is considered to be a white noise process; see Arnold (1974). It can be interpreted as "sales returns," "inventory spoilage," etc., which are random in nature.

The objective function is:

$$\max E\left\{\int_0^T -(P_t^2 + I_t^2)dt + BI_T\right\}. \tag{13.12}$$

It can be interpreted as maximization of the terminal salvage value less the cost of production and inventory assumed to be quadratic. In Exercise 13.1, you will be asked to solve the problem with the objective function given by the expected value of the undiscounted version of the integral in (6.2).

As in Section 6.1.1 we do not restrict the production rate to be nonnegative. In other words, we permit disposal (i.e., $P_t < 0$). While this is done for mathematical expedience, we will state conditions under which a disposal is not required. Note further that the inventory level is allowed to be negative, i.e., we permit backlogging of demand.

The solution of the above model due to Thompson and Sethi (1980) will be carried out via the previous development of the HJB equation satisfied by a certain *value function*.

Let $V(x,t)$ denote the expected value of the objective function from time t to the horizon T with $I_t = x$ and using the optimal policy from t to T. The function $V(x,t)$ is referred to as the value function, and it satisfies the HJB equation

$$0 = \max_u[-(u^2 + x^2) + V_t + V_x(u - S) + \frac{1}{2}\sigma^2 V_{xx}] \tag{13.13}$$

with the boundary condition

$$V(x,T) = Bx. \tag{13.14}$$

Note that these are applications of (13.9) and (13.10) to the production planning problem.

It is now possible to maximize the expression inside the bracket of (13.13) with respect to u by taking its derivative with respect to u and setting it to zero. This procedure yields

$$u^*(x,t) = \frac{V_x(x,t)}{2}. \tag{13.15}$$

Substituting (13.15) into (13.13) yields the equation

$$0 = \frac{V_x^2}{4} - x^2 + V_t - SV_x + \frac{1}{2}\sigma^2 V_{xx}, \tag{13.16}$$

which, after the max operation has been performed, is known as the Hamilton-Jacobi equation. This is a partial differential equation which must be satisfied by the value function $V(x,t)$ with the boundary condition (13.14). The solution of (13.16) is considered in the next section.

Remark 13.1 It is important to remark that if production rate were restricted to be nonnegative, then, as in Remark 6.1, (13.15) would be changed to

$$u^*(x,t) = \max\left[0, \frac{V_x(x,t)}{2}\right]. \tag{13.17}$$

Substituting (13.17) into (13.14) would give us a partial differential equation which must be solved numerically. We will not consider (13.17) further in this chapter.

13.2.1 Solution for the Production Planning Problem

To solve equation (13.16) we let

$$V(x,t) = Q(t)x^2 + R(t)x + M(t). \tag{13.18}$$

Then,

$$\begin{aligned} V_t &= \dot{Q}x^2 + \dot{R}x + \dot{M}, && (13.19)\\ V_x &= 2Qx + R, && (13.20)\\ V_{xx} &= 2Q, && (13.21) \end{aligned}$$

where $\dot{Y}$ denotes dY/dt. Substituting (13.19) in (13.16) and collecting terms gives

$$x^2[\dot{Q} + Q^2 - 1] + x[\dot{R} + RQ - 2SQ] + \dot{M} + \frac{R^2}{2} - RS + \sigma^2 Q = 0. \tag{13.22}$$

Since (13.22) must hold for any value of x, we must have

$$\begin{aligned} \dot{Q} &= 1 - Q^2, \quad Q(T) = 0, && (13.23)\\ \dot{R} &= 2SQ - RQ, \quad R(T) = B, && (13.24)\\ \dot{M} &= RS - \frac{R^2}{4} - \sigma^2 Q, \quad M(T) = 0, && (13.25) \end{aligned}$$

where the boundary conditions for the system of simultaneous differential equations (13.23), (13.24), and (13.25) are obtained by comparing (13.18) with the boundary condition $V(x,T) = Bx$ of (13.14).

To solve (13.23), we expand $\dot{Q}/(1-Q^2)$ by partial fractions to obtain

$$\frac{\dot{Q}}{2}\left[\frac{1}{1-Q} + \frac{1}{1+Q}\right] = 1,$$

which can be easily integrated. The answer is

$$Q = \frac{y-1}{y+1}, \tag{13.26}$$

where

$$y = e^{2(t-T)}. \tag{13.27}$$

Since S is assumed to be a constant, we can reduce (13.24) to

$$\dot{R}^0 + R^0 Q = 0, \quad R^0(T) = B - 2S$$

by the change of variable defined by $R^0 = R - 2S$. Clearly the solution is given by

$$\log R^0(T) - \log R^0(t) = -\int_t^T Q(\tau)d\tau,$$

which can be simplified further to obtain

$$R = 2S + \frac{2(B-2S)\sqrt{y}}{y+1}. \tag{13.28}$$

Having obtained solutions for R and Q, we can easily express (13.25) as

$$M(t) = -\int_t^T [R(\tau)S - (R(\tau))^2/4 - \sigma^2 Q(\tau)]d\tau. \tag{13.29}$$

The optimal control is defined by (13.15), and the use of (13.26) and (13.28) yields

$$u^*(x,t) = V_x/2 = Qx + R/2 = S + \frac{(y-1)x + (B-2S)\sqrt{y}}{y+1}. \tag{13.30}$$

This means that the optimal production rate for $t \in [0,T]$

$$P_t^* = u^*(I_t^*, t) = S + \frac{(e^{2(t-T)} - 1)I_t^* + (B-2S)e^{(t-T)}}{e^{2(t-T)} + 1}, \tag{13.31}$$

where $I_t^*,\ t \in [0,T]$, is the inventory level observed at time t when using the optimal production rate $P_t^*,\ t \in [0,T]$, according to (13.31).

Remark 13.2 The optimal production rate in (13.30) equals the demand rate plus a correction term which depends on the level of inventory and the distance from the horizon time T. Since $(y-1) < 0$ for $t < T$, it is clear that for lower values of x, the optimal production rate is likely to be positive. However, if x is very high, the correction term will become smaller than $-S$, and the optimal control will be negative. In other words, if inventory level is too high, the factory can save money by disposing a part of the inventory resulting in lower holding costs.

Remark 13.3 If the demand rate S were time-dependent, it would have changed the solution of (13.24). Having computed this new solution in place of (13.28), we can once again obtain the optimal control as $u^*(x,t) = Qx + R/2$.

Remark 13.4 Note that when $T \to \infty$, we have $y \to 0$ and

$$u^*(x,t) \to S - x, \tag{13.32}$$

but the undiscounted objective function value (13.12) in this case becomes $-\infty$. Clearly, any other policy will render the objective function value to be $-\infty$. In a sense, the optimal control problem becomes ill-posed. One way to get out of this difficulty is to impose a nonzero discount rate. You are asked to carry this out in Exercise 13.2.

Remark 13.5 It would help our intuition if we could draw a picture of the path of the inventory level over time. Since the inventory level is a stochastic process, we can only draw a typical sample path. Such a sample path is shown in Figure 13.1. If the horizon time T is long enough, the optimal control will bring the inventory level to the goal level $\bar{x} = 0$. It will then hover around this level until t is sufficiently close to the horizon T. During the ending phase, the optimal control will try to build up the inventory level in response to a positive valuation B for ending inventory.

Figure 13.1: A Sample Path of X_t with $X_0 = x_0 > 0$ and $B > 0$

13.3 A Stochastic Advertising Problem

In this section, we will discuss a stochastic advertising model due to Sethi (1983b). The model is:

$$\begin{cases} \max\ E\left[\int_0^\infty e^{-\rho t}(\pi X_t - U_t^2)dt\right] \\ \text{subject to} \\ dX_t = (rU_t\sqrt{1-X_t} - \delta X_t)dt + \sigma(X_t)dZ_t,\ X_0 = x_0, \\ U_t \geq 0, \end{cases} \tag{13.33}$$

where X_t is the market share and U_t is the rate of advertising at time t, and where, as specified in Section 7.2.1, $\rho > 0$ is the discount rate, $\pi > 0$ is the profit margin on sales, $r > 0$ is the advertising effectiveness parameter, and $\delta > 0$ is the sales decay parameter. Furthermore, Z_t is the standard one-dimensional Wiener process and $\sigma(x)$ is the diffusion coefficient function having some properties to be specified shortly. The term in the integrand represents the discounted profit rate at time t. Thus, the term in the square bracket represents the total discounted profits on a sample path. The objective in (13.33) is, therefore, to maximize the expected value of the total discounted profits.

This model is a modification as well as a stochastic extension of the optimal control formulation of the Vidale-Wolfe advertising model presented in (7.43). The Itô equation in (13.33) modifies the Vidale-Wolfe dynamics (7.25) by replacing the term $ru(1-x)$ by $rU_t\sqrt{1-X_t}$ and

adding a diffusion term $\sigma(X_t)dZ_t$ on the right-hand side. Furthermore, we replace the linear cost of advertising u in (7.43) by a quadratic cost of advertising U_t^2 in (13.33). We also relax the control constraint $0 \leq u \leq Q$ in (7.43) to simplify $U_t \geq 0$. The addition of the diffusion term yields a stochastic optimal control problem as expressed in (13.33).

An important consideration in choosing the function $\sigma(x)$ should be that the solution X_t to the Itô equation in (13.33) remains inside the interval $[0, 1]$. Merely requiring that the initial condition $x_0 \in [0, 1]$, as in Section 7.2.1, is no longer sufficient in the stochastic case. Additional conditions need to be imposed. It is possible to specify these conditions by using the theory presented by Gihman and Skorohod (1972) for stochastic differential equations on a finite spatial interval. In our case, the conditions boil down to the following, in addition to $x_0 \in (0, 1)$, which has been assumed already in (13.33):

$$\sigma(x) > 0,\ x \in (0, 1) \text{ and } \sigma(0) = \sigma(1) = 0. \tag{13.34}$$

It is possible to show that for any feedback control $u(x)$ satisfying

$$u(x) \geq 0,\ x \in (0, 1],\ \text{and } u(0) > 0, \tag{13.35}$$

the Itô equation in (13.33) will have a solution X_t such that $0 < X_t < 1$, *almost surely* (i.e., with probability 1). Since our solution for the optimal advertising $u^*(x)$ would turn out to satisfy (13.35), we will have the optimal market share X_t^* lie in the interval $(0, 1)$.

Let $V(x)$ denote the value function for the problem, i.e., $V(x)$ is the expected value of the discounted profits from time t to infinity. When $X_t = x$ and an optimal policy U_t^* is followed from time t onwards. Note that since $T = \infty$, the future looks the same from any time t, and therefore the value function does not depend on t. It is for this reason we have defined the value function as $V(x)$, rather than $V(x, t)$ as in the previous section.

Using now the principle of optimality as in Section 13.1, we can write the HJB equation as

$$\rho V(x) = \max_u \left[\pi x - u^2 + V_x(ru\sqrt{1-x} - \delta x) + V_{xx}(\sigma(x))^2/2\right]. \tag{13.36}$$

Maximization of the RHS of (13.36) can be accomplished by taking its derivative with respect to u and setting it to zero. This gives

$$u^*(x) = \frac{rV_x\sqrt{1-x}}{2}. \tag{13.37}$$

Substituting of (13.37) in (13.36) and simplifying the resulting expression yields the HJB equation

$$\rho V(x) = \pi x + \frac{V_x^2 r^2 (1-x)}{4} - V_x \delta x + \frac{1}{2}\sigma^2(x) V_{xx}. \tag{13.38}$$

As shown in Sethi (1983b), a solution of (13.38) is

$$V(x) = \bar{\lambda} x + \frac{\bar{\lambda}^2 r^2}{4\rho}, \tag{13.39}$$

where

$$\bar{\lambda} = \frac{\sqrt{(\rho+\delta)^2 + r^2\pi} - (\rho+\delta)}{r^2/2}, \tag{13.40}$$

as derived in Exercise 7.37. In Exercise 13.3, you are asked verify that (13.39) and (13.40) solve the HJB equation (13.38).

We can now obtain the explicit formula for the optimal feedback control as

$$u^*(x) = \frac{r\bar{\lambda}\sqrt{1-x}}{2}. \tag{13.41}$$

Note that $u^*(x)$ satisfies the conditions in (13.35).

As in Exercise 7.37, it is easy to characterize (13.41) as

$$U_t^* = u^*(X_t) = \begin{cases} > \bar{u} & \text{if } X_t < \bar{x}, \\ = \bar{u} & \text{if } X_t = \bar{x}, \\ < \bar{u} & \text{if } X_t > \bar{x}, \end{cases} \tag{13.42}$$

where

$$\bar{x} = \frac{r^2\bar{\lambda}/2}{r^2\bar{\lambda}/2 + \delta} \tag{13.43}$$

and

$$\bar{u} = \frac{r\bar{\lambda}\sqrt{1-\bar{x}}}{2}, \tag{13.44}$$

as given in (7.51).

The market share trajectory for X_t is no longer monotone because of the random variations caused by the diffusion term $\sigma(X_t)dZ_t$ in the Itô equation in (13.33). Eventually, however, the market share process hovers around the equilibrium level $\bar{x}$. It is, in this sense and as in the previous section, also a turnpike result in a stochastic environment.

13.4 An Optimal Consumption-Investment Problem

In Example 1.3 in Chapter 1, we had formulated a problem faced by Rich Rentier who wants to consume his wealth in a way that will maximize his total utility of consumption and bequest. In that example, Rich Rentier kept his money in a savings plan earning interest at a fixed rate of $r > 0$.

In this section we will offer Rich, a possibility of investing a part of his wealth in a risky security or stock that earns an expected rate of return that equals $\alpha > r$. The problem of Rich, known now as Rich Investor, is to optimally allocate his wealth between the risk-free savings account and the risky stock over time and consume over time so as to maximize his total utility of consumption. We will assume an infinite horizon problem in lieu of the bequest, for convenience in exposition. One could, however, argue that Rich's bequest would be optimally invested and consumed by his heir, who in turn would leave a bequest that would be optimally invested and consumed by a succeeding heir and so on. Thus, if Rich considers the utility accrued to all his heirs as his own, then he can justify solving an infinite horizon problem without a bequest.

In order to formulate the stochastic optimal control problem of Rich Investor, we must first model his investments. The savings account is easy to model. If S_0 is initial price of a unit of investment in the savings account earning an interest at the rate $r > 0$, then we can write the accumulated amount S_t at time t as

$$S_t = S_0 e^{rt}.$$

This can be expressed as a differential equation, $dS_t/dt = rS_t$, which we will rewrite as

$$dS_t = rS_t dt, \ S_0 \text{ given.} \tag{13.45}$$

Modelling the stock is much more complicated. Merton (1971) and Black and Scholes (1973) have proposed that the stock price P_t can be modelled by an Itô equation, namely,

$$\frac{dP_t}{P_t} = \alpha dt + \sigma dZ_t, \ P_0 \text{ given,} \tag{13.46}$$

or simply,

$$dP_t = \alpha P_t dt + \sigma P_t dZ_t, \ P_0 \text{ given,} \tag{13.47}$$

where α is the average rate of return on stock, σ is the standard deviation associated with the return, and Z_t is a standard Wiener process.

Remark 13.6 The LHS in (13.46) can be written also as $d\ln P_t$. Another name for the process Z_t is *Brownian Motion.* Because of these, the price process P_t given by (13.46) is often referred to as a *logarithmic Brownian Motion.*

In order to complete the formulation of Rich's stochastic optimal control problem, we need the following additional notation:

$$\begin{aligned}
W_t &= \text{the wealth at time } t, \\
C_t &= \text{the consumption rate at time } t, \\
Q_t &= \text{the fraction of the wealth invested in stock at time } t, \\
1 - Q_t &= \text{the fraction of the wealth kept in the savings account} \\
&\quad \text{at time } t, \\
U(c) &= \text{the utility of consumption when consumption is at the} \\
&\quad \text{rate } c\text{; the function } U(c) \text{ is assumed to be increasing} \\
&\quad \text{and concave,} \\
\rho &= \text{the rate of discount applied to consumption utility,} \\
B &= \text{the bankruptcy parameter to be explained later.}
\end{aligned}$$

Next we develop the dynamics of the wealth process. Since the investment decision Q is unconstrained, it means Rich is allowed to buy stock as well as to sell it short. Moreover, Rich can deposit in, as well as borrow money from, the savings account at the rate r.

While it is possible to obtain rigorously the equation for the wealth process involving an intermediate variable, namely, the number N_t of shares of stock owned at time t, we will not do so. Instead, we will write the wealth equation informally as

$$\begin{aligned}
dW_t &= Q_t W_t \alpha dt + Q_t W_t \sigma dZ_t + (1 - Q_t) W_t r dt - C_t dt \\
&= (\alpha - r) Q_t W_t dt + (r W_t - C_t) dt + \sigma Q_t W_t dZ_t, \ W_0 \text{ given,}
\end{aligned} \tag{13.48}$$

and provide an intuitive explanation for it. The term $Q_t W_t \alpha dt$ represents the expected return from the risky investment of $Q_t W_t$ dollars during the period from t to $t+dt$. The term $Q_t W_t \sigma dZ_t$ represents the risk involved in investing $Q_t W_t$ dollars in stock. The term $(1 - Q_t) W_t r dt$ is the amount of interest earned on the balance of $(1 - Q_t) W_t$ dollars in the savings account. Finally, $C_t dt$ represent the amount of consumption during the interval from t to $t + dt$.

In deriving (13.48), we have assumed that Rich can trade continuously in time without incurring any broker's commission. Thus, the change in wealth dW_t from time t to time $t + dt$ is due only to capital gains from change in share price and to consumption. For a rigorous development of (13.48) from (13.45) and (13.46), see Harrison and Pliska (1981).

Since Rich can borrow an unlimited account and invest it in stock, his wealth could fall to zero at some time T. We will say that Rich goes bankrupt at time T, when his wealth falls zero at that time. It is clear that T is a random variable. It is, however, a special type of random variable, called a *stopping time*, since it is observed exactly at the instant of time when wealth falls to zero.

We can now specify Rich's objective function. It is:

$$\max \left\{ J = E\left[\int_0^T e^{-\rho t} U(C_t) dt + e^{-\rho T} B \right] \right\}, \tag{13.49}$$

where we have assumed that Rich experiences a payoff of B, in the units of utility, at the time of bankruptcy. B can be positive if there is a social welfare system in place, or B can be negative if there is remorse associated with bankruptcy. See Sethi (1997a) for a detailed discussion of the bankruptcy parameter B.

Let us recapitulate the optimal control problem of Rich Investor:

$$\begin{cases} \max \left\{ J = E\left[\int_0^T e^{-\rho t} U(C_t) dt + e^{-\rho T} B \right] \right\} \\ \text{subject to} \\ dW_t = (\alpha - r) Q_t W_t dt + (r W_t - C_t) dt + \sigma Q_t W_t dZ_t, \ W_0 \text{ given}, \\ C_t \geq 0. \end{cases} \tag{13.50}$$

As in the infinite horizon problem of Section 13.2, here also the value function is stationary with respect to time t. This is because T is a stopping time of bankruptcy, and the future evolution of wealth, investment, and consumption processes from any starting time t depends only on the wealth at time t and *not* on time t itself. Therefore, let $V(x)$ be the value function associated with an optimal policy beginning with wealth $W_t = x$ at time t. Using the principle of optimality as in Section 13.1, the HJB equation satisfied by the value function $V(x)$ for problem (13.50)

can be written as

$$\begin{cases} \rho V(x) = \max\limits_{c\geq 0, q} & [(\alpha - r)qxV_x + (rx - c)V_x \\ & +(1/2)q^2\sigma^2x^2V_{xx} + U(c)], \\ V(0) = B. \end{cases} \tag{13.51}$$

This problem and number of its generalizations can be solved explicitly; see Sethi (1997a).

For the purpose of this section, we will simplify the problem by making further assumptions. Let

$$U(c) = \ln c, \tag{13.52}$$

as was used in Example 1.3. This utility has an important simplifying property, namely,

$$U'(0) = 1/c|_{c=0} = \infty. \tag{13.53}$$

We also assume $B = -\infty$. See Sethi (1997a, Chapter 2) for solutions when $B > -\infty$.

Under these assumptions, Rich would be sufficiently conservative in his investments so that he does not go bankrupt. This is because bankruptcy at time t means $W_t = 0$, implying "zero consumption" thereafter, and a small amount of wealth would allow Rich to have nonzero consumption resulting in a proportionally large amount of utility on account of (13.53). While we have provided an intuitive explanation, it is possible to show rigorously that condition (13.53) together with $B = -\infty$ implies a strictly positive consumption level at all times and no bankruptcy.

Since Q is already unconstrained, having no bankruptcy and only positive (i.e., interior) consumption level allows us to obtain the form of the optimal consumption and investment policy simply by differentiating the RHS of (13.51) with respect to q and c and equating the resulting expressions to zero. Thus,

$$(\alpha - r)xV_x + q\sigma^2x^2V_{xx} = 0,$$

i.e.,

$$q^*(x) = -\frac{(\alpha - r)V_x}{x\sigma^2V_{xx}}, \tag{13.54}$$

and

$$c^*(x) = \frac{1}{V_x}. \tag{13.55}$$

Substituting (13.54) and (13.55) in (13.51) allows us to remove the max operator from (13.51), and provides us with the equation

$$\rho V(x) = -\frac{\gamma (V_x)^2}{V_{xx}} + \left(rx - \frac{1}{V_x}\right) V_x - \ln V_x, \tag{13.56}$$

where

$$\gamma = \frac{(\alpha - r)^2}{2\sigma^2}. \tag{13.57}$$

This is a nonlinear ordinary differential equation that appears to be quite difficult to solve. However, Karatzas, Lehoczky, Sethi, and Shreve (1986) used a change of variable that transforms (13.56) into a second-order, linear, ordinary differential equation, which has a known solution. For our purposes, we will guess that the value function is of the form

$$V(x) = A \ln x + B, \tag{13.58}$$

where A and B are constants, and obtain the values of A and B by substitution in (13.56). Using (13.58) in (13.56), we see that

$$\begin{aligned} \rho A \ln x + \rho B &= \gamma A + \left(rx - \frac{x}{A}\right) \frac{A}{x} - \ln\left(\frac{A}{x}\right) \\ &= \gamma A + rA - 1 - \ln A + \ln x. \end{aligned}$$

By comparing the coefficients of $\ln x$ and the constants on both sides, we get $A = 1/\rho$ and $B = \frac{r-\rho+\gamma}{\rho^2} + \frac{\ln \rho}{\rho}$. By substituting these values in (13.58), we obtain

$$V(x) = \frac{1}{\rho} \ln(\rho x) + \frac{r - \rho + \gamma}{\rho^2}, \; x \geq 0. \tag{13.59}$$

In Exercise 13.4, you are asked by a direct substitution in (13.56) to verify that (13.59) is indeed a solution of (13.56). Moreover, $V(x)$ defined in (13.59) is strictly concave, so that our concavity assumption made earlier is justified.

From (13.59), it is easy to show that (13.54) and (13.55) yield the following feedback policies:

$$q^*(x) = \frac{\alpha - r}{\sigma^2}, \tag{13.60}$$

$$c^*(x) = \rho x. \tag{13.61}$$

The investment policy (13.60) says that the optimal fraction of the wealth invested in the risky stock is $(\alpha - r)/\sigma^2$, i.e.,

$$Q_t^* = q^*(W_t) = \frac{\alpha - r}{\sigma^2},\ t \geq 0, \tag{13.62}$$

which is a constant over time. The optimal consumption policy is to consume a constant fraction ρ of the current wealth, i.e.,

$$C_t^* = c^*(W_t) = \rho W_t,\ t \geq 0. \tag{13.63}$$

This problem and its many extensions have been studied in great detail. See, e.g., Sethi (1997a).

13.5 Concluding Remarks

In this chapter, we have considered stochastic optimal control problems subject to Itô differential equations. For impulse stochastic control, see Bensoussan and Lions (1984). For stochastic control problems with jump Markov processes or, more generally, martingale problems, see Fleming and Soner (1992), Davis (1993), and Karatzas and Shreve (1998). For problems with incomplete information or partial observation, see Bensoussan (2004), Elliott, Aggoun, and Moore (1995), and Bensoussan, Çakanyildirim, and Sethi (2010).

For applications of stochastic optimal control to manufacturing problems, see Sethi and Zhang (1994a), Yin and Zhang (1997), Sethi, Zhang, and Zhang (2005), Bensoussan et al. (2009,2008a,2008b), and Bensoussan, Çakanyildirim, and Sethi (2007a,2007b,2007c,2009). For applications to problems in finance, see Sethi (1997a), Karatzas and Shreve (1998), and Bensoussan, Keppo, and Sethi (2009). For applications in marketing, see Tapiero (1988), Raman (1990), and Sethi and Zhang (1995). For applications of stochastic optimal control to economics including economics of natural resources, see, e.g., Pindyck (1978a, 1978b), Rausser and Hochman (1979), Arrow and Chang (1980), Derzko and Sethi (1981a), Bensoussan and Lesourne (1980, 1981), Malliaris and Brock (1982), and Brekke and Øksendal (1994).

EXERCISES FOR CHAPTER 13

E 13.1 Solve the production-inventory problem with the state equation (13.11) and the objective function

$$\min\left\{J = E\int_0^T [\frac{h}{2}(I-\hat{I})^2 + \frac{c}{2}(P-\hat{P})^2]dt\right\},$$

where $h > 0, c > 0, \hat{I} \geq 0$ and $\hat{P} \geq 0$; see the objective function (6.2) for the interpretation of these parameters.

E 13.2 Formulate and solve the discounted infinite-horizon version of the stochastic production planning model of Section 13.2. Specifically, assume $B = 0$ and replace the objective function in (13.12) by

$$\max E\left\{\int_0^\infty -e^{-\rho t}(P_t^2 + I_t^2)dt\right\}.$$

E 13.3 Verify by direct substitution that the value function defined by (13.39) and (13.40) solves the HJB equation (13.38).

E 13.4 Verify by direct substitution that the value function in (13.59) solves the HJB equation (13.56).

E 13.5 Solve the consumption-investment problem (13.50) with the utility function

$$U(c) = c^\beta,\ 0 < \beta < 1,$$

and $B = 0$.

E 13.6 Solve Exercise 13.5 when $U(c) = -c^\beta$ with $\beta < 0$ and $B = -\infty$.

E 13.7 Solve the optimal consumption-investment problem:

$$V(x) = \max\left\{J = E\left[\int_0^\infty e^{-\rho t}\ln(C_t - s)dt\right]\right\}$$

subject to

$$\begin{aligned} dW_t &= (\alpha - r)Q_tW_tdt + (rW_t - C_t)dt + \sigma Q_tW_tdZ_t,\ W_0 = x, \\ C_t &\geq s. \end{aligned}$$

Here $s > 0$ denotes a minimal subsistence consumption, and we assume $0 < \rho < 1$. Note that the value function $V(s/r) = -\infty$. Guess a solution of the form

$$V(x) = A\ln(x - s/r) + B.$$

Find the constants A, B, and the optimal feedback consumption and investment allocation policies $c^*(x)$ and $q^*(x)$, respectively. Characterize these policies in words.

E 13.8 Solve the consumption-investment problem:

$$\begin{aligned} V(x) &= \max\left\{J = E\left[\int_0^\infty e^{-\rho t}(C_t - s)^\beta dt\right]\right\} \\ \text{subject to} & \\ dW_t &= (\alpha - r)Q_t W_t dt + (rW_t - C_t)dt + \sigma Q_t W_t dZ_t,\ W_0 = x, \\ C_t &\geq s. \end{aligned}$$

Here $s > 0$ denotes a minimal subsistence consumption and we assume $0 < \rho < 1$ and $0 < \beta < 1$. Note that the value function $V(s/r) = 0$. Therefore, guess a solution of the form

$$V(x) = A(x - s/r)^\beta.$$

Find the constant A and the optimal feedback consumption and investment allocation policies $c^*(x)$ and $q^*(x)$, respectively. Characterize these policies in words.